

TESI DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI NAPOLI “FEDERICO II”

DIPARTIMENTO DI INGEGNERIA ELETTRONICA  
E DELLE TELECOMUNICAZIONI

DOTTORATO DI RICERCA IN  
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

---

THE EXPLOITATION OF  
WIDELY LINEAR FILTERING IN  
MIMO COMMUNICATION SYSTEMS

---

**FABIO STERLE**

Il Coordinatore del Corso di Dottorato

Ch.mo Prof. Giovanni POGGI

Il Tutore

Ch.mo Prof. Davide MATTERA

A. A. 2004–2005



*“Io tengo per fermo ch’ e’ procurasse prima,  
per via de’ sensi, dell’esperienze e dell’osservazioni,  
di assicurarsi quanto fusse possibile della  
conclusione, e che doppo andasse ricercando i  
mezi da poterla dimostrare.”*

*Galileo Galilei, Dialogo sopra i due massimi sistemi.*



# Acknowledgments

First, I want to express my sincere gratitude to Professor Davide Mattera and Professor Luigi Paura for their valuable guidance and support during my time as a Ph.D. student. I think that the results here collected are the fruits of a team-working. I also thank them for proof-reading this manuscript.

I gratefully acknowledge all the people of the Telecommunication Group for creating such a pleasant work environment.

Last but certainly not the least, I want to thank my family for their support and encouragement, and definitely not the least, Manuela, for her patience and understanding.

Fabio Sterle



# Contents

<b>Acknowledgments</b>	<b>v</b>
<b>List of Figures</b>	<b>xi</b>
<b>Introduction</b>	<b>xv</b>
<b>1 MIMO model</b>	<b>1</b>
1.1 Multiple-input multiple-output channel model . . . . .	2
1.2 Multipath fading channel . . . . .	5
1.3 Time-dispersive FIR MIMO channel . . . . .	10
1.4 MIMO in communications . . . . .	14
1.4.1 Multiple antenna systems . . . . .	14
1.4.2 DS-CDMA systems . . . . .	17
1.4.3 OFDM systems . . . . .	19
1.4.4 Fractionally spaced sampling . . . . .	21
1.5 Capacity of MIMO systems . . . . .	23
1.5.1 Capacity of deterministic MIMO channel . . . . .	24
1.5.2 Ergodic capacity of fading MIMO channel . . . . .	25
1.5.3 Outage capacity . . . . .	32
1.6 Multiplexing, diversity and array gain . . . . .	33
1.7 MIMO receiver architectures . . . . .	37
1.7.1 ML receiver . . . . .	38
1.7.2 MMSE FIR equalizer . . . . .	39
1.7.3 ZF FIR equalizer . . . . .	41
1.7.4 Decision-feedback FIR equalizers . . . . .	42
1.8 Transceiver architectures . . . . .	45

<b>2</b>	<b>Decision-feedback equalization</b>	<b>49</b>
2.1	Widely linear processing in the MMSE estimation . . . . .	50
2.2	Widely linear transformations . . . . .	54
2.2.1	Real-valued and complex-valued representation . . . . .	55
2.2.2	WL transformations . . . . .	57
2.3	System model . . . . .	58
2.4	DF-MMSE FIR equalizer . . . . .	60
2.4.1	DF-MMSE for Scenario 1 . . . . .	60
2.4.2	DF-MMSE for Scenario 2 . . . . .	62
2.4.3	DF-MMSE for Scenario 3 . . . . .	63
2.4.4	Detection ordering . . . . .	64
2.5	WL decision-feedback FIR equalizer . . . . .	67
2.6	WL-WDF-MMSE equalizer for Scenario 1 . . . . .	70
2.6.1	Reduced-order design of WL-WDF-MMSE equalizer . . . . .	71
2.6.2	WL-DF-MMSE equalizer for Scenario 1 . . . . .	77
2.7	WL-WDF-MMSE equalizer for Scenario 2 . . . . .	80
2.8	Nonequivalent WL-WDF equalizers . . . . .	84
2.9	Mismatching analysis . . . . .	86
2.10	Numerical results . . . . .	88
2.10.1	Numerical results in Scenario 1 . . . . .	88
2.10.2	Numerical results in Scenario 2 . . . . .	96
<b>3</b>	<b>Transceiver design</b>	<b>103</b>
3.1	System Model . . . . .	104
3.1.1	Real-valued system model . . . . .	105
3.1.2	Complex-valued system model . . . . .	107
3.2	MMSE linear transceiver design . . . . .	108
3.3	MMSE Widely Linear transceiver design . . . . .	111
3.4	MSE Analysis . . . . .	114
3.4.1	Real-valued information symbols and real-valued transmitted symbols ( $b_r = B, n_r = N_i$ ) . . . . .	115
3.4.2	Real-valued information symbols and complex-valued transmitted symbols ( $b_r = B, n_r = 0$ ) . . . . .	118
3.4.3	Complex-valued information symbols and complex- valued transmitted symbols ( $b_r = 0, n_r = 0$ ) . . . . .	121
3.5	SER analysis . . . . .	122
3.6	Numerical results . . . . .	124
	<b>Conclusion</b>	<b>131</b>

---

**Appendix A** **133**

- A.1 (Non)-circular random variables and vectors . . . . . 133
  - A.1.1 Definition of circularity . . . . . 134
  - A.1.2 Second order analysis of a complex-valued random vector . . . . . 135

**Appendix B** **139**



# List of Figures

1.1	Analogic low-pass equivalent MIMO model. . . . .	3
1.2	Channel classification: $T_s$ is the symbol period, $B$ the bandwidth. . . . .	10
1.3	Block diagram of the LTI time-dispersive MIMO channel of order $\nu$ . . . . .	13
1.4	Multiantenna system. . . . .	15
1.5	Illustration of the abstract <i>one ring</i> model. . . . .	16
1.6	Fractionally spaced sampling leads to a stationary multiple output model. . . . .	22
1.7	Capacity of a MISO ( $N_o = 1$ ) system versus the number of inputs $N_i$ for different values of $\mathcal{P}_0$ . . . . .	27
1.8	Capacity of a SIMO ( $N_i = 1$ ) system versus the number of outputs $N_o$ for different values of $\mathcal{P}_0$ . . . . .	28
1.9	Capacity of a square MIMO ( $N_i = N_o = N$ ) system versus $N$ for different values of $\mathcal{P}_0$ . . . . .	29
1.10	The Marčenko and Pastur of density of $\mathbf{H}^H\mathbf{H}$ for $\beta = 1, 0.5, 0.1, 0.01$ . . . . .	31
1.11	One hundred realizations of (1.54) compared to the asymptotic capacity in (1.68) versus $N_i = N_o$ for different values of $\mathcal{P}_0$ . . . . .	32
1.12	Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix $\mathbf{H}^H\mathbf{H}$ with $N_i = 2$ and $N_o = 2$ . . . . .	34
1.13	Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix $\mathbf{H}^H\mathbf{H}$ with $N_i = 2$ and $N_o = 6$ . . . . .	35
1.14	Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix $\mathbf{H}^H\mathbf{H}$ with $N_i = 2$ and $N_o = 12$ . . . . .	36
1.15	Block diagram of a linear receiver. . . . .	40
1.16	The decision feedback equalizer structure. . . . .	42
1.17	The decision feedback equalizer in absence of error propagation. . . . .	44
1.18	The transceiver architecture. . . . .	45
1.19	The transceiver architecture. . . . .	46

2.1	Two stage equalizer structure: the former stage performs the MMSE equalization by utilizing only past decisions (Scenario 1), the latter performs the MMSE equalization by utilizing only current decisions (Scenario 2).	66
2.2	Block diagram of the WL-WDF-MMSE equalizer: the blocks labelled with $L$ are linear filters, whereas the ones labelled with $WL$ are widely linear filters.	77
2.3	Statistical properties of the interference for different decision ordering and complex-valued circularly symmetric transmitted sequences: C.S.I. stands for circularly symmetric interference, whereas C.V.I. stands for circularly variant interference.	84
2.4	ASNR of different equalizers versus $\text{SNR}_i$ .	90
2.5	Performance gain of the WL-WDF-MMSE equalizer over DF-MMSE (a) and WL-DF-MMSE (b) equalizers versus the ratio $\frac{n_c}{N_i}$ for different values of $N_i = N_o$ .	91
2.6	Effects of the number of feedforward taps $N_f$ .	92
2.7	Variation of decision-point ASNR with processing delay $\Delta$ .	92
2.8	Effect of channel mismatching percentage on decision-point ASNR.	93
2.9	Effect of channel mismatch on decision-point ASNR versus $\text{SNR}_i$ .	95
2.10	Effect of channel mismatch on decision-point ASNR versus $N_f$ .	95
2.11	Effect of channel mismatch on decision-point ASNR versus $\Delta$ .	96
2.12	ASNRs of the equalizers versus $\text{SNR}_i$ ( $N_o = 2$ ).	97
2.13	Gains $[G_i]_{i=1,\dots,4}$ versus the number of outputs $n_o$ for $\text{SNR}_i = 15\text{dB}$ (a) and $\text{SNR}_i = 30\text{dB}$ (b).	98
2.14	ASNRs of the equalizers versus $N_o$ for $\text{SNR}_i = 15\text{dB}$ (a) and $\text{SNR}_i = 30\text{dB}$ (b).	99
2.15	SERs of the equalizers versus $\text{SNR}_i$ ; $2 \times 1$ MISO channel.	101
2.16	SERs of the equalizers versus $\text{SNR}_i$ ; $2 \times 2$ MIMO channel.	102
3.1	The transceiver architecture.	104
3.2	Block diagram of the widely linear transceiver.	106
3.3	Decomposition of the MIMO channel into <i>eigen subchannels</i> : the gains $\beta_\ell$ are optimized according to the MMSE criterion.	111
3.4	Average gain $G$ in dB versus $N_o \geq 8$ for $\text{SNR}_i = 0\text{dB}, 5\text{dB}, 10\text{dB}, 20\text{dB}$ ( $B = N_i = 8$ ).	125
3.5	Average gain $G$ in dB versus $N_i \geq 8$ for $\text{SNR}_i = 0\text{dB}, 5\text{dB}, 10\text{dB}, 20\text{dB}$ ( $B = N_o = 8$ ).	126

---

3.6	Probability of allocating all the $B$ information symbols versus $N_o$ , for $\text{SNR}_i = 0\text{dB}, 5\text{dB}, 10\text{dB}, 20\text{dB}$ . . . . .	127
3.7	Symbol error rates of the considered transceivers versus $\text{SNR}_i$ , for different values of $N_o$ . . . . .	129
3.8	Symbol error rates of the considered transceivers versus $\text{SNR}_i$ , for different values of $N_i$ . . . . .	130
3.9	Symbol error rate of the considered transceiver over a square MIMO channel of size $N = 6, 7, 8$ . . . . .	130



# Introduction

Handling the still increasing amount of digital traffic is today a large and important research area. As it is well known, the information capacity of the conventional single-input single-output (SISO) systems grows in a logarithmic fashion with the signal-to-noise ratio. Since the signal-to-noise ratio is related to the transmitted power and the ambient noise, one way to answer the demand of high bit-rate services is to increase the transmitted power. Unfortunately, in the modern multi-user environments, the ambient noise is due to other users transmitting within the same frequency bands. It follows that the increase in the transmit power does not guarantee a capacity gain because it implies a proportional increase in the overall disturbance. Moreover, higher power level comes at the cost of a nonlinearity in the power amplifier. For some years, the single-input multiple-output (SIMO) systems, as well as multiple-input single-output (MISO) one, has represented a possible solution in attempting to increase the signal-to-noise ratio, and consequently the capacity of the SISO system. The multiple output (input) character of such communication systems is usually related to the adoption of receive (transmit) diversity techniques. For instance, the use of multi element array at the receiver (transmitter) and one element at the transmitter (receiver), has been extensively studied. Unfortunately, the capacity achieved by the SIMO systems increases very slowly with the number of system outputs, whereas the capacity achieved by the MISO ones rapidly reaches saturation as the number of input increases. The natural conjunction between SIMO and MISO systems leads to the more attractive multiple-input multiple-output (MIMO) systems. Foschini and Telatar showed that a huge capacity gain can be achieved over the SISO, SIMO, and MISO systems: more specifically, the capacity of a MIMO system can grow, in principle, linearly with the minimum over the number of inputs and outputs.

For this reason, over the last decade, there has been a growth of research activity in the area of MIMO systems. The advantages of using a such systems

lie in the following achievable gains:

- *multiplexing gain*
- *diversity gain*
- *array gain*.

The *multiplexing gain* is related to the possibility of breaking up the data stream to be transmitted into several parallel data streams and of transmitting such streams over as many orthogonal subchannels simultaneously and within the same frequency band. The *diversity gain* is related to possibility of transmitting and receiving several replicas of the information signal so that, with high probability, at least one or more of them will not be in a deep fade at any given instant. The diversity techniques have been widely adopted in communications to counteract fading: owing to a large number of inputs and outputs, the MIMO systems allow one to achieve a diversity order higher than the one achieved by the SISO, SIMO and MISO systems. Finally, the *array gain* is related to the increased signal-to-noise ratio measured at the output of the receiver, and it is achieved by coherently combining (at the transmitter side) the signals to be transmitted and by coherently combining (at the receiver side) the received signals.

In the first part of this thesis, the focus is on the design of receiving architectures which allow one to achieve a *multiplexing gain* over linear time-dispersive MIMO channels. As well known, the optimum (in the Maximum Likelihood sense) MIMO channel equalization mainly suffers from the computational complexity; in fact, it exponentially increases both with the number of input sequences and with the channel memory length. For such a reason, many suboptimal equalizers have been proposed in order to achieve an acceptable compromise between performance and computational complexity. In addition, the remarkable hostility of wireless communication channels justifies the continuous efforts for optimizing computationally feasible equalization techniques. In the class of symbol-by-symbol equalizers, the decision-feedback (DF) equalization plays an important role since it performs almost as well as the optimum equalizer, but it requires a computational complexity slightly higher than the linear equalizer. According to the DF equalizer structure, the intersymbol interference (ISI) and the co-channel interference are removed via the output of a feedback filter, which processes the past decisions provided by a decision device on the basis of the equalizer output, from the output of the feedforward filter, which processes the received signals. The DF equalizer structure usually employs linear feedforward and feedback filters. In this

thesis, we propose to combine the DF strategy with the widely linear (WL) filtering, which generalizes linear filtering by linearly processing both real and imaginary part of the input signals. More specifically, with reference to MIMO dispersive environment, an equalizer employing WL feedforward and feedback finite impulse response filters is presented. The equalization structure will be synthesized according to the MMSE criterion. It will be shown that, in presence of rotationally variant inputs, such an equalizer outperforms the linear-filtering based DF equalizer by exploiting the statistical redundancy characterizing the channel input.

Moreover, the thesis deals with an important issue concerning the choice between two alternative WL receiver structures. The former performs the linear processing of the real and the imaginary parts of the input signals (as state above), whereas the latter performs the linear processing of the input signals and their conjugate version. In the literature, such receiver structures have been proposed for detection scenarios where they result to be equivalent. In this thesis, we recognize that, if the more general structure of DF equalizer which utilizes also the decisions belonging to the same time step is considered, then the two choices are not anymore equivalent.

The second part of this thesis considers the design of MIMO transceivers with channel state information at both sides of the link. In such a scenario, channel-dependent linear transmit and receive processing of data streams can improve the system performances by optimally allocating resources such as power and bits over the multiple inputs. From an information-theoretic viewpoint, the optimum design in terms of capacity diagonalizes the MIMO channel into *eigen subchannels* over which ideal Gaussian codes have to be transmitted with a water-filling power allocation. With reference to non-dispersive MIMO channels, we propose a WL filtering based transceiver structure, i.e., the one which performs a WL transmit processing as well as a WL receive processing. It will be shown that, if the signals are known to be circularly variant, then the transceiver which employs linear filters is not optimal. The performance comparison between the linear transceiver structure and the WL one is also carried out by using the asymptotic results in random matrix theory, which has drawn considerable attention in the last few years.

The outline of the thesis is the following:

**Chapter 1** presents the general framework. The MIMO system model is introduced and the improvement provided MIMO system over conventional

SISO one are shown from an information-theoretic view-point. Finally, the basic properties of MIMO receivers as well as MIMO transceivers are described.

**Chapter 2** addresses the equalization of time-dispersive MIMO channels in presence of possibly rotationally variant transmitted symbols. The DF equalization strategy and the WL processing are combined to obtain the widely-linear/widely-decision-feedback equalizer. The optimum equalizer is derived according to the minimum mean square error criterion. An analytical calculation of the performance loss due unavoidable mismatches between the actual values of the channel parameters, which are assumed known in the derivation, and those estimated in practice, is provided. Finally, the chapter reports an extensive set of experiments mainly aimed at comparing the performance of the proposed WL equalizer and the linear one.

**Chapter 3** addresses the MIMO transceiver optimization when channel knowledge is available at both sides of the link and when possibly rotationally variant symbols have to be transmitted. A new transceiver structure employing WL transmit and receive filters is presented, and its performances are compared with those of the linear transceiver in terms of mean square error measured at the output of the receiver and in terms of symbol error rate.

# Chapter 1

## MIMO model

The increasing requirements on data rate and quality of service for wireless communications systems call for new techniques to improve spectrum efficiency as well as link reliability. In this context, much attention has been focused on multiple-input multiple-output (MIMO) communication channel models mainly for the following reasons:

- a) the increasing exploitation of multiple antennas both at the transmitter and at the receiver side to introduce spatial redundancy as well as to utilize the recent space-time coding techniques;
- b) the widespread use of multiplexing and multiple access techniques which require to resort to a MIMO channel model, since it can describe the mutual interferences among the different symbol streams. More specifically, resorting to a MIMO model is mandatory in modern communication systems that utilize code-division multiple-access (CDMA) techniques such as direct sequence (DS) CDMA, multi carrier (MC) CDMA and orthogonal frequency division multiplexing (OFDM);
- c) the use of fractionally-spaced equalization where each received signal, after analog filtering, is oversampled with sampling frequency  $p$ -times the symbol rate. In this case, an equivalent baud-rate MIMO model can be adopted where the number of channel outputs becomes  $p$  times larger than that associated with the baud-rate sampling.

In this chapter, after a brief introduction of the mathematical model describing the radio channel propagation, we derive the bandpass equivalent linear-time-invariant (LTI) finite-impulse-response (FIR) MIMO channel

model, whose equalization represents the main subject of the thesis. The versatility of the adopted model is demonstrated by showing that a FIR MIMO model arises in many communication systems such as multiple antenna systems, OFDM systems, CDMA systems, and systems which resort to fractionally spaced sampling techniques.

Following the Telatar's paper [1], the extraordinary improvements in the capacity provided by MIMO systems over the conventional single-input single-output (SISO) systems, the multiple-input single-output (MISO) systems, and single-input multiple output (SIMO) systems, are shown. In addition to *diversity gain* and *array gain*, MIMO links can offer the so called *multiplexing gain* by opening parallel data pipes (usually called *spatial mode* or *eigen subchannels*) within the same frequency band at no additional power expense. In the presence of rich scattering, MIMO links offer capacity gains that are proportional to the minimum of the number of channel inputs and outputs.

Finally, both basic MIMO receiver and transceiver architectures are introduced to provide the general frameworks utilized in Chapter 2 and Chapter 3, respectively. More specifically, when the MIMO channel transfer function is known at only the receiver side, the linear and the decision-feedback equalizers are considered rather than the optimum (in the Maximum Likelihood Sense) due to their lower computational complexity. If the channel state information is available at both ends of the link, the transmitter and the receiver structures can be jointly designed to improve the system performances. At the end of this chapter, the linear transceiver design procedure is described.

## 1.1 Multiple-input multiple-output channel model

The block diagram of the equivalent low-pass continuous-time linear<sup>1</sup> MIMO channel with  $N_i$  input signals and  $N_o$  output signals is depicted in Fig. 1.1. At the transmitter side,  $s_k^{(\ell)}$  denotes the  $k$ th complex-valued symbol to be transmitted by utilizing the  $\ell$ th channel input; the symbol  $T_s$  denotes the symbol period, and  $\psi_T(t)$  is the time-invariant unit-energy impulse response of the transmit filters. The continuous-time linear MIMO channel is characterized by the  $N_o N_i$  time-variant impulse responses  $g^{(\ell,k)}(t, \tau)$  of the linear time-variant (LTV) subchannel connecting the  $\ell$ th channel output with the  $k$ th channel input: note that  $g^{(\ell,k)}(t, \tau)$  is defined as the response at time  $t$  to a unitary Dirac impulse applied to the considered subchannel at the time  $t - \tau$  [2]. At the

<sup>1</sup>The modeling of the channel as a linear system agrees with the observed behavior of a large number of communication channels.

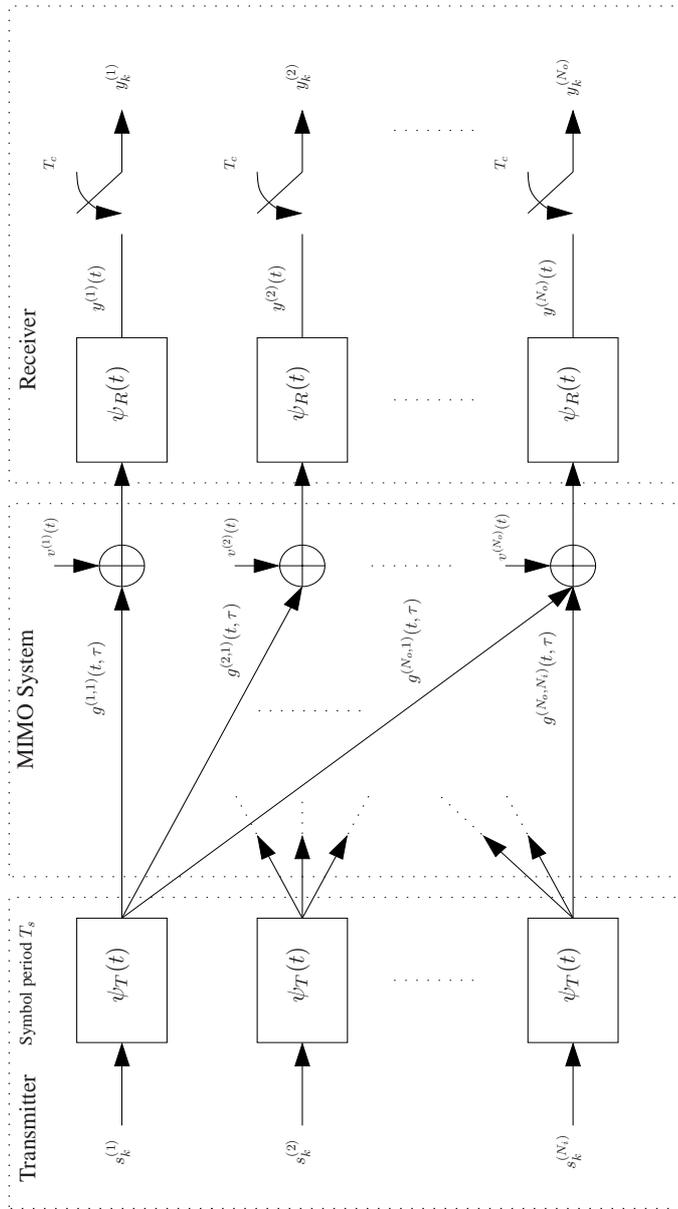


Figure 1.1: Analog low-pass equivalent MIMO model.

receiver side,  $v^{(\ell)}(t)$  represents the additive ambient noise at the  $\ell$ th output, and  $\psi_T(t)$  denotes the time-invariant unit-energy impulse response of the receive filters. The  $\ell$ th channel output  $y^{(\ell)}(t)$  is sampled with sampling period of  $T_c \triangleq T_s/q$ , with  $q \in \mathbb{N}$ . The choice  $q = 1$  accounts for the case where the symbol rate at the transmitter and the sampling rate at the receiver are equal.

According to such a model, the input-output relationship of the MIMO channel can be written as follows:

$$y^{(\ell)}(t) = \sum_{n=-\infty}^{\infty} \sum_{i=1}^{N_i} s_k^{(i)} h^{(\ell,i)}(t, t - nT_s) + n^{(\ell)}(t) \quad \ell = 1, \dots, N_o \quad (1.1)$$

where

$$h^{(\ell,k)}(t, \tau) \triangleq \psi_T(\tau) * g^{(\ell,k)}(t, \tau) * \psi_R(\tau) \quad (1.2)$$

$$n^{(\ell)}(t) \triangleq v^{(\ell)}(t) * \psi_R(t) \quad (1.3)$$

with  $*$  denoting the continuous-time convolution operator. In the following, we will refer to  $h^{(\ell,k)}(t, \tau)$  as the channel impulse response of the subchannel  $(\ell, k)$ , and we will refer to the  $N_o \times N_i$  matrix

$$\mathbf{H}(t, \tau) \triangleq \begin{bmatrix} h^{(1,1)}(t, \tau) & h^{(1,2)}(t, \tau) & \dots & h^{(1,N_i)}(t, \tau) \\ h^{(2,1)}(t, \tau) & h^{(2,2)}(t, \tau) & \dots & h^{(2,N_i)}(t, \tau) \\ \vdots & \vdots & \ddots & \vdots \\ h^{(N_o,1)}(t, \tau) & h^{(N_o,2)}(t, \tau) & \dots & h^{(N_o,N_i)}(t, \tau) \end{bmatrix} \quad (1.4)$$

as the MIMO channel matrix, whereas the column vector  $[h^{(1,i)}(t, \tau) \ h^{(2,i)}(t, \tau) \ \dots \ h^{(N_o,i)}(t, \tau)]^T$  is usually referred to as the signature induced by the  $i$ th input across the channel outputs. In the current literature, the special cases ( $N_i = 1, N_o > 1$ ) and ( $N_i > 1, N_o = 1$ ) are referred to as single-input multiple-output (SIMO) system and multiple-input single-output (MISO) system, respectively; if  $N_i = N_o = 1$ , then the MIMO system degenerates into the conventional SISO system.

The sampled version  $y_k^{(\ell)}$  of the channel output  $y^{(\ell)}(t)$  is equal to

$$\begin{aligned} y_k^{(\ell)} &\triangleq y^{(\ell)}(kT_c) && \ell = 1, \dots, N_o \\ &= \sum_{n=-\infty}^{\infty} \sum_{i=1}^{N_i} s_k^{(i)} h^{(\ell,i)}(kT_c, kT_c - qnT_c) + n_k^{(\ell)} \end{aligned} \quad (1.5)$$

where  $n_k^{(\ell)} \triangleq n^{(\ell)}(kT_c)$  denotes the sampled version of the noise signal  $n^{(\ell)}(t)$ . The discrete-time model in (1.5) is written as a function of the two data rates

$1/T_s$  and  $1/T_c$ . A suitable model, expressed as a function of a single data rate, can be provided by defining the symbol sequences:

$$x_k^{(\ell)} \triangleq \begin{cases} s_{k/q}^{(\ell)} & \text{if } \frac{k}{q} \text{ is integer} \\ 0 & \text{otherwise} \end{cases} . \quad (1.6)$$

Note that  $x_k^{(\ell)}$  represents the oversampled version of  $s_k^{(\ell)}$  obtained by inserting  $q - 1$  zeros between the symbols  $s_k^{(\ell)}$  and  $s_{k-1}^{(\ell)}$ . By utilizing (1.6), the input-output relationship (1.5) can be rewritten as follows

$$y_k^{(\ell)} = \sum_{n=-\infty}^{\infty} \sum_{i=1}^{N_i} x_{k-n}^{(i)} h_{k,n}^{(\ell,i)} + n_k^{(\ell)} \quad \ell = 1, \dots, N_o \quad , \quad (1.7)$$

where  $h_{k,n}^{(\ell,i)} \triangleq h^{(\ell,i)}(kT_c, nT_c)$ . The equation in (1.7) provides the  $T_c$ -space sampled discrete-time MIMO channel model; without loss of generality, in this thesis we will assume  $T_c = T_s$  unless specified.

Let us note that  $g^{(\ell,k)}(t, \tau)$  and  $n^{(\ell)}(t)$  are, in general, complex-valued random processes. More specifically, the channel impulse responses can be modelled as white processes independent of each other, as well as they can be modelled as structured processes and/or correlated with each other when special scenarios are considered. On the other hand, the noise processes can be reasonably modelled as white processes independent of each other and independent of the transmitted signals when they take into account for the effects of the only thermal noise present at the output of the communication channel, while they are typically structured when they take into account for the effects of the co-channel interference.

## 1.2 Multipath fading channel

The characterization of the communication channels  $h^{(\ell,k)}(t, \tau)$  as randomly linear time-variant systems allows one to model many radio channels such as ionospheric radio communication in the HF band, tropospheric scatter radio communications in both the UHF and SHF bands, and ionospheric forward scatter in the VHF band [3]. The time-varying nature of such channels is due to the time-variant physical characteristics of the media and the time variations are described in statistical terms. Because of such unpredictable behavior of the channel, the resultant signal at the receiver will experience *fading*, defined as the changes in the received signal level in time. A widely adopted fading

channel model is the *multipath fading* channel, which allows one to simply describe and analyze propagation situations that include reflection, refraction, and scattering of radio waves. The model assumes that the electromagnetic energy carrying the modulated signal propagates along more than one path connecting the transmitter and receiver. The amplitude, the phase shift, and the time delay introduced by each path vary in time and produce a time-variant destructive and constructive interference which causes fading. In order to provide a characterization of the MIMO channel in (1.4), in this section, we briefly present the main features of the communication channel  $h^{(\ell,k)}(t, \tau)$  when it is modeled as a *multipath fading* channel.

According to the multipath model, the impulse response  $h^{(\ell,k)}(t, \tau)$  can be represented [3, 4] in the form

$$h^{(\ell,k)}(t, \tau) = \sum_{i=1}^{N(t)} \alpha_i^{(\ell,k)}(t) e^{-j2\pi f_0 \tau_i^{(\ell,k)}(t)} \delta\left(\tau - \tau_i^{(\ell,k)}(t)\right) \quad (1.8)$$

where  $\alpha_i^{(\ell,k)}(t)$  denotes the attenuation of the  $i$ th path,  $f_0$  is the carrier frequency,  $\tau_i^{(\ell,k)}(t)$  denotes the time delay that affects the replica of the transmitted signal along the  $i$ th propagation path, and where  $N(t)$  is the number of paths at the time instant  $t$ ; note that, except for the carrier frequency, all these quantities are modeled as random processes. It is well known that large dynamic changes in the medium are required to get a significant variation of  $\alpha_i^{(\ell,k)}(t)$ . On the other hand, the phase shift  $\theta_i^{(\ell,k)}(t) \triangleq 2\pi f_0 \tau_i^{(\ell,k)}(t)$  introduced along the  $i$ th path can noticeably change also in presence of small motions of the medium: in fact,  $\theta_i^{(\ell,k)}(t)$  will change by  $2\pi$  rad whenever  $\tau_i^{(\ell,k)}(t)$  changes by  $1/f_0$ .

According to (1.8), the output of the channel  $h^{(\ell,k)}(t, \tau)$  corresponding to the input signal  $u(t)$  is equal to

$$\begin{aligned} y_u(t) &\triangleq \int_{-\infty}^{\infty} h^{(\ell,k)}(t, \tau) u(t - \tau) d\tau \\ &= \sum_{i=1}^{N(t)} \alpha_i^{(\ell,k)}(t) e^{-j\theta_i^{(\ell,k)}(t)} u\left(\tau - \tau_i^{(\ell,k)}(t)\right), \end{aligned} \quad (1.9)$$

i.e., it is the sum of different replicas of the input waveform affected by different attenuations, phase shifts, and time delays. The fast variations of  $\theta_i^{(\ell,k)}(t)$  produce fast amplitude variations in the received signal, termed *fading*. When  $N(t)$  is sufficiently large, the central limit theorem can be applied, and the

channel response  $h^{(\ell,k)}(t, \tau)$  may be modeled as a complex-valued Gaussian random process<sup>2</sup> in the  $t$  variable. The absence of direct and/or dominant paths allows one to assume that the channel statistics are Gaussian with zero-mean: in such a case the channel is said to be a Rayleigh fading channel, since the envelope  $|h^{(\ell,k)}(t, \tau)|$  at any instant  $t$  is Rayleigh distributed. On the other hand, in presence of a line of sight (LOS) or fixed scatterers, the envelope  $|h^{(\ell,k)}(t, \tau)|$  displays Rice statistics; in such a case, the channel is said to be a Rice fading channel. Finally, a more general fading channel model is described by a Nakagami- $m$  distribution for the envelope of the channel response (see [3] for details).

For some channels, such as the tropospheric scatter channel, it is more appropriate to consider the received signal as consisting of a continuum of multipath components. For such a reason, in the following, we assume  $h^{(\ell,k)}(t, \tau)$  not necessarily expressed as the finite sum of impulsive terms<sup>3</sup> as in (1.8).

The main properties characterizing the fading channel can be studied by introducing the autocorrelation function of the channel impulse response. To this aim, let us assume that  $h^{(\ell,k)}(t, \tau)$  is a Gaussian wide sense stationary (WSS) processes; in such a case it can be fully characterized by its autocorrelation function:

$$E[(h^{(\ell,k)}(t, \tau_1))^* h^{(\ell,k)}(t + \Delta t, \tau_2)] \triangleq \phi_h^{(\ell,k)}(\tau_1, \tau_2, \Delta t) \quad , \quad (1.10)$$

where  $E[\cdot]$  is the statistical expectation operator, and the superscript  $*$  denotes the complex conjugate. Note that, although the fluctuations in the channel are due to non-stationary statistical phenomena, the interest here is in short-term fading which allows us to reasonably assume that  $h^{(\ell,k)}(t, \tau)$  is stationary in a time sense. The function in (1.10) can be seen as the correlation between the two channel impulse responses evaluated at the time instant  $t$  and  $t + \Delta t$ , and corresponding to two different paths  $\tau_1$  and  $\tau_2$ .

The autocorrelation function is usually slightly simplified by assuming that the channel propagation is affected by uncorrelated scattering (US): according to such an assumption, the attenuation and the phase shift introduced by the propagation path with time delay  $\tau_1$  are uncorrelated with the attenuation and the phase shift introduced by the path with time delay  $\tau_2$ . According to the assumption of Gaussian WSSUS channel, the autocorrelation function in (1.10)

<sup>2</sup>The modeling of the wireless channel impulse response as Gaussian process agrees with the observed behavior of many communication channels.

<sup>3</sup>The model in (1.8) is an idealization of the actual behavior of a multipath channel, which would not have such a sharply defined impulse response.

is rewritten as follows:

$$E[(h^{(\ell,k)}(t, \tau_1))^* h^{(\ell,k)}(t + \Delta t, \tau_2)] \triangleq \phi_h^{(\ell,k)}(\tau_1, \Delta t) \delta(\tau_1 - \tau_2) \quad . \quad (1.11)$$

The (autocorrelation) function  $\phi_h^{(\ell,k)}(\tau_1, \Delta t)$  evaluated for  $\Delta t = 0$  is referred to as the *multipath intensity profile* or *delay power spectrum* since it is the average output power of the channel as a function of the time delay  $\tau$ . In fact, by accounting for (1.9), one has that the instantaneous output power is equal to:

$$\begin{aligned} K_{y_u}(t, t) &\triangleq E[y_u(t)y_u^*(t)] & (1.12) \\ &= \int_{-\infty}^{\infty} \phi_h^{(\ell,k)}(\tau_1, 0) |u(t - \tau_1)|^2 \\ &= \phi_h^{(\ell,k)}(t, 0) * |u(t)|^2 \quad , \end{aligned}$$

i.e., the instantaneous input power  $|u(t)|^2$ , which is nonzero over the time interval  $T_u$ , is spread over an interval of duration  $T_u + T_m^{(\ell,k)}$ , with  $T_m^{(\ell,k)}$  denoting the duration of  $\phi_h^{(\ell,k)}(t, 0)$ . For such a reason,  $T_m^{(\ell,k)}$  is called *multipath delay spread*. It follows that, as long as  $T_m^{(\ell,k)} > T_u$ , the received signal is spread in time, and the corresponding channel is said to be time-dispersive: as it is well known, the effect of the time dispersion is the intersymbol interference (ISI). On the contrary, when  $T_u \gg T_m^{(\ell,k)}$ , no linear distortion is present in the received signal and the channel is said to be flat in the frequency domain.

A completely analogous characterization of channel can be provided by defining the transfer function

$$H^{(\ell,k)}(t, f) \triangleq \int_{-\infty}^{\infty} h^{(\ell,k)}(t, \tau) e^{-j2\pi f\tau} d\tau \quad , \quad (1.13)$$

which exhibits the same statistics of  $h^{(\ell,k)}(t, \tau)$ . Under the assumption of WSSUS channel, we define the space-time space-frequency autocorrelation function

$$\phi_H^{(\ell,k)}(\Delta f, \Delta t) = E \left[ \left( H^{(\ell,k)}(t, f_1) \right)^* H^{(\ell,k)}(t + \Delta t, f_2) \right] \quad (1.14)$$

where  $\Delta f = f_2 - f_1$ ; note that  $\phi_H^{(\ell,k)}(\Delta f, \Delta t)$  is the Fourier transform of  $\phi_h^{(\ell,k)}(\tau, \Delta t)$  from  $\tau$  to  $\Delta f$ . The value of  $\phi_H^{(\ell,k)}(\Delta f, 0)$  allows one to understand the behavior of the channel when two sinusoids with frequency separation  $\Delta f$  represent the channel input. When the frequency separation is

such that  $\phi_H^{(\ell,k)}(\Delta f, 0)$  is very near to the maximum value  $\phi_H^{(\ell,k)}(0, 0)$  for all  $\Delta f < (\Delta f)_c^{(\ell,k)}$ , all the transmitted frequencies less than  $(\Delta f)_c^{(\ell,k)}$  will be received fading in highly correlated fashion. For such a reason,  $(\Delta f)_c^{(\ell,k)}$  is called the *coherence bandwidth* and it can be approximated with the inverse of  $T_m^{(\ell,k)}$ . It follows that, as long as the bandwidth  $\Delta f$  of the transmitted signal is larger than  $(\Delta f)_c^{(\ell,k)}$ , the channel exhibits frequency selectivity.

Let us now consider the effects of the time variation of the channel. To this aim, consider the special case where  $u(t) = e^{j2\pi f_1 t}$  is the channel input, and let us evaluate the autocorrelation of the channel output  $y_u(t)$ :

$$\begin{aligned} K_{y_u}(t, \Delta t) &\triangleq E [y_u(t)y_u^*(t + \Delta t)] \\ &= \phi_H^{(\ell,k)}(0, \Delta t)e^{-j2\pi f_1 \Delta t} \end{aligned} \quad (1.15)$$

The correlation  $K_{y_u}(t, \Delta t)$  allows to understand the behavior of the time varying channel in the instants  $t$  and  $t + \Delta t$ , when the tone  $e^{j2\pi f_1 t}$  is transmitted. When time separation  $\Delta t$  is such that  $\phi_H^{(\ell,k)}(0, \Delta t)$  is very near to the maximum value  $\phi_H^{(\ell,k)}(0, 0)$  for all  $\Delta t < (\Delta t)_c^{(\ell,k)}$ , the transmitted signal of duration less than  $(\Delta t)_c$  is not affected by the non linear distortion, or, in other words, the channel is not time selective (say, flat in  $t$ ). For such a reason,  $(\Delta t)_c^{(\ell,k)}$  is called *coherence time*.

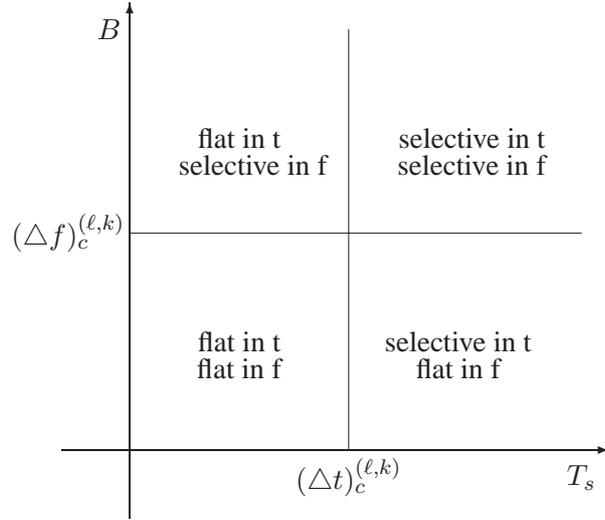
The equation (1.15) can be also studied in the frequency domain. To this aim, define that the Fourier transform of  $\phi_H^{(\ell,k)}(\Delta f, \Delta t)$  with respect to the variable  $\Delta t$ :

$$S_H^{(\ell,k)}(\Delta f, \lambda) \triangleq \int_{-\infty}^{\infty} \phi_H^{(\ell,k)}(\Delta f, \Delta t)e^{-j2\pi \lambda \Delta t} d\Delta t \quad (1.16)$$

The correlation function  $S_H^{(\ell,k)}(\Delta f, \lambda)$  evaluated for  $\Delta f = 0$  is called the *Doppler power spectrum of the channel* since it is a power spectrum that gives the signal intensity as a function of the Doppler frequency  $\lambda$ . By applying the Fourier transform (with respect to the variable  $\Delta t$ ) to both the right hand side and the left hand side of the equation (1.15), one has:

$$\mathcal{F} \{K_{y_u}(t, \Delta t), \Delta t \rightarrow \lambda\} = S_H^{(\ell,k)}(0, \lambda - f_1) \quad (1.17)$$

i.e., the effect of the time variations of the channel result in frequency spread of the received signal. The range of values over which  $S_H^{(\ell,k)}(0, \lambda)$  is sensitively nonzero is called *Doppler spread* and is denoted with  $B_d$ ; being  $S_H^{(\ell,k)}(0, \lambda)$



**Figure 1.2:** Channel classification:  $T_s$  is the symbol period,  $B$  the bandwidth.

the Fourier transform of  $\phi_H^{(\ell,k)}(0, \Delta t)$ ,  $B_d$  can be approximated with the inverse of  $(\Delta t)_c$ .

A complete qualitative classification of the radio channel based on the *coherence bandwidth* and the *coherence time* is reported in Fig 1.2.

### 1.3 Time-dispersive FIR MIMO channel

The characterization of the fading channel reported in the previous section allows one to determine the nature of the communication channel by simply comparing the symbol period  $T_s$  and the bandwidth  $B$  of the transmitted signals with the *coherence time* and the *coherence bandwidth* of the radio channel. In the modern digital communication systems, and especially for mobile applications, the increasing demand of high bit-rate transmission and the need to use stationary channel models require the adoption of a very short symbol period. For such a reason, the condition of flat fading in  $t$  channel is usually satisfied: specifically,  $T_s \lll (\Delta t)_c^{(\ell,k)} \forall \ell, k$ , and, consequently, the  $N_i N_o$  impulse responses of the composite MIMO channel in (1.4) do not introduce non linear

distortion on the transmitted signals. On the other hand, being  $(\Delta f)_c^{(\ell,k)} < B$ , the communication channels  $h^{(\ell,k)}(t, \tau)$  result to be frequency selective or, equivalently, time-dispersive.

According to the above considerations, this thesis considers the transmissions over frequency-selective channel assumed to be invariant within the observation interval. It follows that the output  $y_u(t)$  of the channel  $h^{(\ell,k)}(t, \tau)$  corresponding to the input signal  $u(t)$  can be simplified as follows [see also (1.9)]:

$$\begin{aligned} y_u(t) &= \int_{-\infty}^{\infty} h^{(\ell,k)}(t, \tau) u(t - \tau) t \tau \, d\tau \\ &= \int_{-\infty}^{\infty} h^{(\ell,k)}(\tau) u(t - \tau) t \tau \, d\tau \end{aligned} \quad (1.18)$$

where  $h^{(\ell,k)}(\tau) \triangleq h^{(\ell,k)}(0, \tau)$  is the time-invariant channel impulse response. Accounting for (1.18), the MIMO channel input-output relationship in (1.1) and the corresponding  $T_c$ -space sampled version in (1.7) can be rewritten as follows:

$$y^{(\ell)}(t) = \sum_{n=-\infty}^{\infty} \sum_{i=1}^{N_i} s_k^{(i)} h^{(\ell,i)}(t - nT_s) + n^{(\ell)}(t) \quad (1.19)$$

$$y_k^{(\ell)} = \sum_{n=-\infty}^{\infty} \sum_{i=1}^{N_i} x_{k-n}^{(i)} h_n^{(\ell,i)} + n_k^{(\ell)} \quad , \quad (1.20)$$

respectively, where  $h_n^{(\ell,i)} \triangleq h_{0,n}^{(\ell,i)} = h^{(\ell,i)}(0, nT_c)$ .

It is clear that the total number of  $T_c$ -space discrete-time channel coefficients  $h_n^{(\ell,i)}$  is determined by the maximum delay spread of the physical fading channel  $g^{(\ell,i)}(t, \tau)$  and the time durations of the transmit and receiver filters, which are usually infinite in theory to maintain limited frequency bandwidth. Therefore,  $h_n^{(\ell,i)}$  is a time-invariant filter with infinite impulse response (IIR). However, in practice, the time domain tails of the transmit and receive filters are designed to fall off rapidly, and it is reasonable to assume a finite range for the values of  $n$  over which the amplitude of the channel coefficients  $h_n^{(\ell,i)}$  is essentially nonzero. Thus, by eliminating the coefficients which do not affect significantly (owing to their small power) the channel output, the channel impulse response  $h_n^{(\ell,i)}$  can be truncated to a finite impulse response (FIR).

For the above reason, without loss of generality, we assume:

$$\begin{aligned} h_n^{(\ell,i)} &\neq 0 && \text{if } n = 0, \dots, \nu^{(\ell,i)} \\ h_n^{(\ell,i)} &= 0 && \text{otherwise} \quad , \end{aligned}$$

with  $\nu^{(\ell,i)}$  denoting the memory of the subchannel  $(\ell, i)$ . Therefore, the input-output relationship of the MIMO channel in the discrete time domain can be written as follows:

$$y_k^{(\ell)} = \sum_{n=0}^{\nu^{(\ell,i)}} \sum_{i=1}^{N_i} x_{k-n}^{(i)} h_n^{(\ell,i)} + n_k^{(\ell)} \quad 1 \leq \ell \leq N_o \quad . \quad (1.21)$$

By grouping the received samples from all  $N_o$  channel outputs at the  $k$ th instant into the  $N_o \times 1$  column vector

$$\mathbf{y}_k \triangleq \begin{bmatrix} y_k^{(1)} & y_k^{(2)} & \dots & y_k^{(N_o)} \end{bmatrix}^T \quad , \quad (1.22)$$

one can relate  $\mathbf{y}_k$  to the corresponding input and noise column vectors

$$\mathbf{x}_k \triangleq \begin{bmatrix} x_k^{(1)} & x_k^{(2)} & \dots & x_k^{(N_i)} \end{bmatrix}^T \quad (1.23)$$

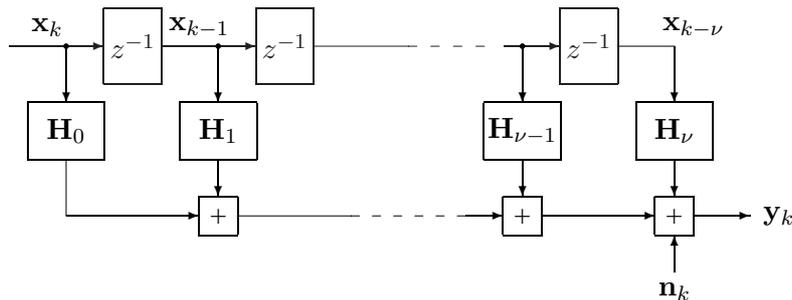
$$\mathbf{n}_k \triangleq \begin{bmatrix} n_k^{(1)} & n_k^{(2)} & \dots & n_k^{(N_o)} \end{bmatrix}^T \quad (1.24)$$

as follows:

$$\begin{aligned} \mathbf{y}_k &= \sum_{n=0}^{\nu} \mathbf{H}_n \mathbf{x}_{k-n} + \mathbf{n}_k \\ &= \mathbf{H}_k \star \mathbf{x}_k + \mathbf{n}_k \end{aligned} \quad (1.25)$$

where  $\mathbf{H}_n$  is the  $N_o \times N_i$  matrix whose  $(\ell, i)$  entry is  $h_n^{(\ell,i)}$ ,  $\nu$  is the maximum length of the  $N_o N_i$  channel impulse responses, i.e.,  $\nu = \max_{(\ell,i)} \nu^{(\ell,i)}$ , and  $\star$  denotes the discrete-time convolution operator. The block diagram of the discrete-time MIMO channel model is depicted in Fig. 1.3, where  $z^{-1}$  denotes the unit-delay block (i.e., the system that respond to the input  $\mathbf{x}_k$  with the output  $\mathbf{x}_{k-1}$ ).

In many applications, the channel matrix entries  $h_n^{(\ell,i)}$  can be modeled as zero-mean jointly WSS complex-valued circularly symmetric (see Chapter 2) Gaussian random sequences with the same variance. However, in real-word



**Figure 1.3:** Block diagram of the LTI time-dispersive MIMO channel of order  $\nu$ .

scenarios, the statistics of  $\mathbf{H}_k$  can vary owing to the spatial and temporal fading correlation, and/or to the presence of a fixed component (e.g., line of sight) in the channel resulting in Ricean fading, and/or to keyholes. These effects have been modeled in [5, 6, 7, 8] and have been shown to have a significant impact on the performance limits of MIMO channels. As regards the noise vector  $\mathbf{n}_k$ , its components  $n_k^{(\ell)}$  are usually modeled as zero-mean jointly WSS complex-valued circularly symmetric Gaussian random sequences uncorrelated with each other and independent of the transmitted signals. Such an assumption is realistic when the noise vector takes into account only for the presence of the thermal noise in the receiver. On the other hand,  $\mathbf{n}_k$  can describe the effects of the narrow-band interference due to overlay applications or cross-talk phenomena: in such cases, the components  $n_k^{(\ell)}$  can exhibit spatial and/or temporal correlation, and they can be correlated with the input signals. However, in the rest of the thesis, we assume the noise vector to be independent of the useful signals.

The time-dispersive MIMO channel model in (1.25), as it will be shown in next section, arises in many applications, including multi-antenna systems, spread-spectrum multiuser communications, multi-carrier systems. On the other hand, we should note that, although it well describes the linear distortion introduced by the radio channel, it does not take into account for other impairments such as the nonlinear distortion introduced by the A/D converters, and/or the phase and time jitter due to imperfect synchronization between the

transmit oscillators and the receive ones. Since the linear distortion represents the most important cause of performance degradation in many communication systems, this thesis will present some receiver and transmitter design methods mainly aimed at reducing the negative effects of the frequency selectivity of the communication channel.

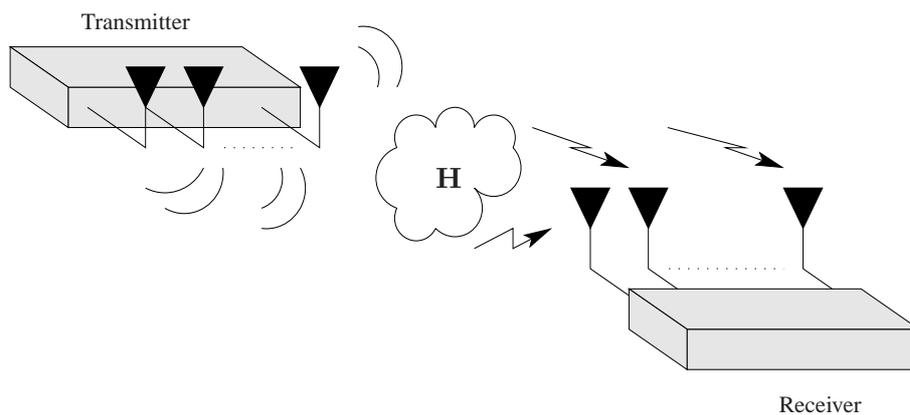
## 1.4 MIMO in communications

The model in (1.25) describes the input-output relationship of an abstract system with several inputs and/or outputs; in fact, its derivation has been carried out with no assumption regarding the specific application scenario. In the literature, the MIMO communication systems are usually identified with the ones that employ multiple antennas at both the transmitter and the receiver. However, we point out that, although it is not always explicit, a “virtual” MIMO model frequently arises in many communication systems. For example, in many applications, a more exact description of the detection scenario requires to take into account for the presence of several signals that, together with the transmitted one, affect the channel output: according to the conventional SISO schemes, the effects of such signals are modeled as additive noise. Otherwise, a MIMO model whose inputs include the undesired signals can be adopted to improve the system performances [9], provided that some a priori information about such undesired signals is available. It is clear that this model is not an actual MIMO one since the MIMO character is generated by an ad hoc representation of the transmitted signals.

For such reasons, in subsection 1.4.1, we introduce the multi-antenna systems by utilizing a simple propagation model, known in the literature as one-ring model: except when specified, in the rest of thesis, we will refer to the MIMO systems as multi-antenna systems. On the other hand, in subsections 1.4.2, 1.4.3, and 1.4.4, we refer to MIMO models which arise in direct-sequence code-division multiple access (DS-CDMA) systems, OFDM systems, and fractionally-spaced SISO systems.

### 1.4.1 Multiple antenna systems

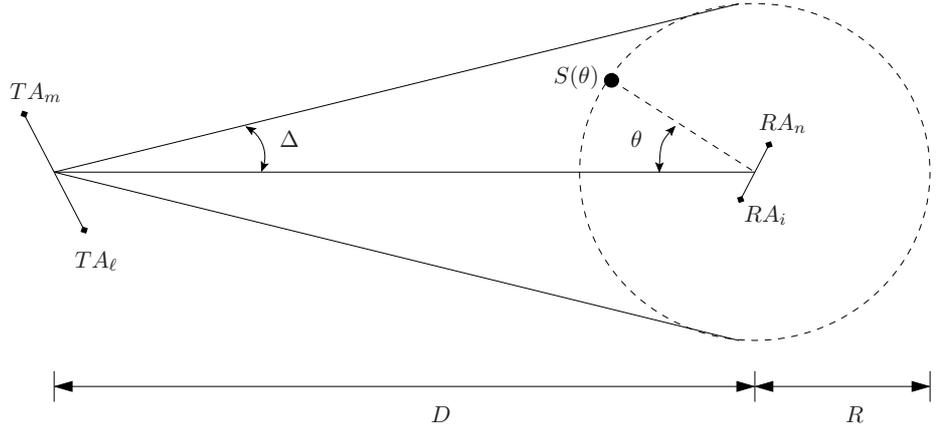
Multiple antenna systems (see Fig. 1.4) have drawn a considerable attention in the last years for their capability to reject interference and to reduce the effect of fading and noise. The earliest form of antenna system for improving the performance of the communication systems was antenna diversity, which



**Figure 1.4:** Multiantenna system.

mitigates the effect of fading. More recently, smart antenna systems, which attempt to actively mitigate the channel impairments, have been developed [10]. The communication systems that employ multiple antennas at the transmitter and at the receiver are popularly known as MIMO systems. Their structure naturally leads to a MIMO model where the signals transmitted by the transmit antennas represent the channel input, while the ones received by the receive antennas represent the channel output. In this section, we derive the channel impulse response of a multiple antenna system through the simple physical scattering model known in literature as *one-ring* model [11], which is appropriate in the fixed communication context, where the transmitter is elevated and seldom obstructed. For simplicity, in the following, we don't account for the time-varying nature of the channel.

Fig. 1.5 shows the *one-ring* model. The transmitting antennas and the receiving ones are denoted with  $TA_i$  and  $RA_\ell$ , and no LOS is present. According to such a model, every actual scatterer that lies at an angle  $\theta$  to the receiver is represented by a corresponding effective scatterer located at the same angle on the ring. In other words, we assume that the receiving antennas are surrounded by an infinite number of local scatterers uniformly distributed over the ring with radius  $R$ , which is usually assumed to be much smaller than the distance  $D$ , denoting the distance between the transmitting array and the receiving one. Each effective scatterer is denoted with  $S(\theta)$ : the rays that are reflected by  $S(\theta)$  are subject to a phase change  $\phi(\theta)$ , which accounts for the dielectric properties of the ring. Moreover, each scatterer is further assumed



**Figure 1.5:** Illustration of the abstract *one ring* model.

to be reflected only once, and all scatterers that reach the receiving antenna are equal in power<sup>4</sup>. According to the above assumptions, and in presence of  $K$  scatterers, the normalized path gain  $G_{\ell,i}$  of the subchannel connecting  $RA_i$  and  $TA_\ell$  is [6]:

$$g^{(\ell,i)}(t) = \frac{1}{\sqrt{2\pi K}} \sum_{k=1}^K e^{j\phi(\theta_k)} \delta \left( t - \frac{D_{TA_\ell \rightarrow S(\theta_k)} + D_{S(\theta_k) \rightarrow RA_i}}{c} \right) \quad (1.26)$$

where  $D_{a \rightarrow b}$  denotes the distance between  $a$  and  $b$ , and  $c$  is the speed of propagation of the electromagnetic field in the medium<sup>5</sup>.

Accounting for the time-invariant unit-energy impulse responses  $\psi_T(t)$  and  $\psi_R(t)$  of the transmit filter and the receive one, respectively, it is easy verified that

$$h^{(\ell,i)}(t) = \psi_T(t) * g^{(\ell,i)}(t) * \psi_R(t) \quad . \quad (1.27)$$

The model in (1.26) and (1.27) allows to simply relate the MIMO subchannel gains to the physical system parameters. Note that, when the reciprocal communication bandwidth is much smaller than delay spread, (1.26) can be approximated as follows

$$g^{(\ell,i)}(t) \simeq \alpha^{(\ell,i)} \delta(t) \quad , \quad (1.28)$$

<sup>4</sup>Such an assumption is equivalent to assume the antenna gains independent of the impinging direction of the plane wave travelling from the scatterer to the receive antenna.

<sup>5</sup>Note the similarity with the multipath model in (1.8).

i.e., a flat-frequency MIMO channel with complex-valued subchannel gains  $\alpha^{(\ell,i)}$  arises. With reference to such a channel model, it has been shown in [6] that when  $K$  is sufficient large, the central limit theorem can be applied and  $\alpha^{(\ell,i)}$  can be modeled as a zero-mean unitary-variance Gaussian random variable: in a way, this motivates the frequent assumption of Gaussian distributed channel coefficients  $h_k^{(\ell,i)}$ . Moreover, accounting for the condition  $D \gg R$ , it is shown that the correlation between  $\alpha^{(\ell,i)}$  and  $\alpha^{(m,n)}$  varies as

$$E \left[ \alpha^{(\ell,i)} \left( \alpha^{(m,n)} \right)^* \right] \sim J_0 \left( \frac{2\pi}{\lambda} d^R(m, n) \right) \quad (1.29)$$

where  $J_0(x) \triangleq (1/2\pi) \int_0^{2\pi} \exp(x \cos(\theta)) d\theta$ , and  $d^R(m, n)$  denotes the distance between the two receiving antennas. If the antenna spacing  $d^R(m, n)$  is sufficiently larger than half wavelength, the fades associated with two different receiving antennas can be considered independent of each other.

### 1.4.2 DS-CDMA systems

The code division multiple access (CDMA) systems have taken on a significant role in communications since they allow all users to use all the available time and frequency resources simultaneously, by assigning a code to each user [12, 13]. In the following, we show that a MIMO system model can be utilized to describe the CDMA system in the general case where the users are asynchronous, and the channel exhibits multipath distortion effects.

In direct sequence (DS) CDMA systems with  $K$  users, the transmitted signal by the  $k$ th user can be written in the following general form:

$$x^{(k)}(t) = \sum_{i=0}^{M-1} b_i^{(k)} s^{(k)}(t - iT) \quad k = 1, 2, \dots, K \quad (1.30)$$

with

$$s^{(k)}(t) \triangleq \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} c_j^{(k)} \varphi(t - jT_c) \quad 0 \leq t < T \quad (1.31)$$

where  $\left\{ b_i^{(k)} \right\}_{i=0}^{M-1}$  denotes the symbol stream to be transmitted by the  $k$ th user,  $T$  is the symbol interval,  $\varphi(\cdot)$  denotes the chip waveform of duration  $T_c = T/N$  and with unit energy, i.e.,  $\int_0^{T_c} \varphi^2(t) dt = 1$ , and, finally,  $\left\{ c_j^{(k)} \right\}_{j=0}^{N-1}$

is the signature sequence assigned to the  $k$ th user. Each signal is transmitted over a time invariant multipath channel whose impulse response is

$$g^{(k)}(t) = \sum_{m=1}^L \alpha_m^{(k)} \delta(t - \tau_m^{(k)}) \quad (1.32)$$

where  $L$  denotes the number of paths,  $\alpha_m^{(k)}$  the complex-valued path gain, and  $\tau_m^{(k)}$  the path delay such that  $\tau_1^{(k)} \leq \tau_2^{(k)} \leq \dots \leq \tau_L^{(k)}$ . At any given receiver, the received continuous-time signal is represented by the sum of the  $K$  channel outputs and the additive noise:

$$y(t) = \sum_{k=1}^K \sum_{i=0}^{M-1} b_i^{(k)} \sum_{m=1}^L \alpha_m^{(k)} s^{(k)}(t - iT - \tau_m^{(k)}) + v(t) \quad (1.33)$$

The received signal  $y(t)$  is processed by the chip matched filter and sampled at the chip-rate: the  $q$ th signal sample during the  $\ell$ th symbol interval is equal to:

$$\begin{aligned} y_\ell^{(q)} &\triangleq \int_{\ell T + qT_c}^{\ell T + (q+1)T_c} y(t) \varphi(t - \ell T - qT_c) dt \quad (1.34) \\ &= \sum_{k=1}^K \sum_{i=0}^{M-1} b_i^{(k)} \\ &\quad \cdot \underbrace{\sum_{m=1}^L \alpha_m^{(k)} \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} c_j^{(k)} \int_0^{T_c} \varphi(z) \varphi(z + ((\ell - i)N + q - j)T_c - \tau_m^{(k)}) dz}_{\triangleq h_{pN+q}^{(k)}} \\ &\quad + n_\ell^{(q)} \\ &= \sum_{k=1}^K \sum_{p=0}^{\xi} b_{\ell-p}^{(k)} h_{pN+q}^{(k)} + n_\ell^{(q)} \end{aligned}$$

where

$$n_\ell^{(q)} \triangleq \int_{\ell T + qT_c}^{\ell T + (q+1)T_c} n(t) \varphi(t - \ell T - qT_c) dt \quad ,$$

and  $\xi \triangleq \max_{1 \leq k \leq K} \left( \left\lceil \frac{\tau_L^{(k)} + T_c}{T} \right\rceil \right)$  (with  $\lceil \cdot \rceil$  denoting the smallest integer part not less than its argument) denotes the maximum delay spread normalized to

$T$ . By collecting the  $N$  samples of the received signal during the  $\ell$ th symbol interval into the column vector

$$\mathbf{y}_\ell \triangleq \left[ y_\ell^{(0)} \quad y_\ell^{(1)} \quad \dots \quad y_\ell^{(N-1)} \right]^T, \quad (1.35)$$

the input-output relationship (1.34) can be rewritten in the matrix form:

$$\mathbf{y}_\ell = \mathbf{H}_\ell \star \mathbf{b}_\ell + \mathbf{n}_\ell \quad (1.36)$$

$$\mathbf{H}_\ell \triangleq \begin{bmatrix} h_{\ell N}^{(1)} & \dots & h_{\ell N}^{(K)} \\ \vdots & \ddots & \vdots \\ h_{\ell N+N-1}^{(1)} & \dots & h_{\ell N+N-1}^{(K)} \end{bmatrix} \quad \mathbf{b}_\ell \triangleq \begin{bmatrix} b_\ell^{(1)} \\ \vdots \\ b_\ell^{(K)} \end{bmatrix} \quad \mathbf{n}_\ell \triangleq \begin{bmatrix} n_\ell^{(1)} \\ \vdots \\ n_\ell^{(K)} \end{bmatrix}.$$

According to (1.36), the asynchronous DS CDMA system is equivalent to a time-dispersive MIMO system, with impulse response  $\mathbf{H}_\ell$ . Let us note that, while the multi-user scenario naturally leads to a multiple-input system, the multiple-output character of the DS CDMA channel is generated by the special transformation of the single received signal.

Finally, it is worthwhile to observe that also the advanced communication systems combining DS CDMA techniques with multicarrier (MC) transmission schemes (see for example [14]) can be described by utilizing an augmented MIMO channel model properly defined.

### 1.4.3 OFDM systems

Orthogonal frequency-division multiplexing (OFDM) is a digital multi-carrier transmission technique that distributes the digitally encoded symbols over several parallel carriers in order to reduce the symbol rate and to achieve robustness against long echoes in a multipath radio channel. Unlike conventional frequency-division multiplexing, the spectra of the OFDM carriers partially overlap. Nevertheless, they exhibit orthogonality on a symbol interval if they are spaced in frequency exactly at the reciprocal of the symbol interval. Such a requirement can be fulfilled by using the discrete Fourier transform, and by introducing a guard interval equal or greater than the delay spread of the channel. In this section, we provide a description of the OFDM communication systems in terms of MIMO systems.

In an OFDM system,  $N$  input symbols (say, OFDM word) are transferred by the serial-to-parallel converter (S/P) to the OFDM modulator. After each symbol is modulated by the corresponding subcarrier, it is sampled and  $D/A$

converted. The discrete-time OFDM signal, implemented by an inverse discrete Fourier transform (IDFT), can be expressed as follows:

$$x_{kN+p} = \frac{1}{N} \sum_{m=0}^{N-1} s_{kN+m} e^{j\frac{2\pi pm}{N}} \quad 0 \leq p \leq N \quad (1.37)$$

$$-\infty \leq k \leq \infty$$

where  $s_{kN+m}$  ( $m = 0, \dots, N-1$ ) denotes the OFDM word to be transmitted,  $x_{kN+p}$  represents the  $(kN+p)$ th output sample of the output of the IDFT block. After pulse shaping and parallel-to-serial (P/S) conversion, the signal is transmitted over a SISO time-variant multipath fading channel that consists of  $L$  propagation paths with complex-valued channel gains  $h_{k,\ell}$  (the apex  $(1,1)$ , which accounts for the  $1 \times 1$  channel, is omitted for clarity). At receiver hand, after matched filtering and removing the cyclic prefix (see for instance [15]), the received signal can be written as

$$y_{kN+p} = \sum_{\ell=0}^{L-1} h_{kN+p,\ell} x_{kN+p-\ell} + n_{kN+p} \quad , \quad (1.38)$$

where  $n_{kN+p}$  denotes the additive noise samples. After S/P conversion of the received samples, the demodulated signal  $Y_k^{(p)}$  is obtained by taking the discrete Fourier transform (DFT) of the vector  $[y_{kN} \dots y_{kN+N-1}]^T$ , i.e.:

$$\begin{aligned} Y_k^{(p)} &= \frac{1}{N} \sum_{i=0}^{N-1} \sum_{\ell=0}^{L-1} \sum_{m=0}^{N-1} s_{kN+m} h_{kN+i,\ell} e^{j\frac{2\pi(i-\ell)m}{N}} e^{-j\frac{2\pi pi}{N}} + \underbrace{\sum_{i=0}^{N-1} n_{kN+i} e^{-j\frac{2\pi pi}{N}}}_{\triangleq N_k^{(p)}} \\ &= \sum_{\ell=0}^{L-1} \sum_{m=0}^{N-1} s_{kN+m} \underbrace{\frac{1}{N} \sum_{i=0}^{N-1} h_{kN+i,\ell} e^{-j\frac{2\pi(p-m)i}{N}} e^{-j\frac{2\pi \ell m}{N}}}_{\triangleq H_k^{(p-m,\ell)}} + N_k^{(p)} \\ &= \sum_{\ell=0}^{L-1} \sum_{m=0}^{N-1} H_k^{(p-m,\ell)} s_{kN+m} e^{-j\frac{2\pi \ell m}{N}} + N_k^{(p)} \quad . \quad (1.39) \end{aligned}$$

The input-output relationship (1.39) can be rewritten in a matrix form as follows:

$$\mathbf{Y}_k = \mathbf{C}_k \mathbf{s}_k + \mathbf{N}_k \quad (1.40)$$

with

$$\mathbf{s}_k \triangleq \begin{bmatrix} s_{kN} \\ s_{kN+1} \\ \vdots \\ s_{kN+(N-1)} \end{bmatrix} \quad \mathbf{Y}_k \triangleq \begin{bmatrix} Y_k^{(0)} \\ Y_k^{(1)} \\ \vdots \\ Y_k^{(N-1)} \end{bmatrix} \quad \mathbf{N}_k \triangleq \begin{bmatrix} N_k^{(0)} \\ N_k^{(1)} \\ \vdots \\ N_k^{(N-1)} \end{bmatrix}, \quad (1.41)$$

and where the  $(m, i)$  entry of the  $N \times N$  matrix  $\mathbf{C}_k$  is defined as

$$\begin{aligned} C_k^{(m,i)} \triangleq & H_k^{(0,m-i)} + H_k^{(1,m-i)} e^{-j\frac{2\pi i}{N}} + \dots \\ & \dots + H_k^{(L-1,m-i)} e^{-j\frac{2\pi(L-1)i}{N}}. \end{aligned} \quad (1.42)$$

It follows that the OFDM system is equivalent to a LTV non-dispersive MIMO system with impulse response  $\mathbf{C}_k$ .

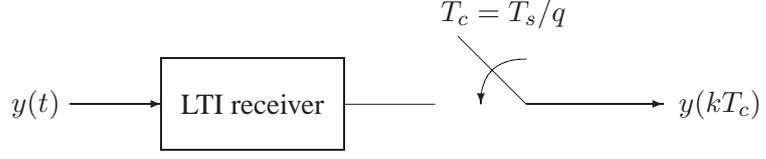
If the channel impulse response remains constant over the word interval, one has  $H_k^{(p-m,\ell)} = 0$  for  $p \neq m$ , implying that  $\mathbf{C}_k$  degenerates into a diagonal matrix, and there exists no intercarrier interference (ICI). In such a case, the received samples  $Y_k^{(p)}$  are affected by only the multiplicative distortion, which can be easily compensated for by a one-tap frequency-domain equalizer: in other words, the OFDM effectively converts a frequency selective fading channel into a set of  $N$  flat fading channels. On the other hand, the variations of the channel impulse response during the word interval, as well as the existence of a frequency offset<sup>6</sup>, destroy the orthogonality among the OFDM subcarriers leading to a non-diagonal matrix  $\mathbf{C}_k$ , which accounts for the presence of ICI.

Recently, multiple antenna solutions and OFDM modulation have been combined to obtain the MIMO-OFDM systems [15]. Also in this case, it is possible to show that the overall system equation can be represented by a MIMO model.

#### 1.4.4 Fractionally spaced sampling

Fractionally spaced sampling is frequently utilized in SISO systems to reduce the sensitivity of the receiver to synchronization errors, or simply to improve the detector performances. According to such a technique, the received signal is sampled  $q$  times during a symbol period  $T_s$  (say  $T_c = T_s/q$  the sampling period), as depicted in Figure 1.6, where an LTI receiver has been considered for simplicity.

<sup>6</sup>Frequency offset in communication systems are caused by the mismatches between the oscillator in the transmitter and in the receiver, by Doppler shifts, etc..



**Figure 1.6:** Fractionally spaced sampling leads to a stationary multiple output model.

Since the sampling rate is not equal to the symbol rate, the signal at the output of the matched filter is not cyclostationary rather that stationary: its moments vary periodically with a period equal to  $q$ . The  $T_c$ -sampled discrete-time input-output relationship can be obtained by utilizing (1.5) with  $N_i = N_o = 1$ . However, a great advantage in terms of detector implementation purposes could be obtained if the output of the matched filter is transformed into a stationary signal. To this aim, we collect  $q$  consecutive samples of the matched filter output in the vector

$$\mathbf{y}_k \triangleq [ y(kT_s + T_c) \quad y(kT_s + 2T_c) \quad \dots \quad y(kT_s + (q-1)T_c) ] \quad (1.43)$$

Each element  $y_k^{(\ell)}$  of  $\mathbf{y}_k$  can be seen as the  $T_s$ -sampled output of a LTI channel with channel impulse response defined as

$$h^{(\ell)}(t) \triangleq h(t + \ell T_c) \quad \ell = 0, \dots, q-1 \quad , \quad (1.44)$$

corrupted by the additive noise  $n(t)$ , and whose input is the transmitted signal  $\sum_{n=-\infty}^{\infty} s_k \delta(t - nT_s)$ , with i.i.d.  $s_k$ . According to the above considerations, we can express  $\mathbf{y}_k$  in (1.43) as the output of a SIMO system defined as follows:

$$\mathbf{y}_k = \mathbf{H}_k \star s_k + \mathbf{n}_k \quad (1.45)$$

where  $\mathbf{H}_k$  is a  $q \times 1$  vector whose  $\ell$ th element is  $h_k^{(\ell)} \triangleq h^{(\ell)}(kT_s)$ , and  $\mathbf{n}_k$  is obtained by stacking  $q$  noise  $T_c$ -space samples at the output of the matched filter. The system model in (1.45) provides stationary input and output sampled at the symbol rate.

For the sake of completeness, let us note that the adoption of the MIMO model, instead of the SISO one, allows one to obtain a stationary model also

in scenarios where the input signal  $s_k$  is cyclostationary (providing so a cyclostationary channel output). For example, the stationarization of pulse amplitude modulated (PAM) cyclostationary sequences transmitted over frequency selective fading channels has been considered in [16].

## 1.5 Capacity of MIMO systems

The MIMO system are here discussed from an information theoretic perspective: the concept of mutual information between the channel input and output gives a guideline to how well our system performs and how close it operates to the ultimate Shannon limit. With regards to such an aspect, recent research on MIMO channels, including the study of channel capacity [1, 10] and the design of communication schemes [17], demonstrates a great improvement of performance over the conventional SISO systems.

For the analysis in this section, we consider a single user frequency flat-fading channel with  $N_i$  inputs and  $N_o$  outputs. According to (1.7), the input-output relationship over a symbol period can be written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k \quad (1.46)$$

where the noise vector  $\mathbf{n}_k$  has i.i.d. complex-valued circularly symmetric (see Chapter 2) zero-mean Gaussian entries with unitary variance, i.e., its correlation matrix is equal to  $\mathbf{R}_n \triangleq [\mathbf{n}_k \mathbf{n}_k^H] = \mathbf{I}_{N_o}$ , with the apex  $H$  denoting the conjugate transpose, and being  $\mathbf{I}_K$  the identity matrix of size  $K$ . The total transmitted average power over a symbol period is constrained to be less or equal to  $\mathcal{P}_0$  by assuming that the input correlation matrix of  $\mathbf{x}_k$ ,  $\mathbf{R}_x \triangleq E[\mathbf{x}_k \mathbf{x}_k^H]$ , satisfies  $\text{trace}(\mathbf{R}_x) = \mathcal{P}_0$ . Let us distinguish two possible scenarios:

- $\mathbf{H}_k = \mathbf{H}$  is deterministic;
- $\mathbf{H}_k$  is modeled as a random matrix with zero-mean circularly symmetric Gaussian entries with unit variance and independent of each other.

In both scenarios, we assume that the channel state information (CSI) is available at the receiver. For convenience, in the following subsections, the subscript  $k$  in (1.46) is omitted, if not stated otherwise.

### 1.5.1 Capacity of deterministic MIMO channel

Assume that the transmitter has knowledge of the channel matrix  $\mathbf{H}$ . The channel capacity (measured in nats/s/Hz) is obtained by maximizing the mutual information between the channel input and output over the input correlation matrices satisfying the power constraint, i.e.<sup>7</sup>:

$$\begin{aligned} \mathcal{C} &= \max_{\mathbf{R}_x, \text{trace}(\mathbf{R}_x) \leq \mathcal{P}_0} \left\{ \log \det (\mathbf{I}_{N_o} + \mathbf{H}\mathbf{R}_x\mathbf{H}^H) \right\} \\ &= \max_{\mathbf{R}_x, \text{trace}(\mathbf{R}_x) \leq \mathcal{P}_0} \left\{ \log \det (\mathbf{I}_{N_i} + \mathbf{R}_x\mathbf{H}^H\mathbf{H}) \right\} . \end{aligned} \quad (1.47)$$

By defining the eigenvalue decomposition (EVD) of the matrix  $\mathbf{H}^H\mathbf{H}$  as

$$\mathbf{H}^H\mathbf{H} \triangleq \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H , \quad (1.48)$$

it can be shown [18] that the input vector maximizing the information rate is a circularly symmetric Gaussian vector with correlation matrix

$$\mathbf{R}_x = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^H \quad (1.49)$$

where the diagonal entries  $\sigma_i$  ( $i = 1, \dots, N_i$ ) of the diagonal matrix  $\mathbf{\Sigma}$  are provided by the well known water-filling procedure, i.e.,

$$\sigma_i = (\mu - \lambda_i^{-1})_+ \quad \text{such that} \quad \sum_i \sigma_i \leq \mathcal{P}_0 , \quad (1.50)$$

with  $\lambda_i$  ( $i = 1, \dots, N_i$ ) denoting the diagonal entries  $\mathbf{\Lambda}$ , and  $(a)_+ \triangleq \max\{0, a\}$ . Thus, the channel capacity can be parametrized as

$$\mathcal{C}(\mu) = \sum_i \log(\mu\lambda_i) . \quad (1.51)$$

It is important to note that, being the nonzero eigenvalues of  $\mathbf{H}^H\mathbf{H}$  equal to those of  $\mathbf{H}\mathbf{H}^H$ , the channels  $\mathbf{H}$  and  $\mathbf{H}^H$  achieves the same capacity (reciprocity).

When CSI is available at the transmitter, the concept of forming an average capacity is somewhat less straightforward, since the transmitter has the additional option of optimizing the power allocation over time as well as over the

<sup>7</sup>Accounting also for the independence between the useful signals and the noise ones, and accounting for  $\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})$ .

eigenvalues (while maintaining the required average power restriction). For such a reason, we shall not consider average capacities for this case.

In some application scenarios, it is reasonable to restrain the computational complexity of the transmitter by allocating the same power to the  $N_i$  channel inputs, i.e.,  $\mathbf{R}_x = (\mathcal{P}_0/N_i)\mathbf{I}_{N_i}$ . In such a case, the achieved capacity is equal to [1]

$$\mathcal{C}_{EP} = \sum_{k=1}^m \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda_k \right) , \quad (1.52)$$

where  $m = \min\{N_i, N_o\}$ , and the subscript  $EP$  stays for equal-power. The above equation expresses the MIMO channel capacity as the sum of the capacities of  $m$  SISO channels with channel gain  $\sqrt{\lambda_k}$ . It follows that multiple *spatial modes* (or *eigen subchannels* or *eigenmodes*) open up between transmitter and receiver resulting in a performance improvement with respect to the conventional SISO system.

### 1.5.2 Ergodic capacity of fading MIMO channel

Let us now consider the more interesting case of random channel matrix, independent of the input and noise vectors. We make the following assumption:

- no CSI is available at the transmitter;
- the distribution of  $\mathbf{H}$  is known at the transmitter;
- the transmission of the symbol-vector  $\mathbf{x}$  corresponds to one use of the channel, and each use of the channel corresponds to an independent realization of  $\mathbf{H}$  (according to the chosen probability distribution); this is known as the *ergodic assumption*.

The channel capacity (called *ergodic capacity*) is obtained by maximizing over the input correlation matrix  $\mathbf{R}_x$ , the mutual information between the channel input and output averaged over  $\mathbf{H}$ ; in symbols, one has:

$$\mathcal{C} = \max_{\mathbf{R}_x, \text{trace}(\mathbf{R}_x) \leq \mathcal{P}_0} \left\{ E_{\mathbf{H}_k} \left[ \log \det \left( \mathbf{I}_{N_o} + \mathbf{H} \mathbf{R}_x \mathbf{H}^H \right) \right] \right\} , \quad (1.53)$$

where  $E_{\mathbf{H}}[\cdot]$  denotes the expectation with respect to  $\mathbf{H}$ . In [1], it has been shown that the capacity is achieved for  $\mathbf{x}$  circularly symmetric zero-mean

complex-valued Gaussian vector with  $\mathbf{R}_x = (\mathcal{P}_0/N_i)\mathbf{I}_{N_i}$ . Accounting for (1.52), it follows:

$$\mathcal{C} = E_{\Lambda} \left[ \sum_{k=1}^m \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda_k \right) \right] . \quad (1.54)$$

The computation of the expectation (1.54) needs the knowledge of the joint density of the unordered strictly positive eigenvalues  $\lambda_k$  of the random non-negative definite Wishart matrix  $\mathbf{H}^H \mathbf{H}$  (or, equivalently,  $\mathbf{H} \mathbf{H}^H$ ). Such a density is known in the literature to be [19]

$$p_{\Lambda}(\lambda_1, \dots, \lambda_m) = \frac{1}{m!} \prod_{i=1}^n \frac{\lambda_i^{n-m}}{(n-i)!(m-i)!} e^{-\sum_{i=1}^m \lambda_i} \prod_{i < k} (\lambda_i - \lambda_k)^2 \quad (1.55)$$

and, by accounting for its symmetry, one has

$$\mathcal{C} = m E_{\lambda_k} \left[ \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda_k \right) \right] , \quad (1.56)$$

where  $\lambda_k$  is one of the unordered eigenvalues. Thus, the capacity of the MIMO system is given by the following theorem [1].

**Theorem 1.1** *The capacity of the fading channel  $\mathbf{H}$  with  $N_i$  input and  $N_o$  outputs, subject to the power constraint  $\text{trace}(\mathbf{R}_x) = \mathcal{P}_0$ , is equal to*

$$\mathcal{C} = m \int_0^{\infty} \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda \right) p_{\lambda}(\lambda) d\lambda , \quad (1.57)$$

where

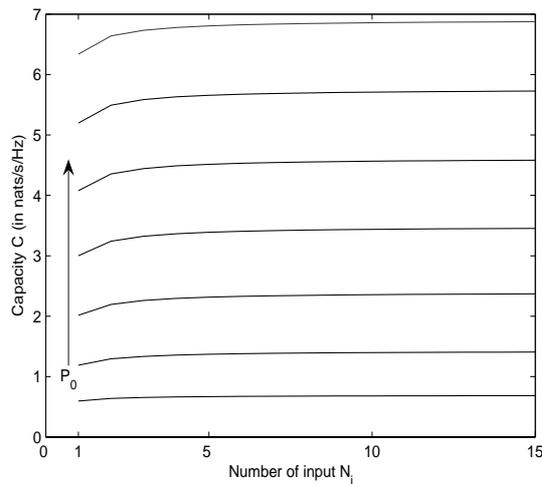
$$p_{\lambda}(\lambda) = \frac{1}{m} \sum_{k=0}^{m-1} \frac{k!}{(k+n-m)!} [L_k^{n-m}(\lambda)]^2 \lambda^{n-m} e^{-\lambda} \quad (1.58)$$

where  $n = \max\{N_i, N_o\}$ , and where  $L_k^{\alpha}(\lambda) = (1/k!) e^{\lambda} \lambda^{-\alpha} \frac{d^k}{d\lambda^k} (e^{-\lambda} \lambda^{\alpha+k})$  is the generalized Laguerre polynomial of order  $k$ .

The expression (1.57) can be specialized to both the cases of MISO and SIMO channels as follows:

$$\mathcal{C}_{MISO} = \frac{1}{(N_i - 1)!} \int_0^{\infty} \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda \right) \lambda^{N_i-1} e^{-\lambda} d\lambda \quad (1.59)$$

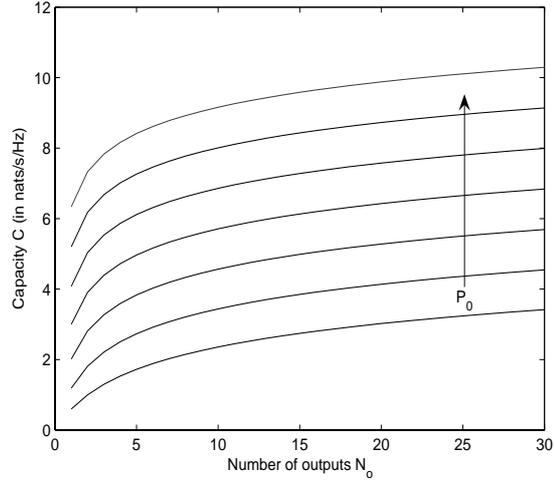
$$\mathcal{C}_{SIMO} = \frac{1}{(N_o - 1)!} \int_0^{\infty} \log(1 + \mathcal{P}_0 \lambda) \lambda^{N_o-1} e^{-\lambda} d\lambda . \quad (1.60)$$



**Figure 1.7:** Capacity of a MISO ( $N_o = 1$ ) system versus the number of inputs  $N_i$  for different values of  $\mathcal{P}_0$ .

By numerically computing the integrals in (1.59) and (1.60), we have plotted in Fig. 1.7 and 1.8 the capacity achieved by the MISO and the SIMO systems, respectively, for different values of the available power ( $\mathcal{P}_0 = 0\text{dB}, 5\text{dB}, 10\text{dB}, 15\text{dB}, 20\text{dB}, 25\text{dB}, 30\text{dB}$ ). As expected, they outperform the conventional SISO systems. However, while  $\mathcal{C}_{SIMO}$  is not bounded as  $N_o$  gets large,  $\mathcal{C}_{MISO}$  rapidly reaches saturation as  $N_i$  increases, i.e., increasing the number of the MISO system inputs provides no further capacity gain. In other words, the reciprocity observed for deterministic  $\mathbf{H}$  does not hold for random  $\mathbf{H}$ . Such a difference can be easily explained by observing that, when CSI is not available at the transmitter, the power allocated to each input  $\mathcal{P}_0/N_i$  decreases as  $N_i$  gets large. Thus, the gain provided by the use of multiple inputs<sup>8</sup> can not be exploited due to the decrease of power (per input). In [20], upper and lower bounds to (1.57) have been provided to show that, for large

<sup>8</sup>Note that the nonzero eigenvalues of the Wishart matrices  $\mathbf{H}^H \mathbf{H}$  and  $\mathbf{H} \mathbf{H}^H$  are equal, as it happen for deterministic  $\mathbf{H}$ .



**Figure 1.8:** Capacity of a SIMO ( $N_i = 1$ ) system versus the number of outputs  $N_o$  for different values of  $\mathcal{P}_0$ .

$\mathcal{P}_0$ , one has

$$\text{finite } N_o, N_i \rightarrow \infty \Rightarrow \mathcal{C} \approx N_o \log(\mathcal{P}_0) \quad (1.61)$$

$$\text{finite } N_i, N_o \rightarrow \infty \Rightarrow \mathcal{C} \approx N_i \log\left(\frac{\mathcal{P}_0}{N_i}\right) + N_i \log(N_o) \quad , \quad (1.62)$$

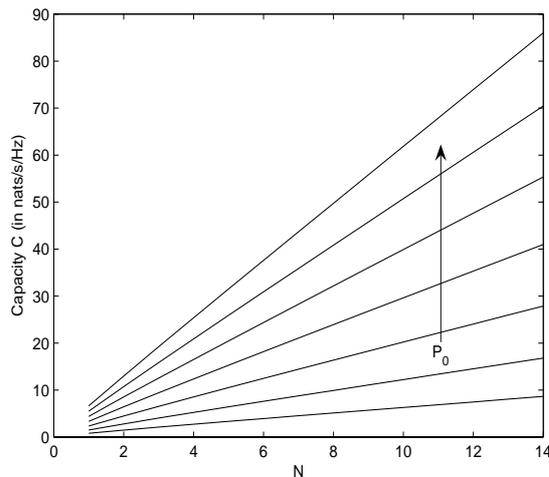
i.e., the system capacity for fixed  $N_i$  grows with  $\log(N_o)$ , while, for fixed  $N_o$ , it remains constant as  $N_i$  increases. This confirms the results reported in Fig. 1.7.

In Fig. 1.9, the capacity of a square ( $N_i = N_o$ ) MIMO system has been plotted versus  $N_i$ , for the different<sup>9</sup> values of  $\mathcal{P}_0$ , by numerically computing the integrals in (1.57). The analysis allows one to state that the system capacity of the square MIMO system is well approximated by a linear function of  $N_i$ . Bounding (1.57), in [20], it has been shown that:

$$\left[ N_i, N_o \rightarrow \infty, \frac{N_i}{N_o} \leq 1 \right] \Rightarrow \frac{1}{N_i} \mathcal{C} \rightarrow \log(\mathcal{P}_0) + \log\left(\frac{N_o}{N_i}\right) \quad (1.63)$$

$$\left[ N_i, N_o \rightarrow \infty, \frac{N_i}{N_o} > 1 \right] \Rightarrow \frac{1}{N_o} \mathcal{C} \rightarrow \log(\mathcal{P}_0) \quad , \quad (1.64)$$

<sup>9</sup>The ones previously considered in Fig. 1.7 and 1.8.



**Figure 1.9:** Capacity of a square MIMO ( $N_i = N_o = N$ ) system versus  $N$  for different values of  $\mathcal{P}_0$ .

i.e., the system capacity grows linearly with the minimum between the number of inputs and outputs  $m$ : this can be explained by observing that the capacity of the MIMO channel has been expressed as the sum of the capacities of  $m$  SISO channels. Moreover, from (1.63) it is straightforward verified that there is no benefit from increasing  $N_i$  beyond  $N_o$  in the asymptotic scenario.

### Asymptotic analysis

In the last few years, a considerable amount of work is available in the literature on the fundamental limits of communication channels that makes substantial use of the asymptotic results in random matrix theory. Here, we briefly recall the landmark contribution<sup>10</sup> of Marčenko and Pastur (1967) about the eigenvalue distribution of random matrices in the form  $\mathbf{H}^H \mathbf{H}$ , and we show its utility in evaluating the capacity in (1.57) of the MIMO system when both the numbers of inputs and outputs grow up to infinity, maintaining finite their ratio.

<sup>10</sup>Other landmark contributions to the theory of random matrices have been previously provided by Wishart (1928) and Wigner (1955).

To this aim, let us first rewrite (1.54) as

$$\begin{aligned} \mathcal{C} &= E_{\Lambda} \left[ \int_0^{\infty} \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda \right) \sum_{k=1}^m \delta(\lambda - \lambda_k) \right] d\lambda \\ &= E_{\Lambda} \left[ m \int_0^{\infty} \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda \right) dF_{\mathbf{H}^H \mathbf{H}}^m(\lambda) \right] \end{aligned} \quad (1.65)$$

where  $F_{\mathbf{H}^H \mathbf{H}}^m(\lambda)$  is the empirical cumulative distribution function and it is defined as

$$F_{\mathbf{H}^H \mathbf{H}}^m(\lambda) \triangleq \frac{1}{m} \sum_{k=1}^m 1\{\lambda_k \leq \lambda\} \quad , \quad (1.66)$$

with  $1\{\cdot\}$  being 1 if the argument is true, and 0 otherwise. Marčenko and Pastur have showed the following theorem [21]:

**Theorem 1.2** Consider an  $N_o \times N_i$  matrix  $\mathbf{H}$  whose entries are zero-mean i.i.d. complex-valued random variables with variance  $\frac{1}{N_o}$  and fourth moments of order  $\mathcal{O}(1/N_o^2)$ . As  $N_i, N_o \rightarrow \infty$  with  $\frac{N_i}{N_o} \rightarrow \beta$ , the empirical asymptotic density function  $dF_{\mathbf{H}^H \mathbf{H}}^m(\lambda)/d\lambda$  converges almost surely to the nonrandom function

$$f_{\beta}(\lambda) = \left(1 - \frac{1}{\beta}\right)_+ \delta(\lambda) + \frac{\sqrt{(\lambda - a)_+(b - \lambda)_+}}{2\pi\beta\lambda} \quad \lambda \in [a, b]$$

with

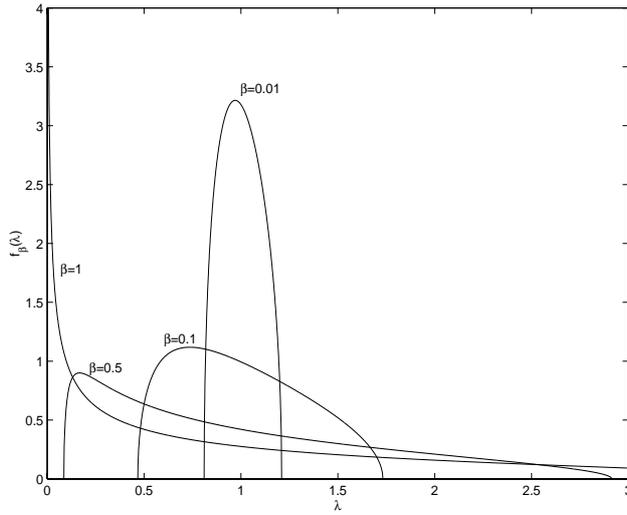
$$a = (1 - \sqrt{\beta})^2 \quad b = (1 + \sqrt{\beta})^2 \quad .$$

The density function  $f_{\beta}(\lambda)$  has been plotted in Fig. 1.10 for different values of  $\beta$ . Note that, when  $\beta > 1$ , the impulsive term in  $f_{\beta}(\lambda)$  accounts for the  $N_i - N_o > 0$  zero eigenvalues of  $\mathbf{H}^H \mathbf{H}$ . Analogously, the empirical asymptotic density function of  $\mathbf{H} \mathbf{H}^H$  converges almost surely to the nonrandom function

$$f'_{\beta}(\lambda) = (1 - \beta)_+ \delta(\lambda) + \frac{\sqrt{(\lambda - a)_+(b - \lambda)_+}}{2\pi\lambda} \quad \lambda \in [a, b] \quad .$$

Thus, according to the asymptotic scenario ( $N_i, N_o \rightarrow \infty$ ), the channel capacity  $\mathcal{C}$  can be written in the following form:

$$\mathcal{C} \rightarrow m \int_a^b \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda \right) f_{\beta}(\lambda) d\lambda \quad (1.67)$$



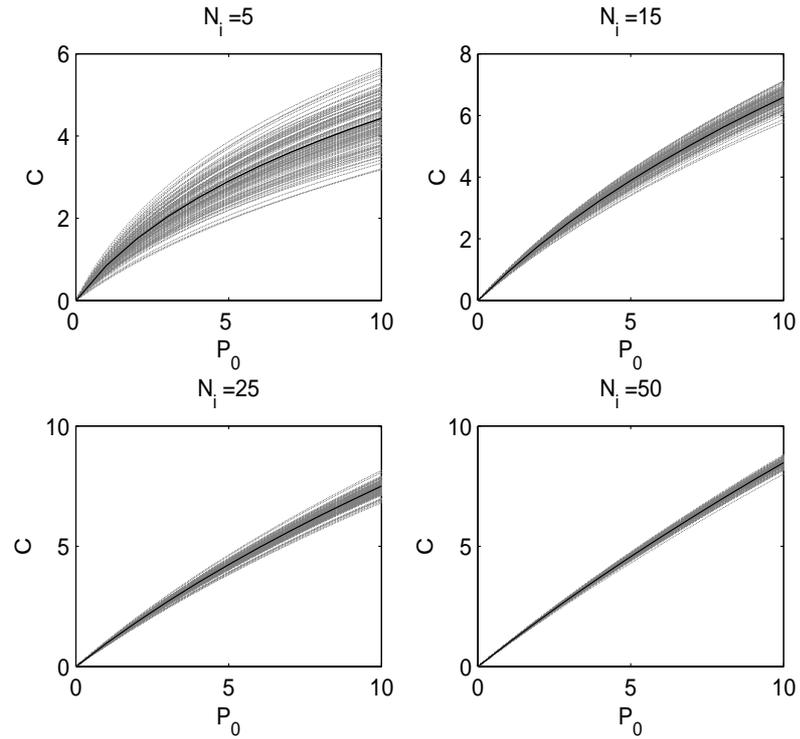
**Figure 1.10:** The Marčenko and Pastur of density of  $\mathbf{H}^H \mathbf{H}$  for  $\beta = 1, 0.5, 0.1, 0.01$ .

where the expectation operator in (1.65) has been removed since  $f_\beta(\lambda)$  is a deterministic function. By specializing such a result in the case of a square MIMO channel, one has [1, 10]

$$C \rightarrow N_o \int_0^4 \log \left( 1 + \frac{\mathcal{P}_0}{N_i} \lambda \right) \frac{1}{2\pi} \sqrt{\frac{4}{\lambda} - 1} d\lambda, \quad (1.68)$$

which confirms the results shown in Fig. 1.9, where the capacity of the MIMO channel grows in a linear fashion with  $N_o$ .

The widespread interest in the asymptotic results concerning the eigenvalue distribution of random matrices is due to its fast convergence to the asymptotic limit. If the convergence is so fast, then, even for small values of the parameters, the asymptotic results come close to the finite-size results. In Fig. 1.11 we have plotted the capacity (black line)  $C$  in (1.68) versus  $0 \leq \mathcal{P}_0 \leq 10$ , for  $N_i = 5, 15, 25, 50$ . The grey points represents the capacity achieved over each one of 100 realizations of the matrix  $\mathbf{H}$ , whose entries have been modeled as zero-mean complex-valued circularly symmetric Gaussian variables with variance  $1/N_o$ , and uncorrelated with each other. As  $N_o$  increases, the asymptotic limit in (1.68) well approximates the capacity of the single channel realization.



**Figure 1.11:** One hundred realizations of (1.54) compared to the asymptotic capacity in (1.68) versus  $N_i = N_o$  for different values of  $P_0$ .

The results concerning the asymptotic scenario will be utilized in Chapter 3 to compare the performances of different MIMO communication systems.

### 1.5.3 Outage capacity

The ergodic assumption considered in the previous subsection will not be satisfied in practical communication systems as, for example, in delay-constrained applications, where the channel matrix  $\mathbf{H}$  is still random, but it remains fixed once it is chosen for all the uses of the channel. In other words, the time interval needed to transmit the information symbols is comparable with the coherence time. In such a case, the Shannon capacity is zero due to the fact that there is always a nonzero probability that the given channel realization will not support

the rate at which we wish to communicate.

The system parameter to be considered is the *out%* outage capacity  $C_{out}$ , defined as the mutual information that is guaranteed for  $(100 - out)\%$  of the channel realization:

$$\text{Prob}(C_{EP} \leq C_{out}) = out\% \quad (1.69)$$

where  $C_{EP}$  has been defined in (1.52). The above definition is reasonable when the channel matrix  $\mathbf{H}$  is just a “snapshot” of the underlying stochastic process. Hence, there is a probability that this particular channel realization is in such a deep fade that the communication system operating with  $C_{out}$  nats/s/Hz will fail to transmit without errors. The zero outage capacity ( $out\% = 0$ ) can be interpreted as the lowest transmission rate that is invariant of the fading. Let us note that, unlike MIMO system, for a SISO one this corresponds to channel inversion, which then makes the observed channel independent of the fading. Since a SISO Rayleigh fading channel is not invertible when the power is finite, the SISO zero outage capacity is zero.

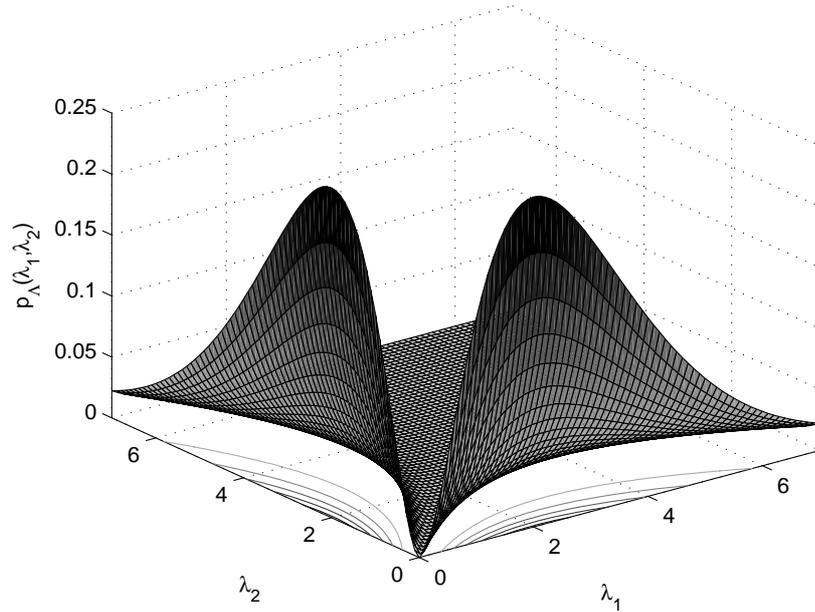
## 1.6 Multiplexing, diversity and array gain

In this section, the concepts of *multiplexing gain*, *diversity gain* and *array gain* are introduced by resorting to an intuitive discussion based on the main results (reported in previous section) about the eigenvalues distribution of random matrices.

Let us note that the channel capacity in (1.51) and in (1.52) is related to the eigenvalues  $\lambda_k$  of the random matrix  $\mathbf{H}^H\mathbf{H}$ . The square roots of such eigenvalues represent the gains of the parallel SISO channels (*spatial modes*) in which  $\mathbf{H}$  is decomposed, and, consequently, give an idea of how much the transmission is good over a certain subchannel.

In Fig. 1.12-1.14, the joint density (1.55) has been depicted for  $N_i = 2$  and  $N_o = 2, 6, 12$ . Note that, as the number of outputs increases, the centers of two lobes of  $p_\lambda(\lambda_1, \lambda_2)$  deviate from the axes  $\lambda_1 = 0$  and  $\lambda_2 = 0$ , i.e., it increases the probability of decomposing the MIMO channel into two *spatial modes* with likely different from zero gain. Such a behavior can be explained by the light of Theorem 1.2, from which it is evident that as  $\beta \rightarrow 0$ , the eigenvalue distribution of  $\mathbf{H}^H\mathbf{H}$  approaches to an impulsive function.

The use of two antennas at the transmitter and  $N_o \geq 2$  antennas at the receiver, in conjunction with rich scattering in the propagation environment, opens up two ( $\min\{N_i, N_o\}$ ) for a  $N_o \times N_i$  MIMO channel, according to

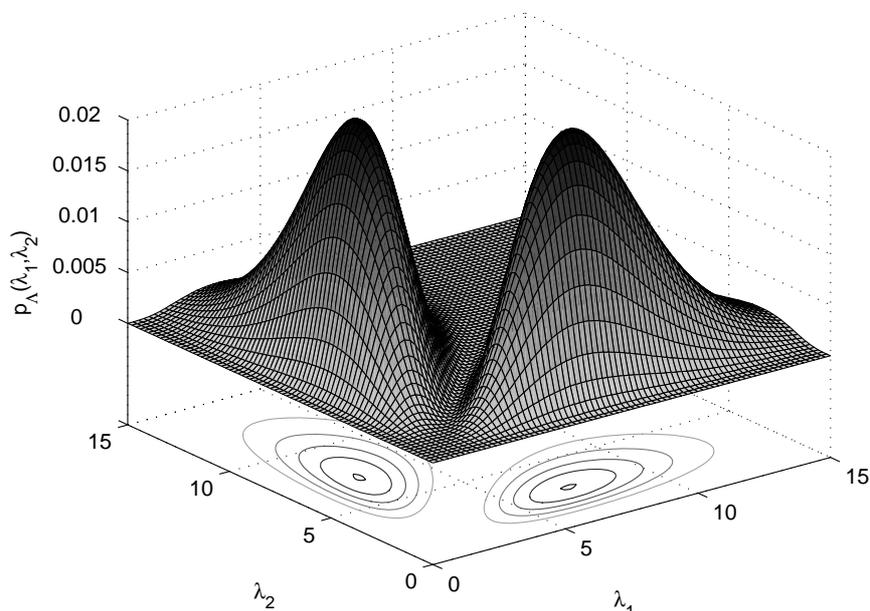


**Figure 1.12:** Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix  $\mathbf{H}^H \mathbf{H}$  with  $N_i = 2$  and  $N_o = 2$ .

(1.57)) data pipes which yield a capacity increase. This effect is called *spatial multiplexing gain*<sup>11</sup>. The basic idea of *spatial multiplexing* lies in the fact that the symbol stream to be transmitted is broken up into several parallel symbol streams which are then transmitted simultaneously from the antennas within the same frequency band. Due to multipath propagation, different spatio-temporal signatures are induced by each transmit antenna across the receive ones. The receiver exploits these signature differences to separate the individual data streams, allowing one to achieve the *multiplexing gain*. Let us note that the *multiplexing gain* can be achieved when the CSI is available at the transmitter and at the receiver, as well as when the CSI is available at only the receiver side. Clearly, the price to be paid for *multiplexing gain* is increased hardware cost due to the use of multiple antennas.

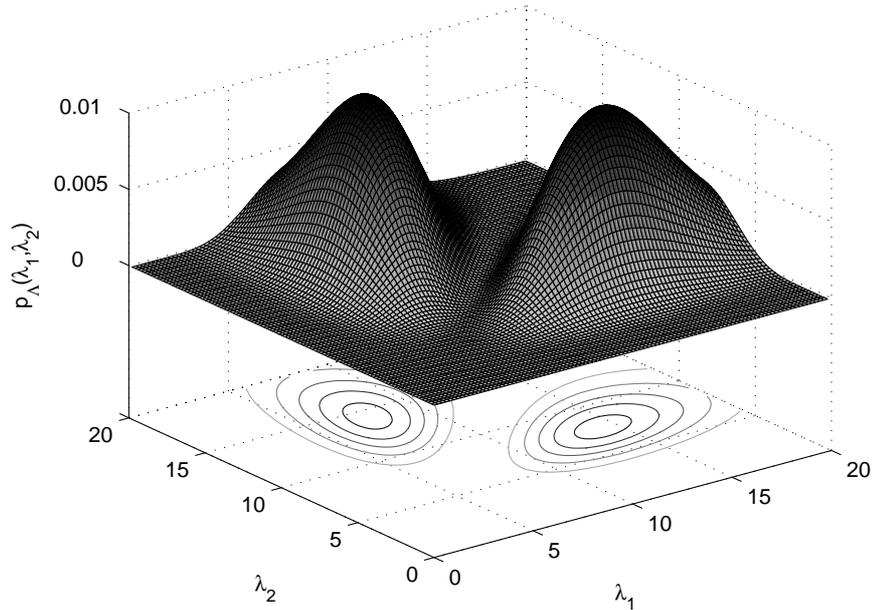
<sup>11</sup>The adjective *spatial* is mainly due to the identification of the MIMO systems with the multi-antenna systems.

By observing the joint densities in Fig. 1.12-1.14 for different values of  $N_o$ , it should be noted that, for the smallest values of  $N_o$ , the *multiplexing gain* is not supported by a good reliability of both the subchannels over which the two data pipes are transmitted. In fact, the two subchannel gains might be quite different, and we expect a good quality transmission over only one of them. In such a case, a possible choice is to give up the idea of maximizing the channel capacity, and maximize the so-called *array gain* by utilizing the subchannel which provides the maximum channel gain [22]. The *array gain* is related to the increased signal-to-noise ratio measured at the output of the receiver, and it is achieved by coherently combining (at the transmitter side) the signals to be transmitted and by coherently combining (at the receiver side) the received signals. It should be noted that, unlike *multiplexing gain*, the *array gain* requires the channel knowledge at the transmitter and at the receiver.



**Figure 1.13:** Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix  $\mathbf{H}^H \mathbf{H}$  with  $N_i = 2$  and  $N_o = 6$ .

Finally, the curves in Fig. 1.12-1.14 allow us to state that, as  $N_o$  increases, the reliability of the subchannels, over which the two data pipes are transmit-



**Figure 1.14:** Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix  $\mathbf{H}^H \mathbf{H}$  with  $N_i = 2$  and  $N_o = 12$ .

ted, become comparable since the corresponding subchannel gains are comparable. Such an effect is ascribed to the *antenna diversity*, which increases when the number of employed antennas grows. The *antenna diversity* is one of the main forms<sup>12</sup> of diversity traditionally exploited in communications. The basic principle of diversity is that if several replicas of the information signal are received through independently fading links (branches), then, with high probability, at least one or more of these links will not be in a deep fade at any given instant. Clearly, this probability will increase if the number of diversity branches increases. For such a reason, regarding to the considered example, the increase of the number of receive antennas allows the receiver to combine the arriving signals to achieve an higher gain over each one of the two *spatial modes*, or, equivalently, to reduce the probability that one of them is in

<sup>12</sup>Other widely used forms of diversity are the temporal diversity, the frequency diversity, and the code diversity. More general forms of diversity are obtained by the combination of the above-mentioned ones.

deep fading. As a consequence, the increased system capacity is supported by a good reliability of both the parallel subchannels. The price to be paid is, also in this case, the hardware cost. Receive diversity, i.e., the use of multiple antennas only at the receiver side, has been widely studied in the past. On the other hand, transmit diversity has become an active area of research in the past few years [23, 24, 25, 26, 27, 28]. More specifically, when the CSI is available at both the transmitter and the receiver side, the transmitter and the receiver can combine the transmitted and the received signals, respectively, to reduce the effects of fading (see Section 1.8). In the case where CSI is available only at the receiver side, transmit diversity techniques require more sophisticated methods such as space-time coding (STC), which uses coding across space and time and allows the receiver to achieve the diversity (see for example [17, 29, 30]).

## 1.7 MIMO receiver architectures

In this section, we provide a very brief overview of the main receiving architectures for MIMO channels, which allow to achieve a *multiplexing gain*. Let us first introduce the working framework. At the transmitter side, the data stream to be transmitted is demultiplexed into  $N_i$  streams  $x_k^{(\ell)}$  ( $\ell = 1, \dots, N_i$ ) which, after coding and modulation, are simultaneously sent over many antennas with symbol period equal to  $T_s$ . At the receiver side,  $N_o$  antennas are employed to recover as many superpositions of the transmitted signals. The received signals  $y^{(\ell)}(t)$  are  $T_s$ -space sampled<sup>13</sup> and, then, are processed to separate the different  $N_i$  transmitted sequences  $x_k^{(\ell)}$ , which are finally remultiplexed to recover the original data stream. The separation step can be performed according to different (optimization) criteria and, clearly, it determines the computational complexity of the receiver: in practical scenarios, the aim to be pursued is represented by the achievement of an acceptable compromise between performance and computational complexity. For such a reason, the maximum likelihood (ML) receiver, that yields the best performance in terms of error rate at the expense of computational complexity, is often replaced by suboptimal equalizers that exhibit a sustainable complexity.

In the following, we introduce some of the main receiver architectures for MIMO receivers:

- ML receiver

---

<sup>13</sup>We have assumed  $T_c = T_s$  in (1.5).

- minimum mean square (MMSE) error linear receiver
- zero-forcing (ZF) linear receiver
- decision-feedback (DF) based receiver.

The MMSE receiver structures and the DF-based ones only introduced here, constitute the main subject of the Chapter 2.

### 1.7.1 ML receiver

Consider a time non-dispersive LTI MIMO channel with  $N_i$  inputs and  $N_o$  outputs. The input-output relationship corresponding to the  $\ell$ th output can be specialized as follows:

$$y_k^{(\ell)} = \sum_{i=1}^{N_i} x_k^{(i)} h^{(\ell,i)} + n_k^{(\ell)} \quad \ell = 1, \dots, N_o \quad (1.70)$$

where  $h^{(\ell,i)} \triangleq h_k^{(\ell,i)}$  ( $\forall k$ ), and where  $x_k^{(i)}$  is drawn from the alphabet<sup>14</sup>  $\mathcal{A}$ . We aim at recovering the transmitted symbol  $x_k^{(\ell)}$  from the observation  $y_k^{(\ell)}$  in the case where the channel impulse response is known at the receiver. By denoting with  $f_n(\cdot)$  the probability density function of the additive noise  $n_k^{(\ell)}$ , the likelihood function of the observation, conditioned on the symbols  $x_k^{(i)}$  ( $i = 1, \dots, N_i$ ) is equal to

$$\mathcal{L} \left( y_k^{(\ell)} / \mathbf{x}_k \right) = f_n \left( y_k^{(\ell)} - \sum_{i=1}^{N_i} x_k^{(i)} h^{(\ell,i)} \right) . \quad (1.71)$$

The ML symbol decision is given simply by the argument that maximizes  $\mathcal{L} \left( y_k^{(\ell)} / \mathbf{x}_k \right)$  over the symbol alphabet

$$\hat{x}_k^{(\ell)} = \operatorname{argmax}_{\mathbf{x} \in \mathcal{A}^{N_i}} \left( \mathcal{L} \left( y_k^{(\ell)} / \mathbf{x}_k = \mathbf{x} \right) \right) . \quad (1.72)$$

Thus, the ML detection requires an exhaustive search over a total of  $\mathcal{A}^{N_i}$  vector symbols, rendering the decoding complexity exponential in the number of channel inputs.

<sup>14</sup>For the sake of clarity, we have assumed the same alphabet for all the input sequences; a more general framework provides to account for the presence of different symbol alphabets as it will be considered in Chapter 2.

In the more general case of time-dispersive LTI MIMO channel, we should consider the likelihood function of the observation conditioned on the frame of symbols  $x_k^{(i)}, x_{k-1}^{(i)}, \dots, x_{k-\nu}^{(i)}$  ( $\forall i$ ). Thus, the exponential complexity of the ML receiver increases simultaneously with the number of inputs and with the channel memory, making its implementation costly for MIMO detection on severe ISI channels, especially as the input signal constellation size increases to improve spectral efficiency.

### 1.7.2 MMSE FIR equalizer

Let us consider the time-dispersive LTI MIMO channel model in (1.25). For a block of  $N_f$  received symbols, rewrite the system equation in the following matrix form:

$$\begin{bmatrix} \mathbf{y}_k \\ \mathbf{y}_{k-1} \\ \vdots \\ \mathbf{y}_{k-(N_f-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_\nu & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_\nu & \dots & \mathbf{0} \\ \vdots & & \ddots & & \ddots & & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_\nu \end{bmatrix} \begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k-1} \\ \vdots \\ \mathbf{x}_{k-(N_f-1)-\nu} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_k \\ \mathbf{n}_{k-1} \\ \vdots \\ \mathbf{n}_{k-(N_f-1)} \end{bmatrix} \quad (1.73)$$

or, more compactly<sup>15</sup>,

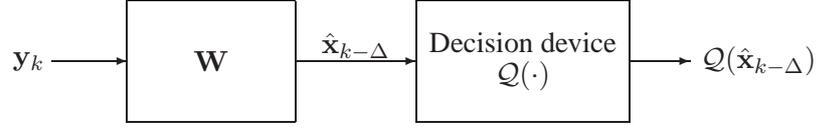
$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1.74)$$

The vector  $\mathbf{y}$  is processed by a linear FIR equalizer to provide an estimate of the transmitted symbol vector; such an estimate is then quantized by the decision device  $\mathcal{Q}(\cdot)$  to produce the symbol constellation (see Fig. 1.15).

The output of the equalizer is equal to

$$\begin{aligned} \hat{\mathbf{x}}_k &= \sum_{\ell=0}^{N_f-1} \mathbf{W}_\ell \mathbf{y}_{k-\ell} \\ &= \underbrace{\left[ \mathbf{W}_0^H \quad \mathbf{W}_1^H \quad \dots \quad \mathbf{W}_{N_f-1}^H \right]}_{\triangleq \mathbf{W}} \mathbf{y} \end{aligned} \quad (1.75)$$

<sup>15</sup>The matrix  $\mathbf{H}$  defined in (1.73) is not related to that defined in Section 1.5; we slightly abuse the notation for the sake of simplicity.



**Figure 1.15:** Block diagram of a linear receiver.

where  $\mathbf{W}_\ell$  denotes the complex-valued matrix taps of size  $N_o \times N_i$ . The equalizer output  $\hat{\mathbf{x}}_k$  is the estimate of the transmitted symbol vector  $\mathbf{x}_{k-\Delta}$ , with  $\Delta$  denoting a processing delay. The value of  $\Delta$  is related to the capability of the equalizer in performing causal processing and anticausal processing: the case  $\Delta = 0$  corresponds to a strictly causal filtering, while the case  $\Delta = N_f + \nu - 1$  corresponds to strictly anticausal filtering.

The MMSE equalizer  $\mathbf{W}$  minimizes the trace of the error correlation matrix<sup>16</sup>

$$\begin{aligned} \mathbf{R}_e &\triangleq E [(\hat{\mathbf{x}}_k - \mathbf{x}_k)(\hat{\mathbf{x}}_k - \mathbf{x}_k)^H] \\ &\triangleq E [\mathbf{e}_k \mathbf{e}_k^H] \end{aligned} \quad (1.76)$$

By resorting to the orthogonality principle, and accounting for the independence of  $\mathbf{x}$  from  $\mathbf{n}$ , the optimum  $\mathbf{W}$  can be written as follows:

$$\mathbf{W}_{MMSE} = \mathbf{R}_y^{-1} \mathbf{H} \mathbf{R}_x \mathbf{e}_{\Delta+1} \quad (1.77)$$

where

$$\begin{aligned} \mathbf{R}_x &\triangleq E [\mathbf{x} \mathbf{x}^H] \\ \mathbf{R}_n &\triangleq E [\mathbf{n} \mathbf{n}^H] \\ \mathbf{R}_y &\triangleq \mathbf{H} \mathbf{R}_x \mathbf{H} + \mathbf{R}_n \\ \mathbf{e}_{\Delta+1} &\triangleq \begin{bmatrix} \mathbf{0}_{N_i \times N_i \Delta} & \mathbf{I}_{N_i} & \mathbf{0}_{N_i \times N_i (N_f + \nu - \Delta - 1)} \end{bmatrix}^T, \end{aligned} \quad (1.78)$$

or, in other words,  $\mathbf{W}_{MMSE}$  is the conventional Wiener filter that processes the observation vector  $\mathbf{y}$  in order to estimate the desired vector  $\mathbf{x}_{k-\Delta}$ . Let us

<sup>16</sup>Let us note that it would be advisable to design the equalizer to adjust the properties of  $Q(\hat{\mathbf{x}}_k)$ , instead of  $\hat{\mathbf{x}}_k$ , for instance to minimize the error rate. However, controlling the properties of  $Q(\hat{\mathbf{x}}_k)$  is much more difficult than controlling the properties of  $\hat{\mathbf{x}}_k$ .

emphasize the importance of optimizing the delay  $\Delta$  which, as it will be shown in Chapter 2, greatly affects the trace of  $\mathbf{R}_e$ : unfortunately, the optimization over  $\Delta$  can be usually carried out only by an exhaustive procedure. Moreover, for the sake of completeness, we point out that, in the MIMO environment, different delays  $\Delta_\ell$  ( $\ell = 1, \dots, N_i$ ) can be chosen for each one of the symbols  $x_{k-\Delta_\ell}^{(\ell)}$  to be estimated. However, since the optimization over  $\Delta_\ell$  is carried out by an exhaustive procedure, the computational complexity can be unsustainable. For such a reason, in the rest of the thesis, we do not explore this variable-delay based detection strategy.

Unfortunately, some channels will still be difficult to be equalized by utilizing only a linear filter. In fact, when the channel exhibits zeros close to the unit circle, the equalizer would need poles outside the unit circle becoming unstable and, at the same time, amplifying received noise, which leads to frequent decision errors. The performance analysis of the linear MMSE FIR equalizer is presented in Chapter 2 where the MMSE equalization is studied in details, accounting also for other (nonlinear) equalizer structures.

### 1.7.3 ZF FIR equalizer

Rewrite the input-output relationship of the LTI time-dispersive MIMO channel in (1.25) as follows:

$$y_k^{(\ell)} = h_0^{(\ell,\ell)} x_k^{(\ell)} + \underbrace{\sum_{i \neq \ell} x_k^{(i)} h_0^{(\ell,i)} + \sum_{n=1}^{\nu^{(\ell,i)}} x_{k-n}^{(i)} h_n^{(\ell,i)}}_{\text{co-channel interference + ISI}} + n_k^{(\ell)} \quad (1.79)$$

where the second term at the right-hand-side (RHS) accounts for the effects of the  $N_i - 1$  inputs over the  $\ell$ th output at the time instant  $k$ , while the third term accounts for the ISI. The ZF FIR equalizer is the linear filter which processes the observation vectors  $\mathbf{y}_k, \mathbf{y}_{k-1}, \dots, \mathbf{y}_{k-(N_f-1)}$  to minimize the (co-channel interference + ISI) power measured at the output of the equalizer.

To this aim, consider the system model (1.74) and assume that the  $x_k^{(1)}$  is the symbol to be estimated<sup>17</sup> on the basis of the observation  $\mathbf{y}$ , which is rewritten in the following equivalent form:

$$\mathbf{y} = \mathbf{H}(:, 1)x_k^{(1)} + \underbrace{\mathbf{H}_{(-1)}\mathbf{x}_{(-1)}}_{\triangleq \mathbf{z}} + \mathbf{n} \quad (1.80)$$

<sup>17</sup>For the sake of clarity, we consider the case  $\Delta = 0$ .

where  $\mathbf{H}(:, k)$  denotes the  $k$ th column of  $\mathbf{H}$ ,  $\mathbf{H}_{(-1)}$  is given by  $\mathbf{H}$  deprived of its first column, and  $\mathbf{x}_{(-1)}$  is given by  $\mathbf{x}$  deprived of its first row; the vector  $\mathbf{z}$  accounts for both the co-channel interference and the ISI. According to (1.75), the first column of the ZF matrix filter  $\mathbf{W}_{ZF}$ , i.e., the vector filter that provides the estimate  $\hat{x}_k^{(1)}$  of  $x_k^{(1)}$  by processing  $\mathbf{y}$ , is obtained by solving the following optimization problem:

$$\begin{aligned} \mathbf{W}_{ZF}(:, 1) &= \underset{\mathbf{w}}{\operatorname{argmin}} |\mathbf{w}^H \mathbf{z}|^2 \\ \text{subject to: } & \mathbf{w}^H \mathbf{H}(:, 1) = \beta^2, \end{aligned} \quad (1.81)$$

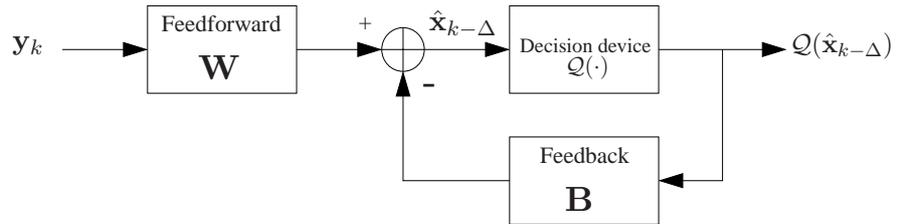
with  $\beta \in \mathbb{R}$ . As known, the optimum filter is derived by exploiting the Lagrangian multiplier method and is given by:

$$\mathbf{W}_{ZF}(:, 1) = \frac{\beta^2}{\mathbf{H}(:, 1)^H \mathbf{R}_z^{-1} \mathbf{H}(:, 1)} \mathbf{R}_z^{-1} \mathbf{H}(:, 1) \quad (1.82)$$

where  $\mathbf{R}_z \triangleq E[\mathbf{z}\mathbf{z}^H]$ . The same reasonings apply to  $\mathbf{W}_{ZF}(:, \ell)$  ( $\ell = 1, \dots, N_i$ ). For ill-conditioned  $\mathbf{H}$ , it is known that the ZF equalizer suffers from the noise enhancement; on the other hand, it is equivalent to the MMSE equalizer in presence of low noise level.

#### 1.7.4 Decision-feedback FIR equalizers

In the class of the non-linear equalizers, the DF FIR equalizer constitutes an attractive compromise between complexity and performance. It can perform almost as well as the ML detector, but it requires a computational complexity only slightly higher than the linear equalizer. Its structure is depicted in Fig. 1.16. The received signal  $\mathbf{y}_k$  is the input of a linear feedforward FIR filter, whose output is denoted with  $\mathbf{z}_k$ . The estimate  $\hat{x}_{k-\Delta}^{(\ell)}$  of the symbol  $x_{k-\Delta}^{(\ell)}$



**Figure 1.16:** The decision feedback equalizer structure.

is obtained by subtracting from  $\mathbf{z}_k$  the output of a linear feedback FIR filter, which processes the past decisions provided by the decision device on the basis of the estimated symbols. In such a way, the output of the feedforward filter can be deprived of the co-channel interference and ISI due to previously transmitted symbols. As long as the decisions are correct, the equalizer provides a good estimate of the transmitted sequences.

Differently from the conventional SISO environment, three MIMO DF equalizer structures can be defined :

**Scenario 1** The DF equalizer provides the estimate of  $x_{k-\Delta}^{(\ell)}$  ( $\ell = 1, \dots, N_i$ ) by resorting to the past decisions  $\mathcal{Q}(\hat{x}_{k-\Delta-n}^{(\ell)})$  with  $n > 0$  and  $\forall \ell$ . Such an equalizer scenario represents the MIMO DF counterpart of the conventional SISO DF equalizer.

**Scenario 2** Assume that the channel inputs are ordered so that lower indexed components of  $\mathbf{x}_k$  are detected first; then, the DF equalizer utilizes, together with past decisions, the current decisions  $\mathcal{Q}(\hat{x}_{k-\Delta}^{(1)})$ ,  $\mathcal{Q}(\hat{x}_{k-\Delta}^{(2)})$ ,  $\dots$ ,  $\mathcal{Q}(\hat{x}_{k-\Delta}^{(\ell-1)})$  to estimate the symbol  $x_{k-\Delta}^{(\ell)}$ . In other words, the decisions are taken sequentially starting with the lower indexed components.

**Scenario 3** When all the current decisions  $\mathcal{Q}(\hat{x}_{k-\Delta}^{(1)})$ ,  $\mathcal{Q}(\hat{x}_{k-\Delta}^{(2)})$ ,  $\dots$ ,  $\mathcal{Q}(\hat{x}_{k-\Delta}^{(N_i)})$  are available from a previous detection stage, then they can be processed together with past decisions to provide the estimate of the symbol of interest  $x_{k-\Delta}^{(\ell)}$ . Such a detection scenario deals with the multistage detection [31].

Accounting for the system model (1.74), the output of the DF FIR equalizer can be written as follows:

$$\hat{\mathbf{x}}_{k-\Delta} = \underbrace{\begin{bmatrix} \mathbf{W}_0^H & \mathbf{W}_1^H & \dots & \mathbf{W}_{N_f-1}^H \end{bmatrix}}_{\triangleq \mathbf{W}^H} \mathbf{y} - \underbrace{\begin{bmatrix} \mathbf{B}_0^H - \mathbf{I}_{N_i} & \mathbf{B}_1^H & \dots & \mathbf{B}_{N_b}^H \end{bmatrix}}_{\triangleq \mathbf{B}^H - \begin{bmatrix} \mathbf{I}_{N_i} & \mathbf{0}_{N_i \times N_i N_b} \end{bmatrix}} \cdot \begin{bmatrix} \mathcal{Q}(\hat{\mathbf{x}}_{k-\Delta}) \\ \mathcal{Q}(\hat{\mathbf{x}}_{k-\Delta-1}) \\ \vdots \\ \mathcal{Q}(\hat{\mathbf{x}}_{k-\Delta-N_b}) \end{bmatrix} \quad (1.83)$$

where  $N_b$  is number of the  $N_i \times N_i$  matrix taps  $\mathbf{B}_\ell$  constituting the feedback filter  $\mathbf{B}$ . The three different equalizer structures previously discussed can be

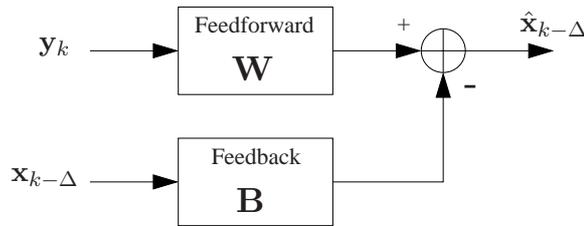
mathematically described by some constraints on the matrix tap  $\mathbf{B}_0$ . Specifically, one has that the constraint  $\mathbf{B}_0 = \mathbf{I}_{N_i}$  holds in Scenario 1 (in the following, Scenario 1), whereas, in Scenario 2 the matrix  $\mathbf{B}_0^H$  is constrained to be monic<sup>18</sup> lower triangular. Finally, in Sc. 3  $\mathbf{B}_0^H$  is constrained to be monic.

The feedforward filter  $\mathbf{W}$  and the feedback one  $\mathbf{B}$  in (1.83) can be designed according to any chosen optimization criterion. Let us note that any optimization procedure should take into account for the non-linearity of the decision device. However, also for simple decision mechanism, the derivation of a closed form for the optimum equalizer is impossible to obtain. For such a reason, in this thesis we adopt the common assumption that the decisions, which affect the current estimate, are correct, i.e.,  $\mathcal{Q}(\hat{x}_k^{(\ell)}) = x_k^{(\ell)}$ . According to such an assumption, the feedback filter can be treated as a feedforward filter which processes a delayed version of the transmitted symbols, as depicted in Fig. 1.17. However, it is clear that, in a realistic environment, the error propagation can not be ignored and the performance loss due to the feeding-back of incorrect decisions has to be measured.

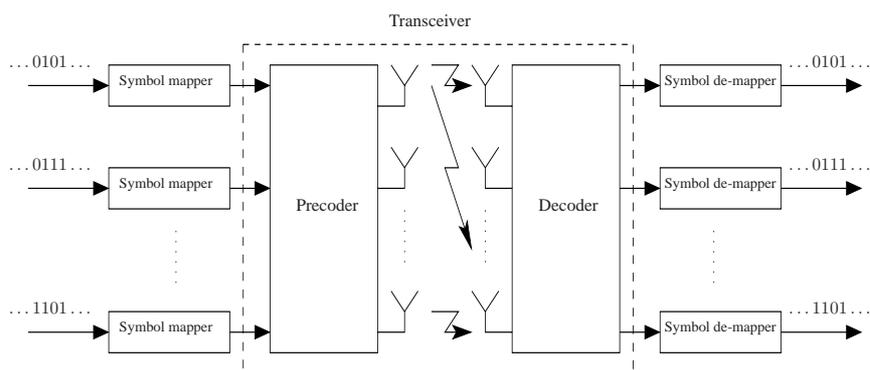
As previously discussed about the linear equalization, in all the three detection scenarios the delay  $\Delta$  has to be optimized, especially for short feedforward filters. Moreover, different delays  $\Delta_\ell$  ( $\ell = 1, \dots, N_i$ ) for each one of the symbols to be estimated can be chosen. However, apart from the computational complexity in optimizing such parameters, allowing different processing delays does not make available all the past decisions in Scenario 1, and all the current ones in Scenario 2 and Sc. 3, leading so to a more complicated mathematical formulation for the DF-based equalization.

The MMSE DF equalization will be considered in details in Chapter 2

<sup>18</sup>A square matrix with diagonal elements all equal 1.



**Figure 1.17:** The decision feedback equalizer in absence of error propagation.



**Figure 1.18:** The transceiver architecture.

with reference to the scenarios Scenario 1 and Scenario 2. Moreover, we will present an equalizer structure that combines the DF strategy with the *widely linear processing*, which allows to improve the performances of the conventional DF equalizers based on the linear filtering.

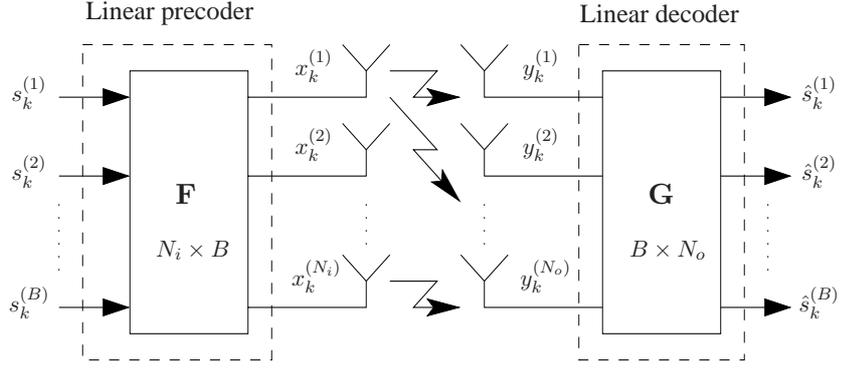
## 1.8 Transceiver architectures

Let us consider the MIMO communication system model depicted in Fig. 1.18. At the transmitter side, the information bit streams are encoded and modulated to generate the information symbol streams. Hence, such streams are processed by a precoder and transmitted over the MIMO channel. At the receiver side, the channel outputs are processed by the decoder which provides an estimate of the precoder inputs. Finally, the Viterbi decoder allows to recover the (estimated) information bit streams. When the channel state information (CSI) is available at both ends of the link, the precoder and the decoder can be jointly designed, according to the chosen optimization criterion, to improve the system performances.

In Fig. 1.19 we have depicted a transceiver structure employing a linear filter as precoder and decoder<sup>19</sup>. For simplicity, we assume the time non-dispersive channel model

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k \quad (1.84)$$

<sup>19</sup>More general structures can be considered (see, for instance, [32] and [33]), but they are out of the scope of this thesis.



**Figure 1.19:** The transceiver architecture.

affected by additive spatially and temporally white noise with correlation matrix  $\mathbf{R}_n = \mathbf{I}_{N_o}$ . The symbol vector to be transmitted is denoted with  $\mathbf{s}_k \triangleq [s_k^{(1)}, s_k^{(2)}, \dots, s_k^{(B)}]^T$  with  $s_k^{(\ell)}$  drawn from the constellation  $\mathcal{A}_\ell$  ( $\ell = 1, \dots, B$ ). Moreover, we assume  $E[\mathbf{s}_k \mathbf{s}_{k-n}^H] = \mathbf{I}_B \delta_{k-n}$ . The precoder  $\mathbf{F} \in \mathbb{C}^{N_i \times B}$  processes  $\mathbf{s}_k$  and provides the channel input vector  $\mathbf{x}_k \triangleq \mathbf{F} \mathbf{s}_k$  of size  $N_i$ . At the receiver side, the equalizer provides the estimate  $\hat{\mathbf{s}}_k$  of  $\mathbf{s}_k$  by processing the received vector  $\mathbf{y}_k$ . The overall system equation is given by:

$$\hat{\mathbf{s}}_k = \mathbf{G} \mathbf{H} \mathbf{F} \mathbf{s}_k + \mathbf{G} \mathbf{n}_k \quad . \quad (1.85)$$

The transceiver can be optimized according to the MMSE criterion as well as the ZF criterion. In addition, the transmitter and the receiver can be jointly designed to maximize the mutual information between precoder input and decoder output, say  $\mathcal{I}(\mathbf{s}_k, \hat{\mathbf{s}}_k)$ : in such a case, it has been shown in [34] that the precoder maximizing the mutual information is unique, whereas the optimum decoder is nonunique and the available degrees of freedom can be utilized to synthesize the receiver according any other optimization criterion.

It can be shown that (see [34, 25]), according to any optimization criterion, the optimum matrices  $\mathbf{F}$  and  $\mathbf{G}$  are given by

$$\begin{aligned} \mathbf{F}^{(opt)} &= \mathbf{V} \mathbf{\Phi} \\ \mathbf{G}^{(opt)} &= \mathbf{\Gamma} \mathbf{\Lambda}^{-1} \mathbf{V}^H \mathbf{H}^H \end{aligned} \quad (1.86)$$

where  $\mathbf{V}$  and  $\mathbf{\Lambda}$  are the eigenvector and the eigenvalue matrices, respectively, of  $\mathbf{H}^H \mathbf{H}$  defined in (1.47), and where  $\mathbf{\Phi}$  and  $\mathbf{\Gamma}$  represent a diagonal matrix

with positive entries and an invertible matrix, respectively, that depend on the chosen optimization criterion.

Let now  $\mathbf{\Gamma}$  be a diagonal matrix. It is straightforward to verify that, in such a special case, the overall MIMO system is described by the diagonal matrix  $\mathbf{\Phi}\mathbf{\Gamma}$  and, hence, the transmission over the MIMO channel  $\mathbf{H}_k$  is equivalent to  $\text{rank}(\mathbf{H}_k)$  transmissions over  $\text{rank}(\mathbf{H}_k)$  parallel non-dispersive subchannels characterized by a transmit gain  $\phi_i$ , corresponding to the  $i$ th diagonal element of  $\mathbf{\Phi}$ , and a receive gain  $\gamma_i$ , corresponding to the  $i$ th diagonal element of  $\mathbf{\Gamma}$ . It can be shown that such a model arises when  $\mathcal{I}(\mathbf{s}_k, \hat{\mathbf{s}}_k)$  has to be maximized, as well as when we adopt the MMSE criterion or the ZF one.

Before concluding this section, it is important to underline two interesting issues recently discussed in [35] about MIMO communication systems:

- According to the transceiver defined in (1.86), let us evaluate the correlation matrix of the estimation error measured at the output of the decoder:

$$\mathbf{R}_e \triangleq E [(\hat{\mathbf{s}}_k - \mathbf{s}_k)(\hat{\mathbf{s}}_k - \mathbf{s}_k)^H] \quad (1.87)$$

$$= \mathbf{\Gamma} (\mathbf{\Phi}^2 + \mathbf{\Lambda}^{-1}) \mathbf{\Gamma}^H - \mathbf{\Phi}\mathbf{\Gamma}^H + \mathbf{\Gamma}\mathbf{\Phi} + \mathbf{I}_B \quad (1.88)$$

The choice of a diagonal  $\mathbf{\Gamma}$  allows one to obtain uncorrelated estimates of  $\mathbf{s}_k$  and, hence, uncorrelated estimations error. This represents an important advantage provided by such a transceiver structure, since the decision device can separately detect the transmitted symbols, requiring so a lower computational complexity<sup>20</sup>.

- It can be simply verified that, given a diagonal  $\mathbf{\Gamma}$ , the mutual information  $\mathcal{I}(\mathbf{s}_k, \hat{\mathbf{s}}_k)$  is equal to:

$$\mathcal{I}(\mathbf{s}_k, \hat{\mathbf{s}}_k) = \sum_{i=1}^{\text{rank}(\mathbf{H}_k)} \log(1 + \lambda_i \phi_i^2) \quad , \quad (1.89)$$

i.e., it is equal to sum of the capacity of  $\text{rank}(\mathbf{H}_k)$  SISO non-dispersive channels. However, note that, due to the different values of the eigenvalues  $\lambda_\ell$ , one has that:

1. since different symbol rates are achieved over each subchannel, different symbol alphabets can be utilized at the transmitter:

---

<sup>20</sup>When  $\mathbf{R}_e$  is not diagonal, the correlation among the different estimates contain useful information to be utilized by the subsequent decoding; hence, the optimum decoding procedure becomes more complicated.

specifically, dense constellations can be transmitted over subchannels corresponding to high  $\lambda_\ell$ , while thin constellations has to be utilized over subchannels corresponding to low  $\lambda_\ell$ . Hence, a more complicated encoder/decoder device is in general required to achieve the capacity;

2. different error rates are achieved over each subchannel. Such an undesirable behavior can be overcome by exploiting non linear processing techniques which allow us to design the transceiver in a more flexible manner. For instance, in [36], the precoder is designed to maximize  $\mathcal{I}(\mathbf{s}_k, \hat{\mathbf{s}}_k)$ , whereas a DF-based decoder, designed according to the MMSE criterion, is employed<sup>21</sup>: it can be shown that such a transceiver structure allows one to reach the same error rate over all the subchannel. An alternative, but expensive, solution is to utilize a large number of receiving antennas, i.e.,  $N_o \gg N_i$ . In fact, according to Theorem 1.2, for small values of  $\beta$  the eigenvalues  $\lambda_i$  are comparable. To this aim, in Chapter 3, we propose a new transceiver design method that, without increasing the number of receiving antennas, allows one to reduce the value of  $\beta$  when real-valued constellation sets are provided by the encoder.

In Chapter 3, the transceiver optimization will be carried out by utilizing *widely linear* filters as precoder and decoder. We will show that such a new transceiver structure generalizes the one based on the conventional linear filtering.

---

<sup>21</sup>Analogously, in [37], the authors synthesize the decoder according to the ZF criterion.

## Chapter 2

# Decision-feedback equalization

**H**igh bit-rate transmissions require to account for the temporal dispersive nature of the communication media, especially for mobile applications where the symbol period must be also short enough to prevent non-stationary channel impairments. Since the linear equalizers perform poorly on severe-ISI channels, the DF-based equalizers, which employ also a linear filter operating on the past decisions, have been extensively proposed. It has been shown that DF strategy allows one to achieve significant performance improvements over linear equalizers both in SISO and in MIMO scenarios. Two DF structures are mainly considered in this chapter. The former exploits (together with the received signals) the past decisions in order to simultaneously estimate all the transmitted signals. Instead, the latter utilizes also the decisions belonging to the same time step providing the estimate of the transmitted signal sequentially: in such a case, the decision ordering represents a degree of freedom with no counterparts in SISO DF-equalization, and it has to be optimized according to a chosen optimization criterion. Unfortunately, the optimization over the decision ordering is NP-hard and hence, suboptimum ordering strategies with polynomial complexity (as example V-BLAST [24]) has to be adopted. With reference to time-dispersive environments, it will be shown that the DF-MMSE equalization can be performed by a two-stage equalizer. The first stage is a time-dispersive equalizer and is unaffected by the decision ordering, whereas the second stage is a zero-memory equalizer depending on the ordering. Owing to such a representation, any decision ordering strategy for non-dispersive environment can be extended to the dispersive one. proposed

The widely linear (WL) filtering [38, 39, 40], as it is well-known, generalizes the linear filtering by exploiting both the correlation and the conju-

gate correlation, i.e., the cross correlation between the signal and its conjugate counterpart to better discriminate useful signals against interference and noise signals. In this chapter, the receiver structure that combines DF strategy with WL filtering, say widely-linear/widely decision feedback (WL-WDF) equalizer, is presented. It will be shown that, without requiring a significant increase in computational complexity, such an equalizer allows one to achieve considerable performance improvements over the DF equalizer based on the conventional linear filtering when circularly variant signals (whose properties are described in Appendix A) are transmitted. The performances of the proposed equalizer are compared with those of the linear feedforward-based equalizer, the DF-MMSE equalizer [9], the WL feedforward-based equalizer [41, 42], and, finally, with those of the DF-based equalizer which employs a WL feedforward filter and a linear feedback one (proposed in [43] with reference to the SISO channels).

The WL-WDF equalizer can be derived according to two alternative representations of the involved signals. The former performs the linear processing of both the real and the imaginary parts of the input vector [39, 44], whereas, the latter performs the linear processing of the input vector and its conjugate version [41, 42, 43, 45, 46]. It is well known that the two representations are equivalent, namely a one-to-one correspondence between each processing and the other. However, it is shown that they are not anymore equivalent when the DF equalizer utilizes both past decisions and decisions belonging to the same time step in order to achieve better performances.

## 2.1 Widely linear processing in the MMSE estimation

In this section, with reference to the mean square error (MSE) estimation of complex data, the Wiener filtering will be generalized by introducing the non strictly linear filtering called *widely linear* (WL), which allows one to exploit the correlation among the data and their conjugate version.

Consider the case where an  $N$ -dimensional zero-mean observation random vector  $\mathbf{r}$  is utilized to estimate the zero-mean  $M$ -dimensional random vector  $\mathbf{d}$ . According to the conventional linear filtering, the estimate  $\hat{\mathbf{d}}$  of  $\mathbf{d}$  can be written in the general form:

$$\hat{\mathbf{d}} = \mathbf{W}^H \mathbf{r} \quad (2.1)$$

where  $\mathbf{W}$  is an  $N \times M$  complex-valued matrix. Define the error vector  $\mathbf{e} \triangleq \mathbf{d} - \hat{\mathbf{d}}$ , whose correlation matrix is denoted with  $\mathbf{R}_{ee} = E[\mathbf{e}\mathbf{e}^H]$ . The

## 2.1. WIDELY LINEAR PROCESSING IN THE MMSE ESTIMATION 51

linear minimum mean square error (MMSE) filter (or Wiener filter) provides an estimate  $\hat{\mathbf{d}}$  of  $\mathbf{d}$  by minimizing the trace of  $\mathbf{R}_{ee}$ , i.e.:

$$\mathbf{W}^{(opt)} = \underset{\mathbf{W} \in \mathbb{C}^{N \times M}}{\operatorname{argmin}} \operatorname{trace}(\mathbf{R}_{ee}) \quad . \quad (2.2)$$

By differentiating<sup>1</sup> the scalar function  $\operatorname{trace}(\mathbf{R}_{ee})$  with respect to  $\mathbf{W}$ , one has that the optimum filter is equal to

$$\mathbf{W}^{(opt)} = \mathbf{R}_{rr}^{-1} \mathbf{R}_{rd} \quad (2.3)$$

with  $\mathbf{R}_{rr} \triangleq E[\mathbf{r}\mathbf{r}^H]$  and  $\mathbf{R}_{rd} \triangleq E[\mathbf{r}\mathbf{d}^H]$ , and the optimum error correlation matrix is given by

$$\mathbf{R}_{ee}^L = \mathbf{R}_{dd} - \mathbf{R}_{rd}^H \mathbf{R}_{rr}^{-1} \mathbf{R}_{rd} \quad , \quad (2.4)$$

with  $\mathbf{R}_{dd} \triangleq E[\mathbf{d}\mathbf{d}^H]$ . Let us show that the linear transformation  $\mathbf{W}$  from  $\mathbf{r}$  to  $\hat{\mathbf{d}}$  in (2.1) does not represent the most general form of a linear transformation when dealing with complex-valued vectors. To this aim, rewrite (2.1) in terms of the real and the imaginary parts of the involved vectors ( $\mathbf{r}$  and  $\hat{\mathbf{d}}$ ):

$$\begin{bmatrix} \Re\{\hat{\mathbf{d}}\} \\ \Im\{\hat{\mathbf{d}}\} \end{bmatrix} = \underbrace{\begin{bmatrix} \Re\{\mathbf{W}^T\} & \Im\{\mathbf{W}^T\} \\ -\Im\{\mathbf{W}^T\} & \Re\{\mathbf{W}^T\} \end{bmatrix}}_{\widetilde{\mathbf{W}}^T} \cdot \begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix} \quad . \quad (2.5)$$

It is easy to understand that the block structure of the matrix  $\widetilde{\mathbf{W}}$  in (2.5) does not allow one to perform the more general linear transformation from  $[\Re\{\mathbf{r}\} \ \Im\{\mathbf{r}\}]^T$  to  $[\Re\{\hat{\mathbf{d}}\} \ \Im\{\hat{\mathbf{d}}\}]^T$ . In fact,  $\Re\{\mathbf{W}\}$  and  $\Im\{\mathbf{W}\}$  simultaneously affect the real and the imaginary part of  $\hat{\mathbf{d}}$ .

The *widely linear* (WL) filtering generalizes the conventional linear filtering by assuming that the estimate  $\hat{\mathbf{d}}$  of  $\mathbf{d}$  is obtained by performing the linear transformation

$$\begin{bmatrix} \Re\{\hat{\mathbf{d}}\} \\ \Im\{\hat{\mathbf{d}}\} \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{W}}_{11}^T & \widetilde{\mathbf{W}}_{21}^T \\ \widetilde{\mathbf{W}}_{12}^T & \widetilde{\mathbf{W}}_{22}^T \end{bmatrix} \cdot \begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix} \quad (2.6)$$

where  $\widetilde{\mathbf{W}}_{\ell k} \in \mathbb{R}^{N \times M}$  ( $\ell, k = 1, 2$ ). In other words, the output vector  $\hat{\mathbf{d}}$  of the WL filter is such that its real and imaginary parts are obtained by separately

<sup>1</sup>When differentiating  $\operatorname{trace}(\mathbf{R}_{ee})$  with respect to  $\mathbf{G}$ , we treat  $\mathbf{G}$  and  $\mathbf{G}^H$  as independent variables, i.e.,  $\partial(\operatorname{trace}(\mathbf{A}\mathbf{X}\mathbf{B})) / (\partial\mathbf{X}) = \mathbf{B}\mathbf{A}$  and  $\partial(\operatorname{trace}(\mathbf{A}\mathbf{X}^H\mathbf{B})) / (\partial\mathbf{X}) = \mathbf{0}$ .

processing the real and the imaginary parts of the input vector  $\mathbf{r}$ . It follows that the linear filters in (2.5) lie in a subset of the WL filters and, consequently, according to any chosen optimization criterion, their performances cannot be better than those obtained by WL filters. The WL transformation in (2.6) can be rewritten in the equivalent form:

$$\begin{bmatrix} \hat{\mathbf{d}} \\ \hat{\mathbf{d}}^* \end{bmatrix} = \begin{bmatrix} \mathbf{F}^H & \mathbf{G}^H \\ \mathbf{G}^T & \mathbf{F}^T \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r} \\ \mathbf{r}^* \end{bmatrix} \quad (2.7)$$

where the  $\mathbf{F}, \mathbf{G} \in \mathbb{C}^{N \times M}$  are given by

$$\begin{aligned} \Re\{\mathbf{F}\} &= \frac{1}{2} [\widetilde{\mathbf{W}}_{11} + \widetilde{\mathbf{W}}_{22}] & \Im\{\mathbf{F}\} &= \frac{1}{2} [\widetilde{\mathbf{W}}_{21} - \widetilde{\mathbf{W}}_{12}] \\ \Re\{\mathbf{G}\} &= \frac{1}{2} [\widetilde{\mathbf{W}}_{11} - \widetilde{\mathbf{W}}_{22}] & \Im\{\mathbf{G}\} &= \frac{1}{2} [-\widetilde{\mathbf{W}}_{21} - \widetilde{\mathbf{W}}_{12}] \end{aligned} \quad (2.8)$$

According to (2.7), the vector  $\hat{\mathbf{d}}$  at the output of a WL filter is obtained by separately processing the input vector  $\mathbf{r}$  and its conjugate version; for such a reason, it is straightforward to observe that the  $\hat{\mathbf{d}}$  is not a linear function in  $\mathbf{r}$ . On the other hand, the conventional linear filter is simply obtained by setting  $\mathbf{G} = \mathbf{0}$ .

We must recognize that the WL processing defined according to (2.7) (in the following, referred to as real-valued representation) is widely adopted in the literature [41, 42, 43, 45, 46] since the second-order moments of  $\hat{\mathbf{d}}$  are directly related to the second-order moments of  $\mathbf{r}$  and  $\mathbf{r}^*$ , i.e., to the correlation matrices  $\mathbf{R}_{rr}$  and  $\mathbf{R}_{rr^*}$  introduced in Section A.1.1 of the Appendix A. Furthermore, the comparison reported below between the MMSE WL estimator and the linear one (obtained by setting  $\mathbf{G} = \mathbf{0}$ ) is less obvious when the real-valued definition in (2.6) is utilized to define the WL filtering. For explanation purposes, in this section, we adopt the complex-valued representation, whereas, in the rest of the thesis, we will adopt the real-valued one. In Section 2.2, however, we introduce some useful operators which allow one to study the main correspondence and differences among the two representations.

The optimum (in the MMSE sense) WL filter can be obtained by solving the optimization problem:

$$\left( \mathbf{F}^{(opt)}, \mathbf{G}^{(opt)} \right) = \underset{\mathbf{F}, \mathbf{G} \in \mathbb{C}^{N \times M}}{\operatorname{argmin}} \operatorname{trace}(\mathbf{R}_{ee}) \quad (2.9)$$

being  $\mathbf{R}_{ee}$  the correlation matrix of the error vector corresponding to the estimation  $\hat{\mathbf{d}} = \mathbf{F}^H \mathbf{r} + \mathbf{G}^H \mathbf{r}^*$ . By differentiating the scalar function  $\operatorname{trace}(\mathbf{R}_{ee})$

with respect to  $\mathbf{F}$  and  $\mathbf{G}$ , one has:

$$\mathbf{F}^{(opt)} = [\mathbf{R}_{rr} - \mathbf{R}_{rr^*} \mathbf{R}_{rr}^{-*} \mathbf{R}_{rr^*}^*]^{-1} \cdot [\mathbf{R}_{rd} - \mathbf{R}_{rr^*} \mathbf{R}_{rr}^{-*} \mathbf{R}_{rd}^*] \quad (2.10)$$

$$\mathbf{G}^{(opt)} = [\mathbf{R}_{rr} - \mathbf{R}_{rr^*} \mathbf{R}_{rr}^{-*} \mathbf{R}_{rr^*}^*]^{-*} \cdot [\mathbf{R}_{rd^*} - \mathbf{R}_{rr^*} \mathbf{R}_{rr}^{-*} \mathbf{R}_{rd}^*]^* \quad (2.11)$$

where  $\mathbf{R}_{rr^*} \triangleq E[\mathbf{r}\mathbf{r}^T]$  and  $\mathbf{R}_{rd^*} \triangleq E[\mathbf{r}\mathbf{d}^T]$ , and where we assume that the involved inverse matrices exist. The optimum error-correlation matrix is equal to

$$\mathbf{R}_{ee}^{WL} = \mathbf{R}_{dd} - \left( \mathbf{F}^{(opt)H} \mathbf{R}_{rd} + \mathbf{G}^{(opt)H} \mathbf{R}_{rd^*} \right) \quad (2.12)$$

The performance advantage of the WL estimator over the linear one is characterized by the matrix  $\Delta_{ee} \triangleq \mathbf{R}_{ee}^{WL} - \mathbf{R}_{ee}^L$  which can be expressed as follows

$$\Delta_{ee} = [\mathbf{R}_{rd^*} - \mathbf{R}_{rr^*} \mathbf{R}_{rr}^{-*} \mathbf{R}_{rd}^*]^T \cdot [\mathbf{R}_{rr}^* - \mathbf{R}_{rr^*}^H \mathbf{R}_{rr}^{-1} \mathbf{R}_{rr^*}]^{-1} \cdot [\mathbf{R}_{rd^*} - \mathbf{R}_{rr^*} \mathbf{R}_{rr}^{-*} \mathbf{R}_{rd}^*]^* \quad (2.13)$$

and whose  $k$ th diagonal entry represents the MSE gain provided by the WL filter over linear one in estimating the  $k$ th entry of  $\mathbf{d}$ . Being  $[\mathbf{R}_{rr}^* - \mathbf{R}_{rr^*}^H \mathbf{R}_{rr}^{-1} \mathbf{R}_{rr^*}]$  non negative definite (see Appendix A Proposition A.1), the linear MMSE filtering can not outperform the WL MMSE filtering.

Note that, if  $\mathbf{r}$  is rotationally variant ( $\mathbf{R}_{rr^*} = \mathbf{0}$ ) and, moreover,  $\mathbf{r}$  and  $\mathbf{d}$  are cross-rotationally variant ( $\mathbf{R}_{rd^*} = \mathbf{0}$ ), then  $\mathbf{F}^{(opt)} = \mathbf{W}^{(opt)}$ ,  $\mathbf{G}^{(opt)} = \mathbf{0}$  and  $\Delta_{ee} = \mathbf{0}$ : in other words the WL MMSE estimator degenerates into the optimum linear one, providing so the same performances. On the other hand,  $\Delta_{ee}$  does not vanish when  $\mathbf{r}$  and  $\mathbf{d}$  are rotationally variant. A special case is represented by the estimation of a real-valued random vector  $\mathbf{d}$  ( $\mathbf{d} = \Re\{\mathbf{d}\}$ ) from a complex-valued observation vector. Being  $\mathbf{R}_{rd} = \mathbf{R}_{rd^*}$ , one has (see (2.10)-(2.11)):

$$\mathbf{F} = \mathbf{G}^* \Rightarrow \hat{\mathbf{d}} = 2\Re\{\mathbf{F}^H \mathbf{r}\} \quad (2.14)$$

Note that the WL filtering leads to a real-valued estimation opposite of the linear filtering which provides a complex-valued vector as estimate of a desired real-valued one.

Finally, let us consider the extreme case where  $\mathbf{r}$  is real-valued ( $\mathbf{r} = \Re\{\mathbf{r}\}$ ). Being  $\mathbf{R}_{rr} = \mathbf{R}_{rr^*}$  and real-valued, the matrix inversion in (2.10)- (2.11) and in (2.14) can not be performed. It is simple to understand that the optimum

MMSE WL filter is not unique, in fact:

$$\begin{aligned}\hat{\mathbf{d}} &= \mathbf{F}^H \mathbf{r} + \mathbf{G}^H \mathbf{r}^* \\ &= \underbrace{(\mathbf{F} + \mathbf{G})^H}_{\triangleq \mathbf{D}} \mathbf{r} .\end{aligned}\tag{2.15}$$

It follows that, given the optimum (linear filter)  $\mathbf{D}$ , there exist infinite optimum filter pairs  $(\mathbf{F}, \mathbf{G})$  whose sum is equal to the optimum  $\mathbf{D}$ . Similar considerations are valid when  $0 < n_r < N$  components of  $\mathbf{r}$  are real-valued. The (non)-uniqueness of the MMSE WL solution clearly follows from the fact that, being  $\mathbf{r} = \mathbf{r}^*$ , the jointly processing of  $\mathbf{r}$  and  $\mathbf{r}^*$  is redundant. With reference to the MIMO channel equalization based on the WL filtering, we address such a problem by processing the only real part of the real-valued observation vector components. ■

### Final remark

Based on the above considerations, we can state that, if the observation vector and/or the desired one are rotationally variant, then the adoption of a linear filter is a suboptimum choice when an MMSE estimate has to be provided on the basis of the second-order statistics. For this reason, WL filters in the MIMO receiver and transceiver architectures will be synthesized in this chapter and in Chapter 3, respectively, in presence of rotationally variant information symbols.

Note that, independently of the chosen optimization criterion, the advantage provided by WL filtering can be intuitively justified by observing that the jointly processing of  $\mathbf{r}$  and  $\mathbf{r}^*$ , in the case where they exhibit statistical correlation, allows one to increase the dimension of the observation space.

## 2.2 Widely linear transformations

The WL processing can be performed by adopting the real-valued representation of the involved vectors, as in (2.6), or the complex-valued one, as in (2.7). For this reason, we introduce some operators that allow us to:

- address the problem of finding the main correspondences and differences between the two representations;

- to synthesize the WL processing-based receiver in Section 2.5 (or the transceiver one in Chapter 3) by utilizing the procedure relative to the conventional linear processing-based structures.

To this aim, in subsection 2.2.1, we define the operators which allow us to represent a complex-valued vector by utilizing a real-valued representation or a complex-valued one, and we provide the correspondence between them. In subsection 2.2.2, we define the WL transformation with reference to both the previously introduced representations.

### 2.2.1 Real-valued and complex-valued representation

Let us define the following operators:

$$\tilde{\mathcal{E}}_p[\mathbf{u}] \triangleq \begin{bmatrix} \Re\{\mathbf{u}(1:n_1, 1:p)\} & \Re\{\mathbf{u}(1:n_1, p+1:n_2)\} \\ \Im\{\mathbf{u}(1:n_1, 1:p)\} & \Im\{\mathbf{u}(1:n_1, p+1:n_2)\} \\ & -\Im\{\mathbf{u}(1:n_1, p+1:n_2)\} \\ & \Re\{\mathbf{u}(1:n_1, p+1:n_2)\} \end{bmatrix} \quad (2.16)$$

$$\tilde{\mathcal{C}}_p[\mathbf{u}] \triangleq \begin{bmatrix} \mathbf{u}(1:n_1, 1:p) & \mathbf{u}(1:n_1, p+1:n_2) \\ \mathbf{u}^*(1:n_1, 1:p) & \mathbf{0}_{n_1 \times (n_2-p)} \\ & \mathbf{0}_{n_1 \times (n_2-p)} \\ & \mathbf{u}^*(1:n_1, p+1:n_2) \end{bmatrix} \quad (2.17)$$

$$\mathcal{E}_p[\mathbf{u}] \triangleq \begin{bmatrix} \Re\{\mathbf{u}\} \\ \Im\{\mathbf{u}(p+1:n_1, 1:n_2)\} \end{bmatrix} \quad (2.18)$$

$$\mathcal{C}_p[\mathbf{u}] \triangleq \begin{bmatrix} \mathbf{u} \\ \mathbf{u}^*(p+1:n_1, 1:n_2) \end{bmatrix} \quad (2.19)$$

where  $\mathbf{u} \in \mathbb{C}^{n_1 \times n_2}$ ,  $0 \leq p \leq n_1$  is an integer value,  $\mathbf{u}(i_1 : \ell_1, i_2 : \ell_2)$  is the submatrix of  $\mathbf{u}$ , whose first and last rows (columns) are the  $i_1$ th ( $i_2$ th) and the  $\ell_1$ th ( $\ell_2$ th) ones, respectively, and where the array  $\mathbf{0}_{n_1 \times n_2}$  is the  $n_1 \times n_2$  matrix containing all null entries (the specification of the size  $n_1 \times n_2$  will be omitted in the sequel for the sake of brevity). Note that, if the first  $p$  components of the  $n_1 \times 1$  vector  $\mathbf{x}$  are real-valued, the augmented vector  $\mathcal{E}_p[\mathbf{x}]$  does not contain

the  $p$  identically null imaginary parts of the first  $p$  components of  $\mathbf{x}$ . Moreover, it is easy verified that the following equivalences hold:

$$\mathbf{x} = \mathbf{y} \Leftrightarrow \mathcal{C}_p[\mathbf{x}] = \mathcal{C}_p[\mathbf{y}] \Leftrightarrow \mathcal{E}_p[\mathbf{x}] = \mathcal{E}_p[\mathbf{y}] \Leftrightarrow \tilde{\mathcal{C}}_p[\mathbf{x}] = \tilde{\mathcal{C}}_p[\mathbf{y}] \Leftrightarrow \tilde{\mathcal{E}}_p[\mathbf{x}] = \tilde{\mathcal{E}}_p[\mathbf{y}] \quad (2.20)$$

$$\mathbf{x} + \mathbf{y} = \mathbf{z} \Leftrightarrow \mathcal{C}_0[\mathbf{x}] + \mathcal{C}_0[\mathbf{y}] = \mathcal{C}_0[\mathbf{z}] \Leftrightarrow \mathcal{E}_0[\mathbf{x}] + \mathcal{E}_0[\mathbf{y}] = \mathcal{E}_0[\mathbf{z}] \quad (2.21)$$

$$\mathbf{h}\mathbf{x} = \mathbf{s} \Leftrightarrow \tilde{\mathcal{C}}_p[\mathbf{h}]\mathcal{C}_p[\mathbf{x}] = \mathcal{C}_0[\mathbf{s}] \Leftrightarrow \tilde{\mathcal{E}}_p[\mathbf{h}]\mathcal{E}_p[\mathbf{x}] = \mathcal{E}_0[\mathbf{s}] \quad (2.22)$$

where  $\mathbf{y}$  and  $\mathbf{z}$  are  $n_1 \times 1$  vectors,  $\mathbf{h}$  is a  $n_2 \times n_1$  matrix and  $\mathbf{s}$  is a  $n_2 \times 1$  vector. The operators  $\tilde{\mathcal{E}}_p[\cdot]$  and  $\mathcal{E}_p[\cdot]$ , as it will be shown in Section 2.5, are useful to rewrite the input-output relationship of a linear FIR MIMO system in terms of the real and the imaginary parts of the involved vectors; analogously, the operators  $\tilde{\mathcal{C}}_p[\cdot]$  and  $\mathcal{C}_p[\cdot]$  are useful to rewrite such a relationship in terms of the involved complex-valued vectors and their conjugate counterparts. Define also

$$\tilde{\mathcal{C}}_p \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \mathbf{u}_1 \quad \tilde{\mathcal{E}}_p \begin{bmatrix} \mathbf{u}_3 \\ \mathbf{u}_4 \\ \mathbf{u}_5 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_3 \\ \mathbf{u}_4 + j\mathbf{u}_5 \end{bmatrix} \quad (2.23)$$

where  $\mathbf{u}_1$  has  $n_1$  rows,  $\mathbf{u}_2, \mathbf{u}_4, \mathbf{u}_5$  has  $n_1 - p$  rows and  $\mathbf{u}_3$  have  $p$  rows. Note that, if the first  $p$  components of  $\mathbf{u}$  are real-valued, then  $\tilde{\mathcal{C}}_p[\mathcal{C}_p[\mathbf{u}]] = \mathbf{u}$  and  $\tilde{\mathcal{E}}_p[\mathcal{E}_p[\mathbf{u}]] = \mathbf{u}$ . We will show that the operators  $\tilde{\mathcal{E}}_p[\cdot]$  and  $\tilde{\mathcal{C}}_p[\cdot]$ , together with  $\mathcal{E}_p[\cdot]$  and  $\mathcal{C}_p[\cdot]$ , allow us to define the input-output relationship of the WL FIR MIMO systems.

Finally, similarly to [40], let us define the matrix transformations

$$\mathbf{T} \triangleq \begin{bmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{2}}\mathbf{I}_{n_1-p} & \frac{j}{\sqrt{2}}\mathbf{I}_{n_1-p} \\ \mathbf{0} & \frac{1}{\sqrt{2}}\mathbf{I}_{n_1-p} & \frac{-j}{\sqrt{2}}\mathbf{I}_{n_1-p} \end{bmatrix} \quad \text{with} \quad \mathbf{T}\mathbf{T}^H = \mathbf{T}^H\mathbf{T} = \mathbf{I}_{2n_1-p} \quad (2.24)$$

and

$$\mathbf{Y} \triangleq \begin{bmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sqrt{2}\mathbf{I}_{n_1-p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sqrt{2}\mathbf{I}_{n_1-p} \end{bmatrix}. \quad (2.25)$$

If  $\mathbf{u}$  is a complex-valued vector with  $n_1$  rows such that the first  $p$  rows are real-valued, then  $\mathcal{C}_p[\mathbf{u}] = \mathbf{Y}\mathbf{T}\mathcal{E}_p[\mathbf{u}]$ .

### 2.2.2 WL transformations

Let  $\mathbf{x}$  be a vector of size  $N_i$  whose  $n_r$  first components are real-valued, and let  $\mathbf{y}$  be a vector of size  $N_o$  whose first  $n_q$  components are real-valued. By adopting the real-valued representation, the WL transformation from  $\mathbf{x}$  to  $\mathbf{y}$  is defined as the linear transformation on the extended vector  $\mathcal{E}_{n_r}[\mathbf{x}]$ , namely:

$$\mathcal{E}_{n_q}[\mathbf{y}] \triangleq \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{13} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} \\ \mathbf{F}_{31} & \mathbf{F}_{32} & \mathbf{F}_{33} \end{bmatrix} \mathcal{E}_{n_r}[\mathbf{x}] = \mathbf{F} \mathcal{E}_{n_r}[\mathbf{x}]. \quad (2.26)$$

where  $\mathbf{F}_{11} \in \mathbb{R}^{n_q \times n_r}$ ,  $\mathbf{F}_{12}, \mathbf{F}_{13} \in \mathbb{R}^{n_q \times (N_i - n_r)}$ ,  $\mathbf{F}_{21}, \mathbf{F}_{31} \in \mathbb{R}^{(N_o - n_q) \times n_r}$ , and where  $\mathbf{F}_{\ell k} \in \mathbb{R}^{(N_o - n_q) \times (N_i - n_r)}$  with  $\ell, k = 2, 3$ . More specifically, the widely linear transformation from  $\mathbf{x}$  to  $\mathbf{y}$  can be written as:  $\mathbf{y} = \bar{\mathcal{E}}_{n_q}[\mathbf{F} \mathcal{E}_{n_r}[\mathbf{x}]]$ .

The linear transformation (2.26) can also be equivalently written as:

$$\mathcal{C}_{n_q}[\mathbf{y}] \triangleq \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \mathbf{G}_{12}^* \\ \mathbf{G}_{21} & \mathbf{G}_{22} & \mathbf{G}_{23} \\ \mathbf{G}_{21}^* & \mathbf{G}_{23}^* & \mathbf{G}_{22}^* \end{bmatrix} \mathcal{C}_{n_r}[\mathbf{x}] = \mathbf{G} \mathcal{C}_{n_r}[\mathbf{x}] \quad (2.27)$$

where  $\mathbf{G}_{11} = \mathbf{F}_{11}$ ,  $\mathbf{G}_{12} = \frac{1}{2}[\mathbf{F}_{12} - j\mathbf{F}_{13}]$ ,  $\mathbf{G}_{21} = \mathbf{F}_{21} + j\mathbf{F}_{31}$ ,  $\mathbf{G}_{22} = \frac{1}{2}[\mathbf{F}_{22} + \mathbf{F}_{33} + j(\mathbf{F}_{32} - \mathbf{F}_{23})]$ , and  $\mathbf{G}_{23} = \frac{1}{2}[\mathbf{F}_{22} - \mathbf{F}_{33} + j(\mathbf{F}_{32} + \mathbf{F}_{23})]$ , or, more compactly,

$$\mathbf{G} = \mathbf{\Upsilon} \mathbf{T} \mathbf{F} \mathbf{T}^H \mathbf{\Upsilon}^{-1}. \quad (2.28)$$

Then, the overall WL processing can also be written as:  $\mathbf{y} = \bar{\mathcal{C}}_{n_q}[\mathbf{G} \mathcal{C}_{n_r}[\mathbf{x}]]$ . In other words, when a matrix  $\mathbf{F}$  for WL processing in real-valued representation is available, then the matrix for the corresponding WL processing in complex-valued representation is  $\mathbf{G} = \mathbf{\Upsilon} \mathbf{T} \mathbf{F} \mathbf{T}^H \mathbf{\Upsilon}^{-1}$  and, *vice versa*,  $\mathbf{F} = \mathbf{T}^H \mathbf{\Upsilon}^{-1} \mathbf{G} \mathbf{\Upsilon} \mathbf{T}$  is the relation for the inverse transformation between the two representations.

The WL transformation in (2.26) (analogously, in (2.27)) becomes a strictly linear transformation in  $\mathbf{x}$  if  $\mathbf{F}_{12} = \mathbf{F}_{13} = \mathbf{0}$ ,  $\mathbf{F}_{22} = \mathbf{F}_{33}$  and  $\mathbf{F}_{32} = -\mathbf{F}_{23}$  (analogously,  $\mathbf{G}_{12} = \mathbf{G}_{23} = \mathbf{0}$ ); in such a case, in fact, the following equivalence holds:

$$\mathbf{y} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{0} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix} \mathbf{x} \Leftrightarrow \mathcal{E}_{n_q}[\mathbf{y}] = \mathbf{F} \mathcal{E}_{n_r}[\mathbf{x}] \Leftrightarrow \mathcal{C}_{n_q}[\mathbf{y}] = \mathbf{G} \mathcal{C}_{n_r}[\mathbf{x}]. \quad (2.29)$$

In the current literature, the two representations are both utilized since they are often equivalent for many application scenarios. However, it can be verified

that such an equivalence does not hold in general. Let us further discuss how it may happen that the choice of the representation can affect the WL processing of the vector  $\mathbf{x}$ . Let us denote with  $S_r^{(1)}$  and  $S_r^{(2)}$  two sets of matrices  $\mathbf{F}$  in the real-valued representation corresponding, by means of the transformation (2.28), to the sets  $S_c^{(1)}$  and  $S_c^{(2)}$  of matrices  $\mathbf{G}$  in the complex-valued representation, respectively. Assume that  $S_r^{(1)}$  satisfies a constraint on the structure of its matrix elements, and  $S_c^{(2)}$  satisfies a constraint on the structure of its elements. Then, if  $S_r^{(1)} \neq S_r^{(2)}$  and, therefore,  $S_c^{(1)} \neq S_c^{(2)}$ , the two representations are not equivalent. In fact, the choice of the real-valued representation implies to define a WL transformation  $\mathbf{F}$  in  $S_r^{(1)}$ , while choosing the complex-valued representation implies to define  $\mathbf{F}$  in  $S_r^{(2)}$ . Obviously, the two representations are equivalent only in the special case where  $\mathbf{F} \in (S_r^{(1)} \cap S_r^{(2)})$ , or, equivalently,  $\mathbf{G} \in (S_c^{(1)} \cap S_c^{(2)})$ .

For example, consider the case where both  $\mathbf{F}$  and  $\mathbf{G}$  are lower triangular matrices. According to the real-valued representation, the imaginary parts of the complex-valued components of  $\mathbf{y}$  are obtained by linearly combining (also) the real parts of the complex-valued components of  $\mathbf{x}$ . On the other hand, according to the complex-valued representation,  $\mathbf{y}$  is obtained by linearly processing  $\mathbf{x}$  being  $\mathbf{G}_{12} = \mathbf{0}$  and  $\mathbf{G}_{23} = \mathbf{0}$ .

### 2.3 System model

Let us consider the FIR baseband equivalent noisy communication channel in (1.25) with  $N_i$  jointly wide-sense stationary (WSS) transmitted signals and  $N_o$  received signals. Therefore, the  $\ell$ th output at the  $k$ th instant is given by

$$y_k^{(\ell)} = \sum_{m=0}^{\nu^{(\ell,i)}} \sum_{i=1}^{N_i} x_{k-m}^{(i)} h_m^{(\ell,i)} + n_k^{(\ell)} \quad 1 \leq \ell \leq N_o \quad (2.30)$$

where  $h_k^{(i,j)}$  ( $k = 0, \dots, \nu^{(i,j)}$ ) accounts for the effect of the  $i$ th input on the  $j$ th output. The noise signals  $n_k^{(j)}$  are assumed to be zero-mean jointly WSS complex-valued processes independent of the useful signals.

Each symbol  $x_k^{(i)}$  belongs to the constellation  $S_i$  ( $i = 1, \dots, N_i$ ): we consider both the two-dimensional constellations (e.g., MPSK with  $M \geq 4$  and QAM) and the one-dimensional constellations (e.g., PAM and its rotated versions) for which it exists  $\theta_i \in [0, 2\pi]$  such that  $e^{-j\theta_i} x_k^{(i)}$  is a real-valued sequence. Since in (2.30)  $h_m^{(i,j)} x_{k-m}^{(i)} = [e^{j\theta_i} h_m^{(i,j)}][e^{-j\theta_i} x_{k-m}^{(i)}]$ , the case  $\theta_i \neq 0$

is easily transformed into an equivalent problem with  $\theta_i = 0$  by substituting, for the considered value of  $i$  and for all values of  $j$  and  $k$ ,  $[x_{k-m}^{(i)}]$  and  $[h_m^{(i,j)}]$  with  $[e^{-j\theta_i} x_{k-m}^{(i)}]$  and  $[e^{j\theta_i} h_m^{(i,j)}]$ , respectively. Note also that two-dimensional constellations as OQPSK and OQAM can be described by two one-dimensional constellations after a suitable sampling [43]. For the above reasons, there is no loss of generality if one assumes that each one-dimensional constellation is real-valued. We order the symbol sequences so that the real-valued constellations have indices  $i \in \{1, \dots, n_r\}$ . The case  $n_r = 0$  accounts for the absence of real-valued constellations whereas  $n_r = N_i$  is the case corresponding to all real-valued constellations.

As shown in Section 1.3, the input-output relationship (2.30) can be expressed using a matrix notation as follows:

$$\mathbf{y}_k = \sum_{m=0}^{\nu} \mathbf{H}_m \mathbf{x}_{k-m} + \mathbf{n}_k \quad (2.31)$$

where  $\nu = \max_{(\ell,i)} \nu^{(\ell,i)}$ ,  $\mathbf{H}_m$  is the  $N_o \times N_i$  matrix whose  $(\ell, i)$  entry is  $h_m^{(\ell,i)}$ , and  $\mathbf{n}_k$  is the  $N_o \times 1$  vector of noise samples at the  $k$ th instant. ■

### Useful definitions

Let us define the set of operators  $\mathbf{T}_N[\cdot, \ell]$  that associate with the sequence of  $n_1 \times n_2$  arrays  $\mathbf{u}_k$  the following sequence of  $n_1 N \times n_2$  arrays:  $\mathbf{T}_N[\mathbf{u}_{k-k_0}, \ell] = [\mathbf{u}_{\ell-k_0}^T \ \mathbf{u}_{\ell-1-k_0}^T \ \dots \ \mathbf{u}_{\ell-N+1-k_0}^T]^T$  where  $\ell$  denotes the observation instant, and  $k_0$  denotes a time-shift. It is also useful to define the correspondence that associates to a sequence of  $n_1 \times n_2$  arrays  $\mathbf{u}_k$  the  $n_1 M \times n_2 N$  matrix  $\tilde{\mathbf{T}}[\mathbf{u}_k, M, N]$  defined as follows

$$\tilde{\mathbf{T}}[\mathbf{u}_k, M, N] \triangleq [\mathbf{T}_M[\mathbf{u}_k, 0] \ \mathbf{T}_M[\mathbf{u}_k, 1] \ \dots \ \mathbf{T}_M[\mathbf{u}_k, N-1]]. \quad (2.32)$$

The operator  $\mathbf{T}_N[\cdot, \ell]$ , when applied to the system input, is useful to represent the output of a linear and time-invariant FIR system, while the operator  $\tilde{\mathbf{T}}[\cdot, M, N]$  applied to the auto-correlation function of a stationary array sequence  $\mathbf{u}_k$  constructs the auto-correlation matrix of  $\mathbf{T}_N[\mathbf{u}_k, k]$ , namely

$$E[\mathbf{T}_M[\mathbf{u}_k, k] \mathbf{T}_N^H[\mathbf{u}_k, k]] = \tilde{\mathbf{T}}[E[\mathbf{u}_k \mathbf{u}_{k-m}^H], M, N] \quad (2.33)$$

## 2.4 DF-MMSE FIR equalizer

In this section, the decision-feedback (DF) FIR equalizer, whose structure has been presented in subsection 1.7.4, is derived according to the MMSE criterion. By utilizing the notation introduced in this chapter, the equalizer output (1.83) can be written as follows:

$$\begin{aligned} \hat{\mathbf{x}}_{k-\Delta} = & \underbrace{\left[ \mathbf{W}_0^H \ \mathbf{W}_1^H \ \dots \ \mathbf{W}_{N_f-1}^H \right]}_{\triangleq \mathbf{W}^H} \mathbf{T}_{N_f}[\mathbf{y}_k, k] \\ & - \underbrace{\left[ (\mathbf{B}_0^H - \mathbf{I}_{N_i}) \ (\mathbf{B}_1^H) \ \dots \ (\mathbf{B}_{N_b}^H) \right]}_{\triangleq \mathbf{B}^H - [\mathbf{I}_{N_i} \ \mathbf{0}_{N_i \times N_i N_b}]} \mathcal{Q}(\mathbf{T}_{N_b+1}[\hat{\mathbf{x}}_k, k - \Delta]) \end{aligned} \quad (2.34)$$

with the integer  $\Delta$  denoting the decision delay, and with  $\mathbf{W}_i$  and  $\mathbf{B}_i$  denoting the matrix taps of size  $N_o \times N_i$  and  $N_i \times N_i$ , respectively. The decision mechanism  $\mathcal{Q}(\cdot)$  is applied to each component of the vector  $\mathbf{T}_{N_b+1}[\hat{\mathbf{x}}_k, k - \Delta]$  to provide an estimate of  $\mathbf{T}_{N_b+1}[\mathbf{x}_k, k - \Delta]$ . According to the MMSE criterion, the optimum FIR filters  $\mathbf{W}$  and  $\mathbf{B}$  in (2.34), say  $\mathbf{W}^{(opt)}$  and  $\mathbf{B}^{(opt)}$ , are derived by minimizing the MSE cost function

$$\min_{\mathbf{W}, \mathbf{B}} E \left[ \|\hat{\mathbf{x}}_{k-\Delta} - \mathbf{x}_{k-\Delta}\|^2 \right] \quad (2.35)$$

and by accounting for the constraints imposed by the considered detection scenario (see subsection 1.7.4). To this aim, let us remind that the filter optimization in Scenario 1 requires to take into account for the constraint  $\mathbf{B}_0 = \mathbf{I}_{N_i}$ , whereas, in Scenario 2,  $\mathbf{B}_0^H$  is constrained to be monic lower triangular. In Scenario 3,  $\mathbf{B}_0^H$  is monic.

In the following, the DF-MMSE equalizer is derived by utilizing the optimization procedure proposed in [9], under the assumption that the available estimates be correct. The procedure is based on the knowledge of the channel impulse response  $\mathbf{H}_m$  ( $m = 0, \dots, \nu$ ) and the matrices  $\mathbf{R}_x(m) \triangleq E[\mathbf{x}_k \mathbf{x}_{k-m}^H]$  and  $\mathbf{R}_n(m) \triangleq E[\mathbf{n}_k \mathbf{n}_{k-m}^H]$ , which describe the spatial and temporal correlation of both the transmitted symbols and the noise, respectively. Such an assumption is reasonable since they can be estimated blindly or by means of a training-based method.

### 2.4.1 DF-MMSE for Scenario 1

The optimum equalizer estimates  $\mathbf{x}_{k-\Delta}$  by resorting to the past decisions, namely it does not exploit also the current decisions. Accounting for (2.32),

let us define the following matrices:

$$\begin{aligned}\mathbf{R}_x &\triangleq \tilde{\mathbf{T}}[\mathbf{R}_x(m), N_f + \nu, N_f + \nu] \\ \mathbf{R}_n &\triangleq \tilde{\mathbf{T}}[\mathbf{R}_n(m), N_f, N_f] \\ \mathbf{H} &\triangleq \tilde{\mathbf{T}}[\mathbf{H}_m, N_f, N_f + \nu]\end{aligned}\quad (2.36)$$

where  $\mathbf{H}_m = \mathbf{0}$  for  $m \notin \{0, \dots, \nu\}$ . Under the assumption that the available estimates be correct, i.e.,  $\mathcal{Q}(\mathbf{T}_{N_b+1}[\hat{\mathbf{x}}_k, k - \Delta]) = \mathbf{T}_{N_b+1}[\mathbf{x}_k, k - \Delta]$ , and that the matrices  $\mathbf{R}_x$ ,  $\mathbf{R}_n$ , and  $\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{R}_n$  be invertible, the optimum  $\mathbf{W}^{(opt)}$  and  $\mathbf{B}^{(opt)}$  are expressed as follows:

#### Scenario 1

$$\begin{cases} \mathbf{W}^{(opt)} = \mathbf{W}_0 \tilde{\mathbf{B}}^{(opt)} \\ \tilde{\mathbf{B}}^{(opt)} \triangleq \begin{bmatrix} \mathbf{0}_{N_i \times N_i \Delta} & \mathbf{B}^{(opt)H} \end{bmatrix}^H \\ \mathbf{B}^{(opt)} = \mathbf{f}_1[\mathbf{R}_x, \mathbf{R}_n, \mathbf{H}] \\ \mathbf{W}_0 \triangleq [\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{R}_n]^{-1} \mathbf{H}\mathbf{R}_x \end{cases} \quad (2.37)$$

The optimum feedback matrix  $\mathbf{B}^{(opt)}$  in (2.37) can be determined as follows:

1. determine  $\mathbf{R} = \mathbf{R}_x^{-1} + \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}$ ;
2. extract from  $\mathbf{R}$  the matrix  $\mathbf{R}_\alpha$  that contains its first  $N_i(\Delta + 1)$  columns and the matrix  $\mathbf{R}_\beta$  that contains the upper-left square sub-block of size  $N_i(\Delta + 1)$ ;
3. determine  $\mathbf{B}^{(opt)}$  as the lower right sub-block of size  $N_i(N_b + 1) \times N_i$  of  $\mathbf{R}_\alpha \mathbf{R}_\beta^{-1}$ .

The corresponding error correlation matrix and achieved MSE are given by

$$\mathbf{R}_e^{(opt,1)} = \Psi^H \mathbf{R}_\beta^{-1} \Psi \quad (2.38)$$

$$MMSE_1 = \text{trace}(\mathbf{R}_e^{(opt,1)}) \quad (2.39)$$

with  $\Psi^H \triangleq [\mathbf{0}_{N_i \times N_i \Delta} \quad \mathbf{I}_{N_i}]$ . Note that the decision delay  $\Delta$  is a parameter to be optimized in the method (in [9] the optimization has been carried out by an exhaustive procedure) and, together with the parameter  $N_f$ , it determines the number of causal taps and the number of anticausal taps in the linear FIR filter  $\mathbf{W}$ . Also the number of taps  $N_b$  of the feedback filter can be optimized; here, however, in accordance with [9], we assume  $N_b = N_f + \nu - \Delta - 1$ .

It can be readily checked that the optimum linear (L) MMSE equalizer, the one which resorts to the only feedforward filter to provide an estimation of  $\mathbf{x}_{k-\Delta}$ , is given by the first equation in (2.37) setting  $\mathbf{B}_k$  in (2.34) equal to  $\mathbf{I}_{N_i}\delta_k$ , i.e.  $\mathbf{B} = [\mathbf{I}_{N_i} \ \mathbf{0}_{N_i \times N_i N_b}]^H$ .

### 2.4.2 DF-MMSE for Scenario 2

The optimum equalizer estimates  $x_{k-\Delta}^{(\ell)}$  by resorting to the past decisions and also the current decisions  $\mathcal{Q}(\hat{x}_{k-\Delta}^{(1)}), \mathcal{Q}(\hat{x}_{k-\Delta}^{(2)}), \dots, \mathcal{Q}(\hat{x}_{k-\Delta}^{(\ell-1)})$  with lower indexed components, namely the decisions are taken sequentially starting with the lower indexed components. It follows that the optimization procedure over the matrix  $\mathbf{W}$  and  $\mathbf{B}$  has to be carried by accounting for the monic lower triangular matrix tap  $\mathbf{B}_0^H$ . The optimum DF-MMSE equalizers given by:

$$\begin{array}{l} \text{Scenario 2} \\ \left\{ \begin{array}{l} \mathbf{W}_2^{(opt)} = \mathbf{W}_0 \tilde{\mathbf{B}}_2^{(opt)} \\ \mathbf{B}_2^{(opt)} = \mathbf{B}^{(opt)} \mathbf{B}_0^{(2)} \\ \mathbf{B}_0^{(2)} = \mathbf{f}_2 [\mathbf{R}_x, \mathbf{R}_n, \mathbf{H}] \end{array} \right. \end{array} \quad (2.40)$$

where  $\mathbf{W}_0$  and  $\mathbf{B}^{(opt)}$  have been defined in (2.37) and where  $\tilde{\mathbf{B}}_2^{(opt)} \triangleq [\mathbf{0}_{N_i \times N_i \Delta} \ \mathbf{B}_2^{(opt)H}]^H$ . The optimum feedback matrix  $\mathbf{B}_0^{(2)}$  in (2.40) can be determined as follows:

1. extract from  $\mathbf{R}$  the upper-left square sub-block  $\mathbf{R}_\beta$  of size  $N_i(\Delta + 1)$  and determine  $\mathbf{R}_b = \mathbf{R}_\beta^{-1}$ ;
2. extract from  $\mathbf{R}_b$  the lower-right square sub-block  $\mathbf{R}_c \triangleq \Psi^H \mathbf{R}_\beta^{-1} \Psi$  of size  $N_i$ ;
3. determine its monic lower triangular Cholesky factor  $\mathbf{L}_c$ , i.e.,  $\mathbf{R}_c = \mathbf{L}_c \mathbf{D} \mathbf{L}_c^H$  where  $\mathbf{D}$  is a diagonal matrix with positive entries and  $\mathbf{L}_c$  has unit diagonal entries;
4. determine the optimum  $\mathbf{B}_0^{(2)} = (\mathbf{L}_c^H)^{-1}$ .

The corresponding error correlation matrix and achieved MSE are given by

$$\mathbf{R}_e^{(opt,2)} = \mathbf{D} \quad (2.41)$$

$$MMSE_2 = \text{trace}(\mathbf{R}_e^{(opt,2)}) \quad (2.42)$$

Note that, being  $\mathbf{D}$  diagonal, the error vector measured at the output of the equalizer has uncorrelated components, i.e., it is spatially white. ■

The procedures in (2.37) and (2.40) to derive the DF-MMSE equalizer exhibit a computational complexity which grows cubically with the system parameters. In presence of spatially and temporally uncorrelated input and noise sequences, (2.37) and (2.40) can be solved (see [9] for details) by means of simpler procedure based on fast factorization algorithms [47, 48] with complexity  $O(N_i^2(N_f + \nu)^2)$ . A procedure alternative to the one in [9] is proposed in Appendix B.

### 2.4.3 DF-MMSE for Scenario 3

For the sake of completeness, we report the DF-MMSE equalizer structure when both past and current decision are available from a previous detection stage. However, in the rest of the thesis, we will not consider anymore such a scenario which deals with the multistage detection.

By accounting for the monic matrix tap  $\mathbf{B}_0$ , the optimum DF-MMSE equalizer is given by:

$$\begin{aligned} & \text{Scenario 3} \\ & \begin{cases} \mathbf{W}_3^{(opt)} = \mathbf{W}_0 \tilde{\mathbf{B}}_3^{(opt)} \\ \mathbf{B}_3^{(opt)} = \mathbf{B}^{(opt)} \mathbf{B}_0^{(3)} \\ \mathbf{B}_0^{(3)} = \mathbf{f}_3 [\mathbf{R}_x, \mathbf{R}_n, \mathbf{H}] \end{cases} \end{aligned} \quad (2.43)$$

where  $\mathbf{W}_0$  and  $\mathbf{B}^{(opt)}$  have been defined in (2.37), and where  $\tilde{\mathbf{B}}_3^{(opt)} \triangleq \begin{bmatrix} \mathbf{0}_{N_i \times N_i \Delta} & \mathbf{B}_3^{(opt)H} \end{bmatrix}^H$ . The optimum feedback matrix  $\mathbf{B}_0^{(3)}$  in (2.43) is determined as follows:

1. according to the steps 1. and 2. of the procedure which provides the optimum DF-MMSE equalizer in Scenario 2, determine  $\mathbf{R}_c$
2. the  $(i, j)$  entry of  $\mathbf{B}_0^{(3)}$  is equal to

$$\left[ \mathbf{B}_0^{(3)} \right]_{(i,j)} = \frac{[\mathbf{R}_d]_{(i,j)}}{[\mathbf{R}_d]_{(j,j)}} \quad (2.44)$$

where  $[\mathbf{R}_d]_{(i,j)}$  denotes the  $(i, j)$  entry of  $\mathbf{R}_d = \mathbf{R}_c^{-1}$ .

The corresponding MSE is equal to

$$MMSE_3 = \sum_{j=1}^{N_i} \frac{1}{[\mathbf{R}_d]_{(j,j)}} . \quad (2.45)$$

#### 2.4.4 Detection ordering

The order of detection in MIMO DF based equalization is a valuable degree of freedom with no counterpart in SISO channels. Note that, while Scenario 1 the decision over the symbols  $x_{k-\Delta}^{(\ell)}$  are taken simultaneously, in Scenario 2 the decisions are taken sequentially according to a chosen order. Changing such an ordering leads to a modification of the equivalent discrete-time channel response to be equalized and, consequently, to different performances. As well known, the optimum detection ordering is NP-hard [31]: in fact,  $N_i!$  DF equalizers exist, one for every order in which the  $N_i$  inputs can be detected. For such a reason, suboptimal algorithms that provide reliable performances with polynomial complexity have been widely studied. Among them, the vertical BLAST (V-BLAST) algorithm<sup>2</sup> ensures significantly performance improvement over the conventional linear equalizer with only  $O(N_i^2)$  complexity [24, 49]. Unfortunately, the issue of the decision ordering has been widely studied in DF equalization over non-dispersive MIMO channel ([24, 49, 50] and references therein) but only few contributions [51, 52] have started its discussion over dispersive MIMO channel.

In this subsection, with reference to the Scenario 2, we prove the dependence of the DF-MMSE equalizer (and of its performances) on the decision ordering by rewriting the system model (2.31) as a function of a permuted input vector  $\mathbf{x}_k$ . Moreover, we show that the optimum equalizer in Scenario 2 can be expressed as a two-stage equalizer: the former is the optimum time-dispersive equalizer of Scenario 1 (which is independent of the decision ordering), the latter performs a non-dispersive equalization and the optimization over the decision ordering. Such a *framework* allows us to extend any suboptimum ordering method proposed for non-dispersive environment to the dispersive one.

Let  $\mathbf{P}$  be a permutation matrix of size  $N_i$ , such that  $\mathbf{P}^T \mathbf{P} = \mathbf{I}_{N_i}$ . By defining the (row) permuted input vector  $\mathbf{x}_k(P) \triangleq \mathbf{P} \mathbf{x}_k$  and, hence, the (column) permuted channel matrix  $\mathbf{H}_m(P) \triangleq \mathbf{H}_m \mathbf{P}^T$ , the channel output

---

<sup>2</sup>BLAST: Bell Laboratories Layered Space-Time

can be equivalently re-written as follows:

$$\mathbf{y}_k = \sum_{m=0}^{\nu} \mathbf{H}_m(P) \mathbf{x}_{k-m}(P) + \mathbf{n}_k. \quad (2.46)$$

In the following the  $\mathbf{P}$ -dependence is omitted for the case  $\mathbf{P} = \mathbf{I}_{N_i}$  (e.g.,  $\mathbf{H}_m(\mathbf{I}_{N_i}) \triangleq \mathbf{H}_m$ ), which corresponds to the input ordering in (2.31). With reference to channel model (2.46), we define the input-correlation matrix

$$\mathbf{R}_x(m, P) \triangleq E [\mathbf{x}_k(P) \mathbf{x}_{k-m}(P)^H] = \mathbf{P} \mathbf{R}_x(m) \mathbf{P}^T. \quad (2.47)$$

For any given permutation  $\mathbf{P}$ , the output correlation matrix  $\mathbf{R}_y(m, P) \triangleq E [\mathbf{y}_k(P) \mathbf{y}_{k-m}^H(P)]$  can be shown to be independent of  $P$ , in fact:

$$\begin{aligned} \mathbf{R}_y(m, P) &= \sum_{m_1=0}^{\nu} \sum_{m_2=0}^{\nu} \mathbf{H}_{m_1}(P) \mathbf{R}_x(m + m_2 - m_1, P) \mathbf{H}_{m_2}^H(P) + \mathbf{R}_n(m) \\ &= \sum_{m_1=0}^{\nu} \sum_{m_2=0}^{\nu} \mathbf{H}_{m_1} \mathbf{R}_x(m + m_2 - m_1) \mathbf{H}_{m_2}^H + \mathbf{R}_n(m) \\ &= \mathbf{R}_y(m) \end{aligned}$$

By replacing the matrices  $\mathbf{R}_x(m)$  and  $\mathbf{H}_m$  with  $\mathbf{R}_x(m, P)$  and  $\mathbf{H}_m(P)$ , the optimization procedure (2.37) can be utilized to determine DF-MMSE equalizer for each one of the  $N_i!$  permutation matrices; the optimum matrix filters and the achieved MMSE are denoted with  $\mathbf{W}_i^{(opt)}(P)$ ,  $\mathbf{B}_i^{(opt)}(P)$  ( $\tilde{\mathbf{B}}_i^{(opt)}(P)$ ), and  $MMSE_i(P)$ , where  $i$  accounts for the detection scenario and  $\mathbf{P}$ -dependence is introduced.

By resorting to the property  $trace(\mathbf{A}) = trace(\mathbf{X}^{-1} \mathbf{A} \mathbf{X})$ , with  $\mathbf{X}$  being a unitary matrix, it can be verified that, for any given  $\mathbf{P}$ , the achieved  $MMSE_1(P)$  in Scenario 1 is such that  $MMSE_1(P) = MMSE_1(\mathbf{I}) \forall \mathbf{P}$ , i.e., the performances of DF-based equalizers are invariant to the decision ordering. On the other hand, if Scenario 2 is considered, the permutation matrix  $\mathbf{P}$  greatly affects the  $MMSE_2(P)$  and, therefore,  $N_i!$  different DF-MMSE equalizers exist. Unfortunately, as previously stated, the optimum decision ordering can be derived only by an exhaustive procedure. For the sake of brevity, in the sequel, if not specified, the problem of input ordering will refer to Scenario 2.

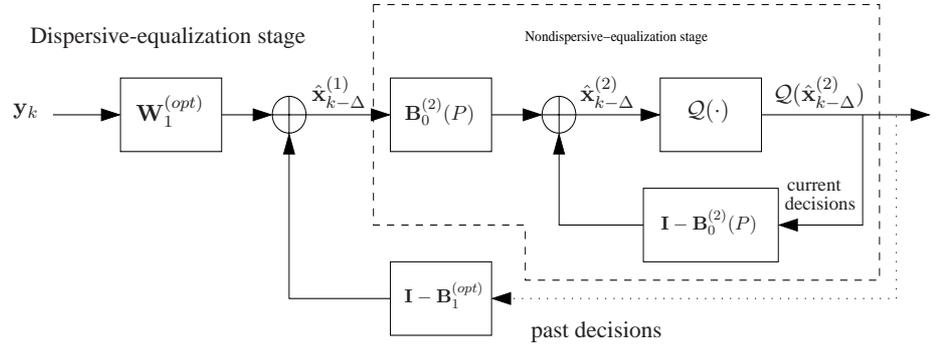
For a given detection ordering  $\mathbf{P}$ , denote with  $\hat{\mathbf{x}}_{k-\Delta}^{(i)}(P)$  the output of the DF-MMSE equalizer when the  $i$ th scenario is considered. Assuming correct

decision, the output vector  $\hat{\mathbf{x}}_{k-\Delta}^{(2)}(P) = \mathbf{P}\hat{\mathbf{x}}_{k-\Delta}^{(2)}$  can be rewritten as follows:

$$\begin{aligned}\hat{\mathbf{x}}_{k-\Delta}^{(2)}(P) &= \mathbf{W}_2^{(opt)H}(P)\mathbf{T}_{N_f}[\mathbf{y}_k, k] + \mathbf{P}\mathbf{x}_{k-\Delta} - \tilde{\mathbf{B}}_2^{(opt)H}(P)\mathbf{T}_{N_f+\nu}[\mathbf{x}_k(P), k] \\ &= \mathbf{B}_0^{(2)H}(P)\left[\mathbf{W}_1^{(opt)H}(P)\mathbf{T}_{N_f}[\mathbf{y}_k, k] - \tilde{\mathbf{B}}_1^{(opt)H}(P)\mathbf{T}_{N_f+\nu}[\mathbf{x}_k(P), k]\right] \\ &\quad + \mathbf{x}_{k-\Delta}(P) \\ &= \mathbf{B}_0^{(2)H}(P)\mathbf{P}\left[\underbrace{(\mathbf{P}^T\mathbf{W}_1^{(opt)H}(P))}_{=\mathbf{W}_1^{(opt)H}}\mathbf{T}_{N_f}[\mathbf{y}_k, k] \right. \\ &\quad \left. - \underbrace{(\mathbf{P}^T\tilde{\mathbf{B}}_1^{(opt)H}(P)\tilde{\mathbf{P}}_{N_b+1})}_{=\tilde{\mathbf{B}}_1^{(opt)H}}\mathbf{T}_{N_f+\nu}[\mathbf{x}_k, k]\right] + \mathbf{P}\mathbf{x}_{k-\Delta}\end{aligned}$$

$$\hat{\mathbf{x}}_{k-\Delta}^{(2)}(P) = \mathbf{B}_0^{(2)H}(P)\mathbf{P}\hat{\mathbf{x}}_{k-\Delta}^{(1)} + [\mathbf{I}_{N_i} - \mathbf{B}_0^{(2)H}(P)]\mathbf{P}\mathbf{x}_{k-\Delta} \quad (2.48)$$

i.e., it can be expressed as the output of a two-stage equalizer (see Fig. 2.1): the output of the former stage coincides with  $\hat{\mathbf{x}}_{k-\Delta}^{(1)}$ , which is also the final output in Scenario 1 (only past decisions are fed back) and is unaffected by a possible ordering; the latter stage, instead, performs a non-dispersive DF-MMSE equalization depending on the permutation matrix  $\mathbf{P}$ .



**Figure 2.1:** Two stage equalizer structure: the former stage performs the MMSE equalization by utilizing only past decisions (Scenario 1), the latter performs the MMSE equalization by utilizing only current decisions (Scenario 2).

Note that (2.48) allows us to simply express the error vector  $\mathbf{e}_{k-\Delta}^{(2)}(P) \triangleq \hat{\mathbf{x}}_{x-\Delta}^{(2)} - \mathbf{x}_{x-\Delta}$  at the output the second stage as follows:

$$\begin{aligned} \mathbf{e}_{k-\Delta}^{(2)} &= \mathbf{B}_0^{(2)H}(P)\mathbf{P}\left[\hat{\mathbf{x}}_{k-\Delta}^{(1)} - \mathbf{x}_{k-\Delta}\right] \\ &\triangleq \mathbf{B}_0^{(2)H}(P)\mathbf{P}\mathbf{e}_{k-\Delta}^{(1)}. \end{aligned} \quad (2.49)$$

where  $\mathbf{e}_{k-\Delta}^{(1)}$  denotes the error vector measured at the output of the first stage and its correlation matrix is equal to  $\mathbf{R}_e^{(opt,1)}$  in (2.38). From (2.40) and (2.49), it is clear that the optimum second stage performs the spatial whitening (see also [53] and [54]) of the error vector  $\mathbf{e}_{k-\Delta}^{(1)}$  by utilizing the monic upper triangular filter  $\mathbf{B}_0^{(2)}(P)$ , which is constrained to be upper triangular with unit diagonal entries.

The decision ordering optimization in a time-dispersive environment has been, therefore, recognized to be equivalent to the decision ordering optimization in such a latter stage. Such a result allows us to extend every suboptimum ordering method, already proposed for DF equalization of non-dispersive MIMO channel, to the MIMO dispersive environment. For the sake of completeness, we point out that:

- the feedforward filtering of the noise and the pre-cursor ISI makes the resulting noise at the output of the first stage spatially correlated;
- the derivation of the second stage needs to take into account for input and noise signals correlated with each other.

Consequently, when designing the second stage, the DF-MMSE procedure in (2.40) has to be modified to take into account for possibly correlated (with each other) input and noise vectors.

## 2.5 WL decision-feedback FIR equalizer

In this section, with reference to the MIMO dispersive channel equalization, we introduce the structure of the widely-linear/widely-decision feedback (WL-WDF) FIR equalizer, i.e., the DF-based equalizer which utilizes both widely linear feedforward filter and a widely linear feedback filter. Owing to the input-output channel model and to the equalizer structure proposed here, the WL-WDF-MMSE in Scenario 1 and Scenario 2 will be carried out (in Section 2.6 and in Section 2.7, respectively) by utilizing the existing procedure

for DF-MMSE equalization. As discussed in Section 2.2, the WL filtering can be performed by utilizing the real-valued representation of the involved vectors, as well as the complex-valued one: it follows that two WL-WDF equalizer structures can be defined according to each representation. When Scenario 1 is considered, the two representations are shown to be equivalent [44], i.e., a one-to-one correspondence (by means of (2.28)) exists between the two structures and, therefore, for any chosen optimization criterion, they perform equivalently. On the other hand, in Scenario 2, the correspondence between the two representations no longer holds. In fact, when the equalizer obtained by a real-valued representation satisfies the constraint imposed by the scenario, the corresponding equalizer provided by (2.28) may not satisfy the constraint in complex-valued representation, and *vice versa*. The problem of the non-equivalence between the two approaches will be addressed in Section 2.8. Here, we adopt the real-valued representation to describe the main properties of the DF-based equalization combined with the WL filtering. Nevertheless, we point out that the same mathematical framework is required when the complex-valued representation is adopted.

We resort to the operators (2.16) and (2.18) and their properties (2.20)-(2.22) to replace the channel model (2.31) with the following equivalent one:

$$\mathcal{E}_0[\mathbf{y}_k] = \sum_{m=0}^{\nu} \tilde{\mathcal{E}}_{n_r}[\mathbf{H}_m] \mathcal{E}_{n_r}[\mathbf{x}_{k-m}] + \mathcal{E}_0[\mathbf{n}_k]. \quad (2.50)$$

Note that the above input-output relationship is written in terms of the augmented input vector  $\mathcal{E}_{n_r}[\mathbf{x}_k]$  whose first  $n_r$  components are the real-valued transmitted symbols  $x_k^{(\ell)}$  with  $\ell = 1, \dots, n_r$ , and whose last  $2(N_i - n_r)$  components are the real and the imaginary parts of the complex-valued transmitted symbols  $x_k^{(\ell)}$  with  $\ell = n_r + 1, \dots, N_i$ . The output of the DF equalizer operating on the channel model (2.50) can be written as:

$$\mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}] = \mathbf{W}_E^H \mathbf{T}_{N_f} [\mathcal{E}_0[\mathbf{y}_k], k] - \mathbf{B}_E^H \mathcal{Q}(\mathbf{T}_{N_b+1}[\mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}, k - \Delta]]) \quad (2.51)$$

where  $\mathbf{W}_E$  and  $\mathbf{B}_E$  are matrices of size  $(2N_o N_f) \times (2N_i - n_r)$  and  $(2N_i - n_r)(N_b + 1) \times (2N_i - n_r)$ , respectively, and where the decision mechanism is applied to each component of the augmented vector  $\mathbf{T}_{N_b+1}[\mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}, k - \Delta]]$ . Let us note that the first tap of the filter matrix  $\mathbf{B}_E$ , say  $\mathbf{B}_{E,0}$ , is subject to the constraint imposed by the equalization scenario: in other words it represents the WL counterpart of the matrix taps  $\mathbf{B}_0$  in (2.34). The operator  $\tilde{\mathcal{E}}_{n_r}[\cdot]$  allows us to obtain from (2.51) the output

$\hat{\mathbf{x}}_{E,k-\Delta}$  of the WL-WDF equalizer, namely:

$$\hat{\mathbf{x}}_{E,k-\Delta} = \bar{\mathcal{E}}_{n_r} \left[ \mathbf{W}_E^H \mathbf{T}_{N_f} [\mathcal{E}_0[\mathbf{y}_k], k] - \mathbf{B}_E^H \mathcal{Q}(\mathbf{T}_{N_b+1}[\mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}, k-\Delta]]) \right]. \quad (2.52)$$

Relation (2.52) shows that the output  $\hat{\mathbf{x}}_{E,k-\Delta}$  can be obtained by the sum of two contributes: the former is obtained by a feedforward-based processing whereas the latter by a feedback-based one.

In the particular case of  $\mathbf{B}_E = \mathbf{0}$  in (2.52), the subset of the WL-WDF equalizers, usually called the class of WL equalizers [42], is recognized. In fact, accounting for  $\mathbf{B}_E = \mathbf{0}$ , (2.52) becomes:

$$\begin{aligned} \hat{\mathbf{x}}_{E,k-\Delta} &= \bar{\mathcal{E}}_{n_r} \left[ \mathbf{W}_E^H \mathbf{T}_{N_f} [\mathcal{E}_0[\mathbf{y}_k], k] \right] \\ &\triangleq \left[ \mathbf{W}_\alpha^H \quad \mathbf{G}_\alpha^H \right] \mathcal{E}_0 \left[ \mathbf{T}_{N_f} [\mathbf{y}_k, k] \right] \end{aligned} \quad (2.53)$$

where  $\left[ \mathbf{W}_\alpha^H \quad \mathbf{G}_\alpha^H \right] \triangleq \bar{\mathcal{E}}_{n_r} \left[ \mathbf{W}_E^H \right] \mathbf{A}^H$ , with  $\mathbf{W}_\alpha$  and  $\mathbf{G}_\alpha$  of size  $N_o(N_f + \nu) \times N_i$ . The matrices  $\mathbf{W}_\alpha$  and  $\mathbf{G}_\alpha$  represent the impulse responses of the FIR feedforward filters which process the real part of the received vector and the imaginary one, respectively. Since the filters  $\mathbf{W}_\alpha$  and  $\mathbf{G}_\alpha$  are independent of each other, the resulting processing of the received signals is widely linear rather than linear.

As regards to the more general case of  $\mathbf{B}_E \neq \mathbf{0}$ , by defining the real-valued sequence  $\mathbf{x}_k^{(r)} \triangleq \mathbf{x}_k(1 : n_r)$ , the complex-valued sequence  $\mathbf{x}_k^{(c)} = \mathbf{x}_k(n_r + 1 : N_i)$ , and the permutation matrix  $\mathbf{A}_r$  such that

$$\mathbf{A}_r \mathbf{T}_{N_b+1} [\mathcal{E}_{n_r}[\mathbf{x}_k], k - \Delta] = \begin{bmatrix} \mathbf{T}_{N_b+1}[\mathbf{x}_k^{(r)}, k - \Delta] \\ \mathcal{E}_0 \left[ \mathbf{T}_{N_b+1}[\mathbf{x}_k^{(c)}, k - \Delta] \right] \end{bmatrix},$$

one has that the feedback-based estimate component in (2.52) can be re-written as follows:

$$\bar{\mathcal{E}}_{n_r} \left[ \mathbf{B}_E^H \hat{\mathbf{x}}_{E,k-\Delta} \right] \triangleq \begin{bmatrix} \mathbf{B}_\alpha \\ \mathbf{D}_\alpha \end{bmatrix}^H \mathcal{Q} \left( \begin{bmatrix} \mathbf{T}_{N_b+1}[\hat{\mathbf{x}}_{E,k}^{(r)}, k - \Delta] \\ \mathcal{E}_0 \left[ \mathbf{T}_{N_b+1}[\hat{\mathbf{x}}_{E,k}^{(c)}, k - \Delta] \right] \end{bmatrix} \right) \quad (2.54)$$

where  $\hat{\mathbf{x}}_{E,k}^{(r)}$  and  $\hat{\mathbf{x}}_{E,k}^{(c)}$  are the estimates of  $\mathbf{x}_k^{(r)}$  and  $\mathbf{x}_k^{(c)}$ , respectively, and where  $\left[ \mathbf{B}_\alpha^H \quad \mathbf{D}_\alpha^H \right] \triangleq \bar{\mathcal{E}}_{n_r} \left[ \mathbf{B}_E^H \right] \mathbf{A}_r^H$ , with  $\mathbf{B}_\alpha$  and  $\mathbf{D}_\alpha$  of size  $N_i(N_b + 1) \times N_i$  and  $(N_i - n_r)(N_b + 1) \times N_i$ , respectively. The matrices  $\mathbf{B}_\alpha$  and  $\mathbf{D}_\alpha$  represent the impulse responses of the FIR feedback filters that process the real part and the imaginary part of the past decisions, respectively. Analogously to the feedforward filters,  $\mathbf{B}_\alpha$  and  $\mathbf{D}_\alpha$  are independent of each other and, hence, the past decisions are processed by a widely linear filter.

## 2.6 WL-WDF-MMSE equalizer for Scenario 1

By adopting the MMSE criterion, and accounting for the (2.52), one can derive the optimum  $(\mathbf{W}_E^{(opt)}, \mathbf{B}_E^{(opt)})$  and, consequently, the optimum  $(\mathbf{W}_\alpha^{(opt)}, \mathbf{G}_\alpha^{(opt)})$  and  $(\mathbf{B}_\alpha^{(opt)}, \mathbf{D}_\alpha^{(opt)})$  in (2.53) and (2.54), by utilizing the procedure (2.37) with the parameters  $N_i$  and  $N_o$  replaced by  $2N_i - n_r$  and  $2N_o$ , respectively<sup>3</sup>, under the assumption of correct previous estimates. The matrices  $\mathbf{R}_x(m)$ ,  $\mathbf{R}_n(m)$ , and  $\mathbf{H}_m$  in (2.36) are replaced with  $\mathbf{R}_x^E(m) \triangleq E[\mathcal{E}_{n_r}[\mathbf{x}_k]\mathcal{E}_{n_r}[\mathbf{x}_{k-m}]^T]$ ,  $\mathbf{R}_n^E(m) \triangleq E[\mathcal{E}_0[\mathbf{n}_k]\mathcal{E}_0[\mathbf{n}_{k-m}]^T]$ , and  $\tilde{\mathcal{E}}_{n_r}[\mathbf{H}_m]$ , respectively, in order to single out  $(\mathbf{W}_E^{(opt)}, \mathbf{B}_E^{(opt)})$ . Consequently, the matrices  $\mathbf{R}_x$ ,  $\mathbf{R}_n$ , and  $\mathbf{H}$  in (2.37) are substituted with  $\mathbf{R}_x^E \triangleq \tilde{\mathbf{T}}[\mathbf{R}_x^E(m), N_f + \nu, N_f + \nu]$ ,  $\mathbf{R}_n^E \triangleq \tilde{\mathbf{T}}[\mathbf{R}_n^E(m), N_f, N_f]$ , and  $\mathbf{H}_E \triangleq \tilde{\mathbf{T}}[\tilde{\mathcal{E}}_{n_r}[\mathbf{H}_m], N_f, N_f + \nu]$ , respectively, and  $\mathbf{R}$  is replaced with  $\mathbf{R}_E = [(\mathbf{R}_x^E)^{-1} + \mathbf{H}_E^H (\mathbf{R}_n^E)^{-1} \mathbf{H}_E]$ .

Unlike the DF optimization, which is based only on the knowledge of  $\mathbf{R}_x(m)$ , the construction of the matrix  $\mathbf{R}_x^E(m)$  in WL-WDF optimization requires also the pseudo-correlation matrix  $\tilde{\mathbf{R}}_x(m) \triangleq E[\mathbf{x}_k \mathbf{x}_{k-m}^T]$ , or, equivalently, the three matrices  $E[\Re\{\mathbf{x}_k\}\Re\{\mathbf{x}_{k-m}\}^T]$ ,  $E[\Im\{\mathbf{x}_k\}\Im\{\mathbf{x}_{k-m}\}^T]$ , and  $E[\Re\{\mathbf{x}_k\}\Im\{\mathbf{x}_{k-m}\}^T]$  (see Section A.1 for details). Of course, the same reasoning applies to the noise statistics.

The application of the procedure (2.37) to the channel model (2.50), namely

$$\begin{bmatrix} \mathbf{B}_\alpha^{(opt)H} & \mathbf{D}_\alpha^{(opt)H} \end{bmatrix}^H = \mathbf{A}_r \bar{\mathcal{E}}_{n_r} [\mathbf{f}_1 [\mathbf{R}_x^E, \mathbf{R}_n^E, \mathbf{H}_E]] \quad (2.55)$$

implies the utilization of augmented matrices with a consequent increase in the computational complexity. In order to apply the procedure  $\mathbf{f}_1[\cdot, \cdot, \cdot]$ , the matrices  $\mathbf{R}_x^E$ ,  $\mathbf{R}_n^E$ , and  $\mathbf{H}_E \mathbf{R}_x^E \mathbf{H}_E^H + \mathbf{R}_n^E$  are obviously required to be non-singular. Note, however, that the possible introduction of the augmented vector  $\mathcal{E}_0[\mathbf{x}_k]$  in (2.50) (rather than  $\mathcal{E}_{n_r}[\mathbf{x}_k]$ , here utilized) implies that the matrix  $\mathbf{R}_x^E$  is singular when  $n_r \geq 1$ . In fact, if the feedback filter is interpreted as a feedforward filter which processes a delayed version of the transmitted symbols<sup>4</sup>, the couple  $(\mathbf{W}_E, \mathbf{B}_E)$  can be seen as a WL filter that processes

<sup>3</sup>According to (2.34), the feedback filter in (2.51) can be re-written as  $[\mathbf{B}_{E,0}^H - \mathbf{I}_{2N_i - n_r} \mathbf{B}_{E,1}^H \dots \mathbf{B}_{E,N_b}^H]^H$ , and, therefore, the utilization of (2.37) provides the optimum filter  $\mathbf{B}_E = [\mathbf{B}_{E,0}^H \mathbf{B}_{E,1}^H \dots \mathbf{B}_{E,N_b}^H]^H$  subject to the constraint  $\mathbf{B}_{E,0} = \mathbf{I}_{2N_i - n_r}$ .

<sup>4</sup>Such an interpretation is possible when no decision errors occur.

an observation vector with both complex-valued (i.e., the complex-valued past decisions  $\mathbf{x}_k^{(c)}$ ) and real-valued components (i.e., the real-valued past decisions  $\mathbf{x}_k^{(r)}$ ). It follows that, according to (2.16), the MMSE solution is not unique when  $\mathcal{E}_0[\mathbf{x}_k]$  in (2.50).

### 2.6.1 Reduced-order design of WL-WDF-MMSE equalizer

By exploiting the particular structures of the augmented matrices, we propose an equivalent procedure exhibiting a lower computational complexity than that of the procedure (2.55) which is based on the inversion of the augmented matrices. Moreover, such a simplified procedure provides also a better understanding of the optimum WL-WDF equalizer structure.

To this aim, define the  $n_r(N_f + \nu) \times 1$  real-valued vector  $\tilde{\mathbf{x}}_k^{(r)} \triangleq \mathbf{T}_{N_f+\nu} \left[ \mathbf{x}_k^{(r)}, k \right]$  and the  $(N_i - n_r)(N_f + \nu) \times 1$  vector  $\tilde{\mathbf{x}}_k^{(c)} \triangleq \mathbf{T}_{N_f+\nu} \left[ \mathbf{x}_k^{(c)}, k \right]$ , and re-order the matrices  $\mathbf{R}_x^E$ ,  $\mathbf{R}_n^E$ , and  $\mathbf{H}_E$  as follows

$$\mathbf{R}_x^A \triangleq \mathbf{A}_r \mathbf{R}_x^E \mathbf{A}_r^H = \begin{bmatrix} \mathbf{R}_{x^{(r)}} & \mathbf{R}_{x^{(rc)}} & \hat{\mathbf{R}}_{x^{(rc)}} \\ \mathbf{R}_{x^{(rc)}}^T & \mathbf{R}_{x^{(c)}} & \hat{\mathbf{R}}_{x^{(cc)}} \\ \hat{\mathbf{R}}_{x^{(rc)}}^T & \hat{\mathbf{R}}_{x^{(cc)}}^T & \hat{\mathbf{R}}_{x^{(c)}} \end{bmatrix} \quad (2.56)$$

$$\mathbf{R}_n^A \triangleq \mathbf{A} \mathbf{R}_n^E \mathbf{A}^H = \begin{bmatrix} \mathbf{R}_{n^{(c)}} & \hat{\mathbf{R}}_{n^{(cc)}} \\ \hat{\mathbf{R}}_{n^{(cc)}}^T & \hat{\mathbf{R}}_{n^{(c)}} \end{bmatrix} \quad (2.57)$$

$$\mathbf{H}_A \triangleq \mathbf{A} \mathbf{H}_E \mathbf{A}_r^H = \begin{bmatrix} \mathbf{H}_R^{(r)} & \mathbf{H}_R^{(c)} & -\mathbf{H}_I^{(c)} \\ \mathbf{H}_I^{(r)} & \mathbf{H}_I^{(c)} & \mathbf{H}_R^{(c)} \end{bmatrix} \quad (2.58)$$

where

$$\mathbf{R}_{x^{(r)}} \triangleq E \left[ \tilde{\mathbf{x}}_k^{(r)} \tilde{\mathbf{x}}_k^{(r)T} \right] \quad \mathbf{R}_{x^{(c)}} \triangleq E \left[ \Re \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\} \Re \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\}^T \right] \quad (2.59)$$

$$\mathbf{R}_{x^{(rc)}} \triangleq E \left[ \tilde{\mathbf{x}}_k^{(r)} \Re \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\}^T \right] \quad \hat{\mathbf{R}}_{x^{(rc)}} \triangleq E \left[ \tilde{\mathbf{x}}_k^{(r)} \Im \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\}^T \right] \quad (2.60)$$

$$\hat{\mathbf{R}}_{x^{(cc)}} \triangleq E \left[ \Re \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\} \Im \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\}^T \right] \quad \hat{\mathbf{R}}_{x^{(c)}} \triangleq E \left[ \Im \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\} \Im \left\{ \tilde{\mathbf{x}}_k^{(c)} \right\}^T \right], \quad (2.61)$$

$\mathbf{R}_{n^{(c)}}$ ,  $\hat{\mathbf{R}}_{n^{(c)}}$ , and  $\hat{\mathbf{R}}_{n^{(cc)}}$  are the auto-correlation matrices and the cross-correlation matrix of the real part and the imaginary part of the noise vector

$\mathbf{T}_{N_f}[\mathbf{n}_k, k]$ , the  $(N_o N_f) \times n_r(N_f + \nu)$  matrices  $\mathbf{H}_R^{(r)}$  and  $\mathbf{H}_I^{(r)}$  contain the columns of  $\Re\{\mathbf{H}\}$  and  $\Im\{\mathbf{H}\}$ , respectively, that perform the processing of the real-valued transmitted sequences and, finally, the  $(N_o N_f) \times (N_i - n_r)(N_f + \nu)$  matrices  $\mathbf{H}_R^{(c)}$  and  $\mathbf{H}_I^{(c)}$  contain the remaining columns of  $\Re\{\mathbf{H}\}$  and  $\Im\{\mathbf{H}\}$ , respectively. Then, from (2.55) and (2.56)-(2.58) one has:

$$\begin{bmatrix} \mathbf{B}_\alpha^{(opt)} \\ \mathbf{D}_\alpha^{(opt)} \end{bmatrix} = \mathbf{A}_r \bar{\mathcal{E}}_{n_r} [\mathbf{f}_1 [\mathbf{A}_r^H \mathbf{R}_x^A \mathbf{A}_r, \mathbf{A}^H \mathbf{R}_n^A \mathbf{A}, \mathbf{A}^H \mathbf{H}_A \mathbf{A}_r]]. \quad (2.62)$$

It can be easily verified that the right-hand side of (2.62) can be re-written so that it requires only additions, multiplications, and inversion of real-valued symmetric matrices in the form

$$\mathbf{g}(\mathbf{L}_{m \times m}, \mathbf{M}_{m \times n}, \mathbf{N}_{m \times n}, \mathbf{P}_{n \times n}, \mathbf{Q}_{n \times n}, \mathbf{S}_{n \times n}) \triangleq \begin{bmatrix} \mathbf{L} & \mathbf{M} & \mathbf{N} \\ \mathbf{M}^T & \mathbf{P} & \mathbf{Q} \\ \mathbf{N}^T & \mathbf{Q}^T & \mathbf{S} \end{bmatrix}. \quad (2.63)$$

The set of symmetric matrices  $\mathbf{g}(\mathbf{L}_{m \times m}, \mathbf{M}_{m \times n}, \mathbf{N}_{m \times n}, \mathbf{P}_{n \times n}, \mathbf{Q}_{n \times n}, \mathbf{S}_{n \times n})$  is a matrix algebra  $\Omega_{m,n}$  closed under addition, multiplication, and inversion. By resorting to the formula for inversion of block-partitioned matrices [55], and omitting the matrix sizes in the notation, it can be shown that the following relation holds

$$[\mathbf{g}(\mathbf{L}, \mathbf{M}, \mathbf{N}, \mathbf{P}, \mathbf{Q}, \mathbf{S})]^{-1} = \mathbf{g}(\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4, \mathbf{G}_5, \mathbf{G}_6) \quad (2.64)$$

where

$$\mathbf{G}_1 = \mathbf{L}^{-1} - [\mathbf{G}_2 \quad \mathbf{G}_3] \begin{bmatrix} \mathbf{M}^T \\ \mathbf{N}^T \end{bmatrix} \mathbf{L}^{-1} \quad (2.65)$$

$$\mathbf{G}_2 = -\mathbf{L}^{-1} (\mathbf{M} \mathbf{G}_4 + \mathbf{N} \mathbf{G}_5^T) \quad (2.66)$$

$$\mathbf{G}_3 = -\mathbf{L}^{-1} (\mathbf{M} \mathbf{G}_5 + \mathbf{N} \mathbf{G}_6) \quad (2.67)$$

$$\mathbf{G}_4 = (\mathbf{P} - \mathbf{M}^T \mathbf{L}^{-1} \mathbf{M})^{-1} - \mathbf{G}_5 (\mathbf{Q}^T - \mathbf{N}^T \mathbf{A}^{-1} \mathbf{M}) (\mathbf{P} - \mathbf{M}^T \mathbf{L}^{-1} \mathbf{M})^{-1} \quad (2.68)$$

$$\mathbf{G}_5 = -(\mathbf{P} - \mathbf{M}^T \mathbf{L}^{-1} \mathbf{M})^{-1} (\mathbf{Q}^T - \mathbf{N}^T \mathbf{A}^{-1} \mathbf{M})^T \mathbf{G}_6 \quad (2.69)$$

$$\mathbf{G}_6 = [(\mathbf{S} - \mathbf{N}^T \mathbf{L}^{-1} \mathbf{N}) - (\mathbf{Q}^T - \mathbf{N}^T \mathbf{A}^{-1} \mathbf{M}) \cdot (\mathbf{P} - \mathbf{M}^T \mathbf{L}^{-1} \mathbf{M})^{-1} (\mathbf{Q}^T - \mathbf{N}^T \mathbf{A}^{-1} \mathbf{M})^T]^{-1}. \quad (2.70)$$

By accounting for the properties of  $\Omega_{m,n}$ , the optimum  $(\mathbf{B}_\alpha^{(opt)}, \mathbf{D}_\alpha^{(opt)})$  can be determined by means of the following alternative procedure:

1. determine the matrices in (2.71) and in (2.72) set below:

$$\begin{aligned} & \mathbf{g}(\mathbf{I}_{N_o N_f}, \mathbf{0}_{N_o N_f}, \mathbf{0}_{N_o N_f}, \mathbf{R}_{1,n}, \mathbf{R}_{2,n}, \mathbf{R}_{3,n}) \\ & \triangleq \left[ \mathbf{g} \left( \mathbf{I}_{N_o N_f}, \mathbf{0}_{N_o N_f}, \mathbf{0}_{N_o N_f}, \mathbf{R}_{n(c)}, \hat{\mathbf{R}}_{n(cc)}, \hat{\mathbf{R}}_{n(c)} \right) \right]^{-1} \quad (2.71) \\ \mathbf{R} & \triangleq \mathbf{g}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \mathbf{R}_4, \mathbf{R}_5, \mathbf{R}_6) \\ & \triangleq \left[ \mathbf{g} \left( \mathbf{R}_{x(r)}, \mathbf{R}_{x(rc)}, \hat{\mathbf{R}}_{x(rc)}, \mathbf{R}_{x(c)}, \hat{\mathbf{R}}_{x(cc)}, \hat{\mathbf{R}}_{x(c)} \right) \right]^{-1} \\ & \quad + \mathbf{H}_A^T \begin{bmatrix} \mathbf{R}_{1,n} & \mathbf{R}_{2,n} \\ \mathbf{R}_{2,n}^T & \mathbf{R}_{3,n} \end{bmatrix} \mathbf{H}_A \quad (2.72) \end{aligned}$$

2. extract from  $\mathbf{R}_1$ ,  $\mathbf{R}_2^T$ , and  $\mathbf{R}_3^T$  the matrices  $\mathbf{R}_{1,e}$ ,  $\bar{\mathbf{R}}_{2,e}$ , and  $\bar{\mathbf{R}}_{3,e}$  (respectively) that contain their first  $n_r(\Delta + 1)$  columns; moreover, extract from  $\mathbf{R}_2$ ,  $\mathbf{R}_3$ ,  $\mathbf{R}_4$ ,  $\mathbf{R}_5$ ,  $\mathbf{R}_5^T$ , and  $\mathbf{R}_6$  the matrices  $\mathbf{R}_{2,e}$ ,  $\mathbf{R}_{3,e}$ ,  $\mathbf{R}_{4,e}$ ,  $\mathbf{R}_{5,e}$ ,  $\bar{\mathbf{R}}_{5,e}$ , and  $\mathbf{R}_{6,e}$  (respectively) that contain their first  $(N_i - n_r)(\Delta + 1)$  columns;
3. extract from  $\mathbf{R}_1$  the upper-left square sub-block  $\mathbf{R}_{1,f}$  of size  $n_r(\Delta + 1)$  and from  $\mathbf{R}_2$  and  $\mathbf{R}_3$  the upper-left sub-blocks  $\mathbf{R}_{2,f}$  and  $\mathbf{R}_{3,f}$  (respectively) of size  $n_r(\Delta + 1) \times (N_i - n_r)(\Delta + 1)$ . Finally, extract from  $\mathbf{R}_4$ ,  $\mathbf{R}_5$ , and  $\mathbf{R}_6$  the upper-left square sub-blocks  $\mathbf{R}_{4,f}$ ,  $\mathbf{R}_{5,f}$ , and  $\mathbf{R}_{6,f}$  (respectively) of size  $(N_i - n_r)(\Delta + 1)$ ;
4. determine the matrix

$$\begin{aligned} \mathbf{R}_\Delta & \triangleq \mathbf{g}(\mathbf{R}_{1,\Delta}, \mathbf{R}_{2,\Delta}, \mathbf{R}_{3,\Delta}, \mathbf{R}_{4,\Delta}, \mathbf{R}_{5,\Delta}, \mathbf{R}_{6,\Delta}) \\ & \triangleq \left[ \mathbf{g}(\mathbf{R}_{1,f}, \mathbf{R}_{2,f}, \mathbf{R}_{3,f}, \mathbf{R}_{4,f}, \mathbf{R}_{5,f}, \mathbf{R}_{6,f}) \right]^{-1} \quad (2.73) \end{aligned}$$

5. extract from the matrix

$$\mathbf{R}_{1,e} \mathbf{R}_{1,\Delta} + \mathbf{R}_{2,e} \mathbf{R}_{2,\Delta}^T + \mathbf{R}_{3,e} \mathbf{R}_{3,\Delta}^T \quad (2.74)$$

the lower-right sub-block  $\mathbf{B}^{(1,1)}$  of size  $n_r(N_b + 1) \times n_r$  and from the matrices

$$\hat{\mathbf{R}}_{1,e} \mathbf{R}_{2,\Delta} + \mathbf{R}_{2,e} \mathbf{R}_{4,\Delta} + \mathbf{R}_{3,e} \mathbf{R}_{5,\Delta}^T \quad (2.75)$$

$$\hat{\mathbf{R}}_{1,e} \mathbf{R}_{3,\Delta} + \mathbf{R}_{2,e} \mathbf{R}_{5,\Delta} + \mathbf{R}_{3,e} \mathbf{R}_{6,\Delta}^T \quad (2.76)$$

the lower-right sub-blocks  $\mathbf{B}^{(1,2)}$  and  $\mathbf{B}^{(1,3)}$  (respectively) of size  $n_r(N_b + 1) \times (N_i - n_r)$ . Moreover, extract from the matrices

$$\bar{\mathbf{R}}_{2,e}\mathbf{R}_{1,\Delta} + \mathbf{R}_{4,e}\mathbf{R}_{2,\Delta}^T + \mathbf{R}_{5,e}\mathbf{R}_{3,\Delta}^T \quad (2.77)$$

$$\bar{\mathbf{R}}_{3,e}\mathbf{R}_{1,\Delta} + \bar{\mathbf{R}}_{5,e}\mathbf{R}_{2,\Delta}^T + \mathbf{R}_{6,e}\mathbf{R}_{3,\Delta}^T \quad (2.78)$$

the lower-right sub-blocks  $\mathbf{B}^{(2,1)}$  and  $\mathbf{B}^{(3,1)}$  (respectively) of size  $(N_i - n_r)(N_b + 1) \times n_r$  and from the matrices

$$\bar{\mathbf{R}}_{2,e}\mathbf{R}_{2,\Delta} + \mathbf{R}_{4,e}\mathbf{R}_{4,\Delta} + \mathbf{R}_{5,e}\mathbf{R}_{5,\Delta}^T \quad (2.79)$$

$$\bar{\mathbf{R}}_{2,e}\mathbf{R}_{3,\Delta} + \mathbf{R}_{4,e}\mathbf{R}_{5,\Delta} + \mathbf{R}_{5,e}\mathbf{R}_{6,\Delta} \quad (2.80)$$

$$\bar{\mathbf{R}}_{3,e}\mathbf{R}_{2,\Delta} + \bar{\mathbf{R}}_{5,e}\mathbf{R}_{4,\Delta} + \mathbf{R}_{6,e}\mathbf{R}_{5,\Delta}^T \quad (2.81)$$

$$\bar{\mathbf{R}}_{3,e}\mathbf{R}_{3,\Delta} + \bar{\mathbf{R}}_{5,e}\mathbf{R}_{5,\Delta} + \mathbf{R}_{6,e}\mathbf{R}_{6,\Delta} \quad (2.82)$$

the lower-right sub-blocks  $\mathbf{B}^{(2,2)}$ ,  $\mathbf{B}^{(2,3)}$ ,  $\mathbf{B}^{(3,2)}$ , and  $\mathbf{B}^{(3,3)}$  (respectively) of size  $(N_i - n_r)(N_b + 1) \times (N_i - n_r)$ ;

Finally, the optimum  $(\mathbf{B}_\alpha^{(opt)}, \mathbf{D}_\alpha^{(opt)})$  are given by the following matrices:

$$\mathbf{B}_\alpha^{(opt)} = \begin{bmatrix} \mathbf{B}^{(1,1)} & \mathbf{B}^{(1,2)} - j\mathbf{B}^{(1,3)} \\ \mathbf{B}^{(2,1)} & \mathbf{B}^{(2,2)} - j\mathbf{B}^{(2,3)} \end{bmatrix} \quad (2.83)$$

$$\mathbf{D}_\alpha^{(opt)} = \begin{bmatrix} \mathbf{B}^{(3,1)} & \mathbf{B}^{(3,2)} - j\mathbf{B}^{(3,3)} \end{bmatrix}. \quad (2.84)$$

The properties of the matrix algebra  $\Omega_{m,n}$  and the formula for inversion of block-partitioned matrices also allow us to express the optimum  $(\mathbf{W}_\alpha^{(opt)}, \mathbf{G}_\alpha^{(opt)})$  as follows:

$$\begin{aligned} \mathbf{W}_\alpha^{(opt)} &= \mathbf{R}_{y^{(c)}}^{-1} \left( \mathbf{R}_{yx^{(r)}} \tilde{\mathbf{B}}_\alpha + \mathbf{R}_{yx^{(rc)}} \tilde{\mathbf{D}}_\alpha \right) \\ &\quad - \mathbf{R}_{y^{(c)}}^{-1} \hat{\mathbf{R}}_{y^{(cc)}} \mathbf{G}_\alpha^{(opt)} \end{aligned} \quad (2.85)$$

$$\begin{aligned} \mathbf{G}_\alpha^{(opt)} &= \left( \hat{\mathbf{R}}_{y^{(c)}} - \hat{\mathbf{R}}_{y^{(cc)}}^T \mathbf{R}_{y^{(c)}}^{-1} \hat{\mathbf{R}}_{y^{(cc)}} \right)^{-1} \\ &\quad \cdot \left[ \left( \mathbf{R}_{yx^{(cr)}} - \hat{\mathbf{R}}_{y^{(cc)}}^T \mathbf{R}_{y^{(c)}}^{-1} \mathbf{R}_{yx^{(r)}} \right) \tilde{\mathbf{B}}_\alpha \right. \\ &\quad \left. + \left( \mathbf{R}_{yx^{(c)}} - \hat{\mathbf{R}}_{y^{(cc)}}^T \mathbf{R}_{y^{(c)}}^{-1} \mathbf{R}_{yx^{(rc)}} \right) \tilde{\mathbf{D}}_\alpha \right] \end{aligned} \quad (2.86)$$

where

$$\begin{aligned} \tilde{\mathbf{B}}_\alpha &\triangleq \begin{bmatrix} \mathbf{0}_{n_r \Delta \times N_i} \\ \text{-----} \\ \mathbf{B}^{(1,1)} & \mathbf{B}^{(1,2)} - j\mathbf{B}^{(1,3)} \\ \text{-----} \\ \mathbf{0}_{(N_i - n_r)\Delta \times N_i} \\ \text{-----} \\ \mathbf{B}^{(2,1)} & \mathbf{B}^{(2,2)} - j\mathbf{B}^{(2,3)} \\ \text{-----} \\ \mathbf{0}_{(N_i - n_r)\Delta \times N_i} \\ \text{-----} \\ \mathbf{B}^{(3,1)} & \mathbf{B}^{(3,2)} - j\mathbf{B}^{(3,3)} \end{bmatrix} \\ \tilde{\mathbf{D}}_\alpha &\triangleq \begin{bmatrix} \mathbf{0}_{(N_i - n_r)\Delta \times N_i} \\ \text{-----} \\ \mathbf{B}^{(3,1)} & \mathbf{B}^{(3,2)} - j\mathbf{B}^{(3,3)} \end{bmatrix} \end{aligned} \quad (2.87)$$

and  $\mathbf{R}_{y^{(c)}}$  and  $\hat{\mathbf{R}}_{y^{(c)}}$  are the correlation matrices of the real part and the imaginary part, respectively, of the output vector  $\mathbf{T}_{N_f} [\mathbf{y}_k, k]$ ,  $\hat{\mathbf{R}}_{y^{(cc)}}$  denotes their cross-correlation matrix,  $\mathbf{R}_{yx^{(r)}} [\mathbf{R}_{yx^{(cr)}}]$  is the cross-correlation matrix between the real [imaginary] part of the output vector  $\mathbf{T}_{N_f} [\mathbf{y}_k, k]$  and the real part of the input vector  $\mathbf{T}_{N_f + \nu} [\mathbf{x}_k, k]$  and, analogously,  $\mathbf{R}_{yx^{(rc)}} [\mathbf{R}_{yx^{(c)}}]$  is the cross-correlation matrix between the real [imaginary] part of the output vector  $\mathbf{T}_{N_f} [\mathbf{y}_k, k]$  and the imaginary part of the input vector  $\tilde{\mathbf{x}}_k^{(c)}$ .

This procedure, by exploiting the special structure (2.63) of the correlation matrices, exhibits a lower computational complexity with respect to the procedure (2.55): in fact, the maximum size of the matrices to be inverted in feedback filter reduced-order design is  $\max(n_r(N_f + \nu), (N_i - n_r)(N_f + \nu))$  instead of  $(2N_i - n_r)(N_f + \nu)$  required in (2.55).

In the particular case where both the transmitted and the noise vector are complex-valued rotationally invariant, i.e.  $n_r = 0$ ,  $\mathbf{R}_{x^{(c)}} = \hat{\mathbf{R}}_{x^{(c)}}$ ,  $\mathbf{R}_{n^{(c)}} = \hat{\mathbf{R}}_{n^{(c)}}$ , and the matrices  $\hat{\mathbf{R}}_{x^{(cc)}}$  and  $\hat{\mathbf{R}}_{n^{(cc)}}$  are skew-symmetric [56], consequently,  $\mathbf{R}_{4,e} = \mathbf{R}_{6,e}$ ,  $\hat{\mathbf{R}}_{5,e} = -\mathbf{R}_{5,e}$ ,  $\mathbf{R}_{4,\Delta} = \mathbf{R}_{6,\Delta}$ , and  $\mathbf{R}_{5,\Delta}$  is skew-symmetric; moreover,  $\mathbf{R}_{y^{(c)}} = \hat{\mathbf{R}}_{y^{(c)}}$ ,  $\hat{\mathbf{R}}_{y^{(cc)}}$  is skew-symmetric,  $\mathbf{R}_{yx^{(r)}} = \mathbf{R}_{yx^{(c)}}$ , and  $\mathbf{R}_{yx^{(rc)}} = -\mathbf{R}_{yx^{(cr)}}$ , and, hence,  $\mathbf{D}_\alpha^{(opt)} = -j\mathbf{B}_\alpha^{(opt)}$  and  $\mathbf{G}_\alpha^{(opt)} = -j\mathbf{W}_\alpha^{(opt)}$ , i.e. the optimum WL-WDF-MMSE equalizer degenerates into DF-MMSE one proposed in [9]. Note that, when the noise vector is rotationally variant, the optimum WL-WDF equalizer does not reduce to the linear one also when the transmitted sequences are rotationally invariant: for such a reason, in order to examine some particular structures of the WL-WDF-MMSE equalizer, in the sequel we assume a rotationally invariant noise. When all the  $N_i$  transmitted signals  $x_k^{(i)}$  are real-valued, i.e.  $n_r = N_i$ , the op-

imum feedback filter is linear and real-valued, in fact  $\mathbf{B}_\alpha^{(opt)} = \mathbf{B}^{(1,1)}$ , while the optimum feedforward filter is widely linear. Such an equalizer, here renamed WL-DF-MMSE, has been proposed in [43] with reference to the SISO channel.

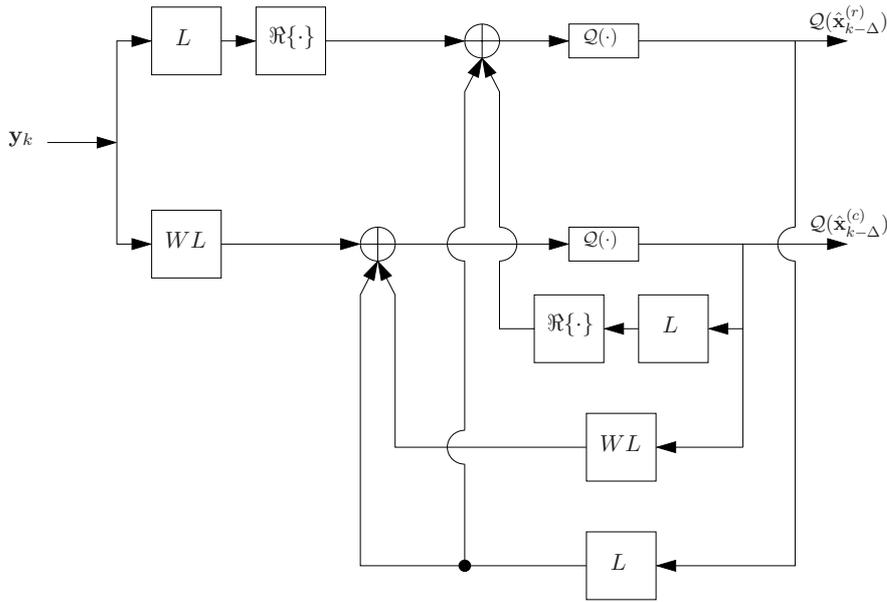
In the case  $0 < n_r < N_i$ , it is simple to verify that the optimum matrices  $(\mathbf{W}_\alpha^{(opt)}, \mathbf{G}_\alpha^{(opt)})$  and  $(\mathbf{B}_\alpha^{(opt)}, \mathbf{D}_\alpha^{(opt)})$  satisfy the following conditions:

- A) the feedforward processing of the received vector is widely linear, except for the extreme case  $n_r = 0$ ;
  - a1) the feedforward sub-processing that provides the contributes to the estimate of the real-valued components of  $\mathbf{x}_{k-\Delta}$  exhibits an halved computational complexity with respect to the processing which provides the last  $N_i - n_r$  ones;
- B) the processing of the past decisions  $\hat{\mathbf{x}}_{E,k-\Delta}$  is widely linear unless  $n_r = 0$ ;
  - b1) the sub-processing of the real-valued components of  $\hat{\mathbf{x}}_{E,k-\Delta}$  is linear rather than widely linear;
  - b2) the sub-processing that provides the real-valued components of  $\hat{\mathbf{x}}_{E,k-\Delta}$  exhibits an halved computational complexity;
  - b3) the sub-processing that provides the complex-valued components of  $\hat{\mathbf{x}}_{E,k-\Delta}$  is widely linear, although the complex-valued components of the transmitted vector  $\mathbf{x}_{k-\Delta}$  are complex-valued circularly symmetric.

The above properties are summarized in the overall structure reported in Fig. 2.2. Note that the synthesis of the WL-WDF equalizer can be performed both in the presence of uncorrelated input sequences and in the case where space-time coding introduces spatial and temporal correlation (with  $\mathbf{R}_x^E$  non-singular) in the transmitted sequences [26].

It is possible to show that the error-correlation matrix of the WL-WDF-MMSE equalizer is equal to

$$\mathbf{R}_e^{(opt)} \triangleq E \left[ (\mathbf{x}_{k-\Delta} - \hat{\mathbf{x}}_{E,k-\Delta}) (\mathbf{x}_{k-\Delta} - \hat{\mathbf{x}}_{E,k-\Delta})^H \right] = \mathbf{\Phi}^H \mathbf{R}_\Delta \mathbf{\Phi} \quad (2.88)$$



**Figure 2.2:** Block diagram of the WL-WDF-MMSE equalizer: the blocks labelled with  $L$  are linear filters, whereas the ones labelled with  $WL$  are widely linear filters.

where

$$\Phi \triangleq \left[ \begin{array}{c|c} \mathbf{0}_{n_r \Delta \times n_r} & \mathbf{0}_{(n_r + N_i \Delta) \times (N_i - n_r)} \\ \mathbf{I}_{n_r} & \mathbf{I}_{N_i - n_r} \\ \mathbf{0}_{2(N_i - n_r)(\Delta + 1) \times n_r} & \begin{array}{c} \mathbf{0}_{(N_i - n_r)\Delta \times (N_i - n_r)} \\ -j \cdot \mathbf{I}_{N_i - n_r} \end{array} \end{array} \right]. \quad (2.89)$$

Moreover, it is straightforward to recognize that the optimum WL equalizer in (2.53), as shown in [42, 57] (by using the complex-valued notation), is given by (2.85) and (2.86) setting  $\mathbf{B}^{(1,1)} = [\mathbf{I}_{n_r} \ \mathbf{0}_{n_r \times n_r N_b}]^T$ ,  $\mathbf{B}^{(1,2)} = \mathbf{B}^{(1,3)} = \mathbf{0}$ ,  $\mathbf{B}^{(2,1)} = \mathbf{B}^{(3,1)} = \mathbf{0}$ ,  $\mathbf{B}^{(2,2)} = \mathbf{B}^{(3,3)} = [\mathbf{I}_{N_i - n_r} \ \mathbf{0}_{(N_i - n_r) \times (N_i - n_r) N_b}]^T$ , and  $\mathbf{B}^{(2,3)} = \mathbf{B}^{(3,2)} = \mathbf{0}$ .

### 2.6.2 WL-DF-MMSE equalizer for Scenario 1

Since it is interesting to evaluate the advantage of the widely linear feedback filtering over linear feedback filtering, in this section we present the deriva-

tion of the WL-DF-MMSE equalizer (which employs a linear feedback filter rather than a widely linear one). Following the same guidelines of the previous section, it can be shown that the optimum WL-DF-MMSE equalizer can be determined by the following procedure:

1. determine the matrix

$$\begin{aligned}\dot{\mathbf{R}} &\triangleq \mathbf{g} \left( \dot{\mathbf{R}}_1, \dot{\mathbf{R}}_2, \dot{\mathbf{R}}_3, \dot{\mathbf{R}}_4, \dot{\mathbf{R}}_5, \dot{\mathbf{R}}_6 \right) \\ &\triangleq \mathbf{R}_x^A - \mathbf{R}_x^A \mathbf{H}_A^H \left( \mathbf{H}_A \mathbf{R}_x^A \mathbf{H}_A^H + \mathbf{R}_n^A \right)^{-1} \mathbf{H}_A \mathbf{R}_x^A\end{aligned}\quad (2.90)$$

2. determine the matrix  $\mathbf{R}^\beta$  defined as follows:

$$\begin{aligned}\mathbf{R}^\beta &\triangleq \begin{bmatrix} \dot{\mathbf{R}}_1 & \mathbf{R}_2^\beta \\ \mathbf{R}_2^{\beta H} & \mathbf{R}_3^\beta \end{bmatrix} \\ &\triangleq \begin{bmatrix} \dot{\mathbf{R}}_1 & \dot{\mathbf{R}}_2 - j\dot{\mathbf{R}}_3 \\ \dot{\mathbf{R}}_2^T + j\dot{\mathbf{R}}_3^T & \dot{\mathbf{R}}_4 + \dot{\mathbf{R}}_6 - j \left( \dot{\mathbf{R}}_5 - \dot{\mathbf{R}}_5^T \right) \end{bmatrix}\end{aligned}\quad (2.91)$$

3. determine the following matrices:

$$\mathbf{R}_1 \triangleq \left( \dot{\mathbf{R}}_1 \right)^{-1} + \left( \dot{\mathbf{R}}_1 \right)^{-1} \mathbf{R}_2^\beta \mathbf{R}_3^\beta \mathbf{R}_2^{\beta H} \left( \dot{\mathbf{R}}_1 \right)^{-1}\quad (2.92)$$

$$\mathbf{R}_2 \triangleq - \left( \dot{\mathbf{R}}_1 \right)^{-1} \mathbf{R}_2^\beta \mathbf{R}_3\quad (2.93)$$

$$\mathbf{R}_3 \triangleq \left[ \mathbf{R}_3^\beta - \mathbf{R}_2^{\beta H} \left( \dot{\mathbf{R}}_1 \right)^{-1} \mathbf{R}_2^\beta \right]^{-1}\quad (2.94)$$

4. extract from  $\mathbf{R}_1$  and  $\mathbf{R}_2^H$  the matrices  $\mathbf{R}_{1,e}^\beta$  and  $\bar{\mathbf{R}}_{2,e}^\beta$  (respectively) that contain the first  $n_r (\Delta + 1)$  columns; analogously, extract from  $\mathbf{R}_2$  and  $\mathbf{R}_3$  the matrices  $\mathbf{R}_{2,e}^\beta$  and  $\mathbf{R}_{3,e}^\beta$  (respectively) that contain the first  $(N_i - n_r) (\Delta + 1)$  columns;
5. extract from  $\mathbf{R}_1$  the upper-left square sub-block  $\mathbf{R}_{1,f}^\beta$  of size  $n_r (\Delta + 1)$ , from  $\mathbf{R}_2$  the upper-left sub-block  $\mathbf{R}_{2,f}^\beta$  of size  $n_r (\Delta + 1) \times (N_i - n_r) (\Delta + 1)$ , and, finally, from  $\mathbf{R}_3$  the upper-left square sub-block  $\mathbf{R}_{3,f}^\beta$  of size  $(N_i - n_r) (\Delta + 1)$ ;

6. determine the matrices

$$\mathbf{R}_{1,\Delta}^\beta \triangleq \left(\mathbf{R}_{1,f}^\beta\right)^{-1} + \left(\mathbf{R}_{1,f}^\beta\right)^{-1} \mathbf{R}_{2,f}^\beta \mathbf{R}_{3,\Delta}^\beta \mathbf{R}_{2,f}^{\beta H} \left(\mathbf{R}_{1,f}^\beta\right)^{-1} \quad (2.95)$$

$$\mathbf{R}_{2,\Delta}^\beta \triangleq - \left(\mathbf{R}_{1,f}^\beta\right)^{-1} \mathbf{R}_{2,f}^\beta \mathbf{R}_{3,\Delta}^\beta \quad (2.96)$$

$$\mathbf{R}_{3,\Delta}^\beta \triangleq \left[ \mathbf{R}_{3,f}^\beta - \mathbf{R}_{2,f}^{\beta H} \left(\mathbf{R}_{1,f}^\beta\right)^{-1} \mathbf{R}_{2,f}^\beta \right]^{-1} \quad (2.97)$$

7. extract from the matrix

$$\mathbf{R}_{1,e}^\beta \mathbf{R}_{1,\Delta}^\beta + \mathbf{R}_{2,e}^\beta \mathbf{R}_{2,\Delta}^\beta \quad (2.98)$$

the lower-right sub-block  $\mathbf{B}_\beta^{(1,1)}$  of size  $n_r(N_b + 1) \times n_r$  and from the matrix

$$\mathbf{R}_{1,e}^\beta \mathbf{R}_{2,\Delta}^\beta + \mathbf{R}_{2,e}^\beta \mathbf{R}_{3,\Delta}^\beta \quad (2.99)$$

the lower-right sub-block  $\mathbf{B}_\beta^{(1,2)}$  of size  $n_r(N_b + 1) \times (N_i - n_r)$ . Moreover, extract from the matrix

$$\bar{\mathbf{R}}_{2,e}^\beta \mathbf{R}_{1,\Delta}^\beta + \mathbf{R}_{3,e}^\beta \mathbf{R}_{2,\Delta}^\beta \quad (2.100)$$

the lower-right sub-block  $\mathbf{B}_\beta^{(2,1)}$  of size  $(N_i - n_r)(N_b + 1) \times n_r$  and from the matrix

$$\bar{\mathbf{R}}_{2,e}^\beta \mathbf{R}_{2,\Delta}^\beta + \mathbf{R}_{3,e}^\beta \mathbf{R}_{3,\Delta}^\beta \quad (2.101)$$

the lower-right sub-block  $\mathbf{B}_\beta^{(2,2)}$  of size  $(N_i - n_r)(N_b + 1) \times (N_i - n_r)$ .

Therefore, the optimum feedback filters are:

$$\mathbf{B}_\beta^{(opt)} = \begin{bmatrix} \mathbf{B}_\beta^{(1,1)} & \mathbf{B}_\beta^{(1,2)} \\ \mathbf{B}_\beta^{(2,1)} & \mathbf{B}_\beta^{(2,2)} \end{bmatrix} \quad (2.102)$$

$$\mathbf{D}_\beta^{(opt)} = \begin{bmatrix} -j\mathbf{B}_\beta^{(2,1)} & -j\mathbf{B}_\beta^{(2,2)} \end{bmatrix}. \quad (2.103)$$

The optimum feedforward filters are given by equations (2.85) and (2.86) by replacing  $\mathbf{B}_\alpha^{(opt)}$  in (2.83) with  $\mathbf{B}_\beta^{(opt)}$  in (2.102) and  $\mathbf{D}_\alpha^{(opt)}$  in (2.84) with  $\mathbf{D}_\beta^{(opt)}$  in (2.103).

## 2.7 WL-WDF-MMSE equalizer for Scenario 2

Analogously to Scenario 1, the optimum WL-WDF equalizer in Scenario 2 can be determined by applying the procedure (2.40) to the input-output channel model (2.50), and accounting for the augmented correlation matrices  $\mathbf{R}_x^E(m)$  and  $\mathbf{R}_n^E(m)$  (as already shown in Section 2.6 with reference to Scenario 1). Note that, unlike Scenario 1 where  $\mathbf{B}_{E,0}^H = \mathbf{0}$ , the term  $\left(\mathbf{B}_{E,0}^H \mathcal{E}_{n_r}[\hat{\mathbf{x}}_{k-\Delta}]\right)$  is included in (2.52) and the lower triangular matrix tap  $\mathbf{B}_{E,0}^H$  is to be optimized according to the MMSE criterion.

As it happens for DF equalization, it is easy to understand that the performances of the WL-WDF-MMSE equalizer are affected by the decision ordering. In fact, let  $\mathbf{P}$  be a permutation matrix of size  $(2N_i - n_r)$ , such that  $\mathbf{P}^T \mathbf{P} = \mathbf{I}_{2N_i - n_r}$ . By defining the (row) permuted input vector  $\mathcal{E}_{n_r}[\mathbf{x}_k](P) \triangleq \mathbf{P} \mathcal{E}_{n_r}[\mathbf{x}_k]$  and, hence, the (column) permuted channel matrix  $\tilde{\mathcal{E}}_{n_r}[\mathbf{H}_m](P) \triangleq \tilde{\mathcal{E}}_{n_r}[\mathbf{H}_m] \mathbf{P}^T$ , the channel output (2.50) can be equivalently re-written as follows:

$$\mathcal{E}_0[\mathbf{y}_k] = \sum_{m=0}^{\nu} \tilde{\mathcal{E}}_{n_r}[\mathbf{H}_m](P) \mathcal{E}_{n_r}[\mathbf{x}_{k-m}](P) + \mathbf{n}_k. \quad (2.104)$$

The reordering of the columns of  $\tilde{\mathcal{E}}_{n_r}[\mathbf{H}_m]$  determines different WL-WDF equalizer structures with different performances. Note that the optimization over the decision ordering requires to account for  $(2N_i - n_r)!$  permutation matrices  $\mathbf{P}$ , instead of  $N_i!$ , as it happens in DF equalization. It is straightforward verified that also the WL-WDF-MMSE equalizer can be implemented by resorting to two equalization stages:

- 1<sup>st</sup> the former stage is the optimum WL-WDF-MMSE equalizer in Scenario 1, maybe synthesized by resorting to the low computational complexity algorithm proposed in subsection 2.6.1. Specifically, such a stage performs a dispersive WL-WDF equalization by processing the received vector  $\mathbf{y}_k$  over  $N_f$  symbol period and by utilizing only past decisions  $\mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta-n}]$  with  $n > 0$  to provide the estimate  $\hat{\mathbf{x}}_{E,k-\Delta}^{(1)}$  of  $\mathbf{x}_{k-\Delta}$ ;
- 2<sup>nd</sup> the latter stage is the optimum WL-WDF non-dispersive equalizer in Scenario 2 that performs the non-dispersive equalization of  $\hat{\mathbf{x}}_{E,k-\Delta}^{(1)}$  by utilizing current decisions and whose performances depend on the decision ordering. Clearly, any suboptimal ordering algorithm designed for DF equalization can be used in WL-WDF equalization.

Note that, when Scenario 2 is considered, the decision over  $\Re\{x_{k-\Delta}^{(\ell)}\}$  with  $\ell > n_r$  can improve the estimation of  $\Im\{x_{k-\Delta}^{(\ell)}\}$  or, *vice versa*, when a different ordering of the components of  $\mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta}]$  is adopted, the decision over  $\Im\{x_{k-\Delta}^{(\ell)}\}$  can improve the estimation of  $\Re\{x_{k-\Delta}^{(\ell)}\}$ . In other words, differently from the DF equalizer, which simultaneously provides the estimates of  $\Re\{x_{k-\Delta}^{(\ell)}\}$  and  $\Im\{x_{k-\Delta}^{(\ell)}\}$ , the WL-WDF equalizer can estimate in a sequential manner the quadrature and the in-phase components of the complex-valued transmitted symbols.

Let us now show that, although not intuitive, the estimate of  $\Re\{x_{k-\Delta}^{(\ell)}\}$  can improve the estimation of  $\Im\{x_{k-\Delta}^{(\ell)}\}$ , also when  $x_{k-\Delta}^{(\ell)}$  is complex-valued circularly symmetric. To this aim, assume (for simplicity) that the components of the transmitted vector  $\mathbf{x}_k$  are uncorrelated with each other and temporally uncorrelated, i.e.,  $\mathbf{R}_x(m) = \Sigma_x \delta(m)$  with  $\Sigma_x$  diagonal, and (assuming correct decisions) rewrite the output of the WL-WDF equalizer as follows:

$$\begin{aligned} \mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}] &= \mathbf{W}_E^H \mathbf{T}_{N_f} [\mathcal{E}_0[\mathbf{y}_k], k] - \mathbf{B}_E^H \mathbf{T}_{N_b+1} [\mathcal{E}_{n_r}[\mathbf{x}_k], k - \Delta] \quad (2.105) \\ &= \mathbf{W}_E^H \tilde{\mathbf{T}}[\tilde{\mathcal{E}}_{n_r}[\mathbf{H}_k], N_f, \Delta + 1] \cdot \mathbf{T}_{\Delta+1} [\mathcal{E}_{n_r}[\mathbf{x}_k], k] \\ &\quad + \mathbf{W}_E^H \tilde{\mathbf{T}}[\tilde{\mathcal{E}}_{n_r}[\mathbf{H}_{k+\Delta+1}], N_f, N_b] \cdot \mathbf{T}_{N_b} [\mathcal{E}_{n_r}[\mathbf{x}_k], k - \Delta - 1] \clubsuit \\ &\quad + \mathbf{W}_E^H \mathbf{T}_{N_f} [\mathcal{E}_0[\mathbf{n}_k], k] - \mathbf{B}_E^H \mathbf{T}_{N_b+1} [\mathcal{E}_{n_r}[\mathbf{x}_k], k - \Delta] \end{aligned}$$

where  $\clubsuit$  is the post-cursor ISI term which affects  $\mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}]$ , and where the degree of freedom lying in channel input ordering have been temporarily disregarded. According to any chosen detection scenario, the optimum feedback filter removes such an ISI term. Therefore, the equalizer output can be equivalently rewritten as follows:

$$\begin{aligned} \mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}^\clubsuit] &\triangleq \mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}] - \clubsuit \quad (2.106) \\ &= \mathbf{W}_E^H \mathbf{r} - \mathbf{B}_{E,0}^H \mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta}] \end{aligned}$$

with  $\mathbf{r} \triangleq \tilde{\mathbf{T}}[\tilde{\mathcal{E}}_{n_r}[\mathbf{H}_k], N_f, \Delta + 1] \cdot \mathbf{T}_{\Delta+1} [\mathcal{E}_{n_r}[\mathbf{x}_k], k] + \mathbf{T}_{N_f} [\mathcal{E}_0[\mathbf{n}_k], k]$ , or, by reordering the rows of  $\mathbf{W}_E$ , as

$$\mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}^\clubsuit] \triangleq \mathbf{W}_E^H \mathcal{E}_0[\bar{\mathbf{r}}] - \mathbf{B}_{E,0}^H \mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta}] \quad (2.107)$$

with  $\bar{\mathbf{r}} \triangleq \tilde{\mathbf{T}}[\mathbf{H}_k, N_f, \Delta + 1] \cdot \mathbf{T}_{\Delta+1} [\mathbf{x}_k, k] + \mathbf{T}_{N_f} [\mathbf{n}_k, k]$ . The matrix<sup>5</sup>  $\mathbf{W}_E^H$  and the lower triangular matrix  $\mathbf{B}_{E,0}^H$  have to be optimized according to the MMSE

<sup>5</sup>Note that, although the rows of  $\mathbf{W}_E$  have been reordered, we have not defined a new matrix filter for simplicity of the notation.

criterion. Since  $\bar{\mathbf{r}}$  does not depend on the channel input ordering, (2.107) can be rewritten by accounting for the dependence on the permutation matrix  $\mathbf{P}$ :

$$\mathcal{E}_{n_r}[\hat{\mathbf{x}}_{E,k-\Delta}^*] \triangleq \mathbf{W}_E^H \mathcal{E}_0[\bar{\mathbf{r}}] - \mathbf{B}_{E,0}^H \underbrace{\mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta}]}_{\triangleq \bar{\mathbf{x}}}(P) \quad (2.108)$$

From the above expression, it can be easily shown<sup>6</sup> that the optimum  $\mathbf{W}_E^H$  and  $\mathbf{B}_{E,0}^H$  are given by:

$$\begin{bmatrix} \mathbf{W}_E^{(opt)}(:, \ell) \\ \mathbf{B}_{E,0}^{(opt)}(1 : \ell - 1, \ell) \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\bar{\mathbf{r}}} & \mathbf{R}_{\bar{\mathbf{r}}\bar{\mathbf{x}}_\ell} \\ \mathbf{R}_{\bar{\mathbf{r}}\bar{\mathbf{x}}_\ell}^T & \mathbf{R}_{\bar{\mathbf{x}}_\ell} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{p}_{\bar{\mathbf{r}}\bar{\mathbf{x}}_\ell} \\ \mathbf{p}_{\bar{\mathbf{x}}_\ell} \end{bmatrix} \quad (2.109)$$

with  $\ell = 1, \dots, 2N_i - n_r$ , and where  $\mathbf{R}_{\bar{\mathbf{r}}} \triangleq E[\mathcal{E}_0[\bar{\mathbf{r}}]\mathcal{E}_0[\bar{\mathbf{r}}^T]]$ ,  $\mathbf{R}_{\bar{\mathbf{r}}\bar{\mathbf{x}}_\ell} \triangleq E[\mathcal{E}_0[\bar{\mathbf{r}}]\bar{\mathbf{x}}_\ell^T(1 : \ell - 1)]$ ,  $\mathbf{R}_{\bar{\mathbf{x}}_\ell} \triangleq E[\bar{\mathbf{x}}_\ell(1 : \ell - 1)\bar{\mathbf{x}}_\ell^T(1 : \ell - 1)]$ ,  $\mathbf{p}_{\bar{\mathbf{r}}\bar{\mathbf{x}}_\ell} \triangleq E[\mathcal{E}_0[\bar{\mathbf{r}}]\bar{\mathbf{x}}_\ell(\ell)]$ , and  $\mathbf{p}_{\bar{\mathbf{x}}_\ell} \triangleq E[\bar{\mathbf{x}}_\ell(1 : \ell - 1)\bar{\mathbf{x}}_\ell(\ell)]$ .

For the sake of simplicity, assume that the permutation matrix  $\mathbf{P}$  in (2.104) is such that  $\{\bar{\mathbf{x}} = \mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta}](P) : \bar{\mathbf{x}}(1) = \Re\{x_{k-\Delta}^{(i)}\}, \bar{\mathbf{x}}(2) = \Im\{x_{k-\Delta}^{(i)}\}$ , i.e., the first two detected components of  $\mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta}](P)$  are the in-phase and the quadrature components of the complex-valued symbol  $x_{k-\Delta}^{(i)}$  ( $i > n_r$ ). From (2.109) and employing the matrix-inversion lemma, it can be verified if the following equality holds true,

$$\mathbf{p}_{11}^T \mathbf{M}_{11} \mathbf{p}_{12} + \mathbf{p}_{21}^T \mathbf{M}_{12}^T \mathbf{p}_{12} + \mathbf{p}_{11}^T \mathbf{M}_{12} \mathbf{p}_{22} + \mathbf{p}_{21}^T \mathbf{M}_{22} \mathbf{p}_{22} = 0 \quad (2.110)$$

with

$$\begin{aligned} \mathbf{p}_{11} &\triangleq E[\Re\{\bar{\mathbf{r}}\}\Re\{x_{k-\Delta}^{(i)}\}] & \mathbf{p}_{21} &\triangleq E[\Im\{\bar{\mathbf{r}}\}\Re\{x_{k-\Delta}^{(i)}\}] \\ \mathbf{p}_{12} &\triangleq E[\Re\{\bar{\mathbf{r}}\}\Im\{x_{k-\Delta}^{(i)}\}] & \mathbf{p}_{22} &\triangleq E[\Im\{\bar{\mathbf{r}}\}\Im\{x_{k-\Delta}^{(i)}\}] \\ \mathbf{P}_{11} &\triangleq E[\Re\{\bar{\mathbf{r}}\}\Re\{\bar{\mathbf{r}}^T\}] & \mathbf{P}_{12} &\triangleq E[\Re\{\bar{\mathbf{r}}\}\Im\{\bar{\mathbf{r}}^T\}] \\ \mathbf{P}_{22} &\triangleq E[\Im\{\bar{\mathbf{r}}\}\Im\{\bar{\mathbf{r}}^T\}] \quad , \end{aligned}$$

and with

$$\begin{aligned} \mathbf{M}_{11} &\triangleq \mathbf{P}_{11}^{-1} + \mathbf{P}_{11}^{-1} \mathbf{P}_{12} \mathbf{M}_{22} \mathbf{P}_{12}^T \mathbf{P}_{11}^{-1} \\ \mathbf{M}_{12} &\triangleq -\mathbf{P}_{11}^{-1} \mathbf{P}_{12} \mathbf{M}_{22} \\ \mathbf{M}_{22} &\triangleq (\mathbf{P}_{22} - \mathbf{P}_{12}^T \mathbf{P}_{11}^{-1} \mathbf{P}_{12})^{-1} \quad , \end{aligned}$$

<sup>6</sup>According to the assumption of correct decisions, treat the feedback filter as a feedforward filter that processes the vector  $\mathbf{x}_{k-\Delta}$ .

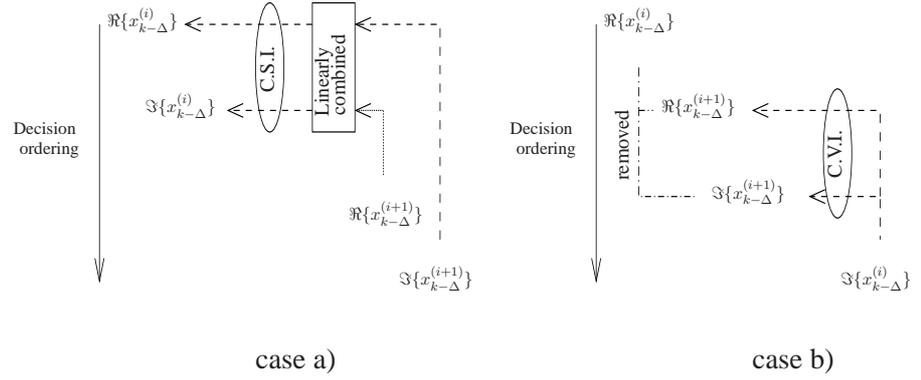
then the decision taken for  $\Re\{x_{k-\Delta}^{(i)}\}$  is not fed back to improve the estimation of  $\Im\{x_{k-\Delta}^{(i)}\}$ . Unfortunately, the condition (2.110) is not so intuitive. Let us investigate the special case where the input sequences are jointly circularly symmetric (i.e.,  $n_r = 0$  and  $E[x_k^{(i)} x_{k-\ell}^{(j)}] = \delta_{ij} \forall \ell, \forall i, j \in \{1, \dots, n_i\}$ ). By exploiting the special structure of the correlation matrices of circularly symmetric vectors [56], and by using the properties of the set of symmetric matrices reported in [44], it can be verified that the following relations hold:

$$\begin{aligned} \mathbf{M}_{11} &= \mathbf{M}_{22} & \mathbf{M}_{12} &= -\mathbf{M}_{12}^T \\ \mathbf{P}_{11} &= \mathbf{P}_{22} & \mathbf{P}_{12} &= -\mathbf{P}_{21}. \end{aligned} \quad (2.111)$$

In such a case, the condition (2.110) is verified<sup>7</sup>. It follows that the estimate of  $\Re\{x_{k-\Delta}^{(i)}\}$  does not improve the estimate of  $\Im\{x_{k-\Delta}^{(i)}\}$ , and it can be verified that if the real part and the imaginary one of each symbol  $x_{k-\Delta}^{(i)}$  are detected successively, then the WL-WDF-MMSE equalizer structure degenerates into the DF. On the other hand, it can be verified that the equality in (2.111) does not hold for every decision ordering and, consequently, the decision over  $\Re\{x_{k-\Delta}^{(i)}\}$  is fed back to improve the estimate of  $\Im\{x_{k-\Delta}^{(i)}\}$ , allowing one to consider the optimization over all the  $(2N_i)!$  permutation matrices. Such a behavior can be explained by studying the variations of the statistical property of the interference with respect to the decision ordering. In Fig. 2.3, we have considered two different decision ordering. Assume that  $x_{k-\Delta}^{(i)}$  is circularly symmetric  $\forall i$ . According to the case<sup>8</sup> *a*), the undetected components  $\Re\{x_{k-\Delta}^{(i+1)}\}$  and  $\Im\{x_{k-\Delta}^{(i+1)}\}$ , by means of the channel matrix weights ( $\mathcal{E}_0[\mathbf{H}_0]$ ), can be seen as a residual circularly symmetric interference when estimating the real and the imaginary parts of  $x_{k-\Delta}^{(i)}$ . Being both  $x_{k-\Delta}^{(i)}$  and the interference jointly circularly symmetric, the resulting processing is linear (see subsection 2.1), i.e.,  $\Re\{x_{k-\Delta}^{(i)}\}$  is not fed back to improve the estimate of  $\Im\{x_{k-\Delta}^{(i)}\}$ . On the other hand, when  $\Re\{x_{k-\Delta}^{(i)}\}$  is detected first (case *b*)), the feedback filter allows to remove the interference term due to  $\Re\{x_{k-\Delta}^{(i)}\}$  in the estimates of  $\Re\{x_{k-\Delta}^{(i+1)}\}$  and  $\Im\{x_{k-\Delta}^{(i+1)}\}$ . In such a case, the last component to be detected  $\Im\{x_{k-\Delta}^{(i)}\}$  can be seen as a residual rotationally variant interference when estimating  $\Re\{x_{k-\Delta}^{(i+1)}\}$  and  $\Im\{x_{k-\Delta}^{(i+1)}\}$ . It follows that the feedback filter performs

<sup>7</sup>If  $\mathbf{A}$  is a real-valued skew-symmetric matrix, then  $\mathbf{v}^T \mathbf{A} \mathbf{v} = 0$  for every real-valued  $\mathbf{v}$ .

<sup>8</sup>The case *a*) corresponds to the detection ordering such that the real part and the imaginary one of each symbol sequence  $x_{k-\Delta}^{(i)}$  are detected successively, i.e., the ordering such that (2.111) is verified.



**Figure 2.3:** Statistical properties of the interference for different decision ordering and complex-valued circularly symmetric transmitted sequences: C.S.I. stays for circularly symmetric interference, whereas C.V.I. stays for circularly variant interference.

a widely linear processing of  $\Re\{x_{k-\Delta}^{(i+1)}\}$  and  $\Im\{x_{k-\Delta}^{(i+1)}\}$ , i.e.,  $\Re\{x_{k-\Delta}^{(i+1)}\}$  is fed back to estimate  $\Im\{x_{k-\Delta}^{(i+1)}\}$ . The same reasoning applies to the case where  $n_r > 0$ .

The above results can be extended to any detection ordering by defining a vector  $\bar{\mathbf{r}}$  that is updated after each cancellation performed by the feedback tap  $\mathbf{B}_{E,0}$ . However, we do not consider such a case since it makes heavy the notation and, consequently, does not allow to simply expose the main concepts discussed here.

The results in this section generalize the ones of [58] for the SISO scenario, and formalize the ones of [59] for the MIMO non-dispersive scenario.

## 2.8 Nonequivalent WL-WDF equalizers

The WL-WDF equalizer structure (presented in Section 2.5) has been introduced by resorting to the real-valued representation ((2.16) and (2.18)) of the transmitted vector  $\mathbf{x}_k$  and the received one  $\mathbf{y}_k$ , as well as of the input vector of the feedback filter  $\mathcal{Q}(\mathcal{E}_{n_r}[\mathbf{x}_{k-\Delta}])$ . An analogous structure of the WL-WDF equalizer can be introduced by resorting to the complex-valued representation ((2.16) and (2.18)) of the involved vectors. The complex-valued counterpart of

the equalizer input-output relationship in (2.52) can be written as follows:

$$\hat{\mathbf{x}}_{C,k-\Delta} = \bar{\mathbf{C}}_{n_r} \left[ \mathbf{W}_C^H \mathbf{T}_{N_f} [\mathbf{C}_0[\mathbf{y}_k], k] - \mathbf{B}_C^H \mathcal{Q}(\mathbf{T}_{N_b+1}[\mathcal{C}_{n_r}[\hat{\mathbf{x}}_{C,k-\Delta}], k - \Delta]) \right]. \quad (2.112)$$

where  $\mathbf{W}_C$  and  $\mathbf{B}_C$  are complex-valued matrices of size  $(2N_o N_f) \times (2N_i - n_r)$  and  $(2N_i - n_r)(N_b + 1) \times (2N_i - n_r)$ , and where  $\hat{\mathbf{x}}_{C,k-\Delta}$  denotes the equalizer output. The first tap of the filter matrix  $\mathbf{B}_C$ , say  $\mathbf{B}_{C,0}$ , is subject to the constraint imposed by the equalization scenario. By means of (2.28), it is simple to show that, when Scenario 1 is the detection scenario, one has

$$\mathcal{C}_{n_r}[\hat{\mathbf{x}}_{C,k-\Delta}] = \mathbf{\Upsilon} \mathbf{T} \mathcal{E}_{n_r}[\hat{\mathbf{x}}_{C,k-\Delta}] \quad (2.113)$$

where  $\mathbf{\Upsilon}$  and  $\mathbf{T}$  are given by (2.25) and (2.24) (with  $n_1 = n_r$ ), respectively, i.e., a one-to-one correspondence exists between the two structures and, consequently, they perform equivalently.

On the other hand, when Scenario 2 is considered, the matrix taps  $\mathbf{B}_{E,0}$  and  $\mathbf{B}_{C,0}$  are constrained to be upper triangular. In such a case, the complex-valued counterpart of  $\mathbf{B}_{E,0}$  provided by (2.28) is not lower triangular in general and, *vice versa*, the real-valued counterpart of  $\mathbf{B}_{C,0}$  provided by (2.28) is not lower triangular. It follows that the real-valued equalizer structure and the complex-valued one are not equivalent and, consequently, they are expected to perform differently. Moreover, let us note that, when  $\mathbf{B}_{C,0}$  is upper triangular, the decision over the complex-valued symbol  $(x_{k-\Delta}^{(\ell)})^*$  can be fed back to improve the estimation of  $x_{k-\Delta}^{(\ell)}$ , and, *vice versa*, when a different ordering for the components of  $\mathcal{C}_{n_r}[\mathbf{x}_{k-\Delta}]$  is adopted, the decision over the complex-valued symbol  $x_{k-\Delta}^{(\ell)}$  can be fed back to improve the estimation of  $(x_{k-\Delta}^{(\ell)})^*$ . It follows that  $\mathbf{B}_{C,0}^H$  defines in general a WL transformation that can not be expressed in the form of (2.27): we call such transformation *generalized widely linear*.

A derivation analogous to the one proposed in Section 2.7 for the real-valued equalizer structure allows to determine the complex-valued counterpart of the condition (2.110), i.e., the condition that is verified when  $(x_{k-\Delta}^{(\ell)})^*$  is not fed back to improve the estimate  $x_{k-\Delta}^{(\ell)}$  (being  $(x_{k-\Delta}^{(\ell)})^*$  and  $x_{k-\Delta}^{(\ell)}$  the first two detected components of the vector  $\mathcal{C}_{n_r}[\mathbf{x}_{k-\Delta}](P)$ ). Specifically, accounting for the vector  $\bar{\mathbf{r}}$  given by (2.107), define the correlation and the pseudo-correlation matrices  $\mathbf{R}_{\bar{\mathbf{r}}\bar{\mathbf{r}}} \triangleq E[\bar{\mathbf{r}}\bar{\mathbf{r}}^H]$  and  $\mathbf{R}_{\bar{\mathbf{r}}\bar{\mathbf{r}}^*} \triangleq E[\bar{\mathbf{r}}\bar{\mathbf{r}}^T]$ , respectively. Moreover, define the cross correlation vectors  $\mathbf{p}_{\bar{\mathbf{r}}x_i} \triangleq E[\bar{\mathbf{r}}(x_{k-\Delta}^{(i)})^*]$  and  $\mathbf{p}_{\bar{\mathbf{r}}x_i^*} \triangleq E[\bar{\mathbf{r}}x_{k-\Delta}^{(i)}]$ . It can be verified that if

$$2 \mathbf{p}_{\bar{\mathbf{r}}x_i^*}^H \mathbf{R}_{11} \mathbf{p}_{\bar{\mathbf{r}}x_i} + \mathbf{p}_{\bar{\mathbf{r}}x_i}^T \mathbf{R}_{12}^* \mathbf{p}_{\bar{\mathbf{r}}x_i} + \mathbf{p}_{\bar{\mathbf{r}}x_i^*}^H \mathbf{R}_{12} \mathbf{p}_{\bar{\mathbf{r}}x_i^*}^* = 0, \quad (2.114)$$

where

$$\begin{aligned}\mathbf{R}_{11} &\triangleq [\mathbf{R}_{\bar{r}\bar{r}} - \mathbf{R}_{\bar{r}\bar{r}^*} \mathbf{R}_{\bar{r}\bar{r}}^{-*} \mathbf{R}_{\bar{r}\bar{r}^*}^*]^{-1} \\ \mathbf{R}_{12} &\triangleq -\mathbf{R}_{\bar{r}\bar{r}}^{-1} \mathbf{R}_{\bar{r}\bar{r}^*} \mathbf{R}_{11} \quad ,\end{aligned}$$

then the decision taken for  $(x_{k-\Delta}^{(i)})^*$  is not fed back to improve the estimation of  $x_{k-\Delta}^{(i)}$ . From (2.114), it follows that if the observation vector  $\bar{r}$  is rotationally variant ( $\mathbf{R}_{\bar{r}\bar{r}^*} \neq \mathbf{0}$ ), then the estimate of  $x_{k-\Delta}^{(i)}$  and the estimate of its conjugate version can be not equivalent, i.e. the optimum feedforward and feedback filters can define two *generalized widely linear* transformations of the received vector and the decision vector, respectively.

The condition (2.114) is always verified when the transmitted sequences are complex-valued circularly symmetric. In such a case the WL-WDF-MMSE equalizer structure obtained by adopting the complex-valued representation degenerates into the linear DF-MMSE one. On the other hand, when  $n_r > 0$ , the condition (2.114) does not hold also when the  $(N_i - n_r)$  complex-valued transmitted sequences are circularly symmetric.

## 2.9 Mismatching analysis

In this section, we derive a closed-form expression that allows one to assess performance degradation of the WL-WDF-MMSE equalizer due to the channel and/or (possibly) noise mismatch conditions. We denote the misestimated channel matrix and the misestimated noise-correlation matrix  $\mathbf{H}^{E^e}$  and  $\mathbf{R}_n^{E^e}$ , respectively. Channel and noise estimation errors give rise to an incorrect estimation of the input-correlation matrices: specifically, denoted with  $\nu^e$  the misestimated channel memory, we call  $\mathbf{R}_{x^{(r)}}^e$ ,  $\mathbf{R}_{x^{(rc)}}^e$ ,  $\hat{\mathbf{R}}_{x^{(rc)}}^e$ ,  $\mathbf{R}_{x^{(c)}}^e$ ,  $\hat{\mathbf{R}}_{x^{(cc)}}^e$ , and  $\hat{\mathbf{R}}_{x^{(c)}}^e$  the misestimated input-correlation matrices of size  $n_r(N_f + \nu^e) \times n_r(N_f + \nu^e)$ ,  $n_r(N_f + \nu^e) \times (N_i - n_r)(N_f + \nu^e)$  and  $(N_i - n_r)(N_f + \nu^e) \times (N_i - n_r)(N_f + \nu^e)$ . It follows that, by permutation of the augmented input and noise correlation matrices and augmented channel

matrix, (2.56)-(2.58) are modified as follows

$$\mathbf{R}_x^{A^e} \triangleq \begin{bmatrix} \mathbf{R}_{x^{(r)}}^e & \mathbf{R}_{x^{(rc)}}^e & \hat{\mathbf{R}}_{x^{(rc)}}^e \\ \mathbf{R}_{x^{(rc)}}^{eT} & \mathbf{R}_{x^{(c)}}^e & \hat{\mathbf{R}}_{x^{(cc)}}^e \\ \hat{\mathbf{R}}_{x^{(rc)}}^{eT} & \hat{\mathbf{R}}_{x^{(cc)}}^{eT} & \hat{\mathbf{R}}_{x^{(c)}}^e \end{bmatrix} \quad (2.115)$$

$$\mathbf{R}_n^{A^e} \triangleq \begin{bmatrix} \mathbf{R}_{n^{(c)}}^e & \hat{\mathbf{R}}_{n^{(cc)}}^e \\ \hat{\mathbf{R}}_{n^{(cc)}}^{eT} & \hat{\mathbf{R}}_{n^{(c)}}^e \end{bmatrix} \quad (2.116)$$

$$\mathbf{H}_A^e \triangleq \begin{bmatrix} \mathbf{H}_R^{(r)e} & \mathbf{H}_R^{(c)e} & -\mathbf{H}_I^{(c)e} \\ \mathbf{H}_I^{(r)e} & \mathbf{H}_I^{(c)e} & \mathbf{H}_R^{(c)e} \end{bmatrix} \quad (2.117)$$

where  $\mathbf{H}_R^{(r)e}$  and  $\mathbf{H}_I^{(r)e}$  are of size  $(N_o N_f) \times n_r(N_f + \nu^e)$  and where  $\mathbf{H}_R^{(c)e}$  and  $\mathbf{H}_I^{(c)e}$  are of size  $(N_o N_f) \times (N_i - n_r)(N_f + \nu^e)$ . Consequently, by substituting the input and noise correlation matrices and the channel matrices in (2.56)-(2.58) with the corresponding misestimated ones, the optimum  $(\mathbf{B}_\alpha^e, \mathbf{D}_\alpha^e)$  can be derived by using the procedure described in Section 2.6 from the steps 2)-6). Note that the misestimated channel memory  $\nu^e$  gives rise to a mismatched decision delay  $\Delta^e$  and to a mismatched order of the feedback filter  $N_b^e = N_f + \nu^e - \Delta^e - 1$ . The optimum feedforward filters  $(\mathbf{W}_\alpha^e, \mathbf{G}_\alpha^e)$  are determined by using in (2.85)-(2.86) the misestimated matrices defined in this section and taking into account for the misestimated decision delay  $\Delta^e$ .

The error-correlation matrix corresponding to the misestimated feedback filter  $\tilde{\mathbf{B}}_F^e \triangleq [\tilde{\mathbf{B}}_\alpha^{eH} \tilde{\mathbf{D}}_\alpha^{eH}]^H$ , with  $\tilde{\mathbf{B}}_\alpha^e$  and  $\tilde{\mathbf{D}}_\alpha^e$  defined according to (2.87), and to the feedforward filter  $\mathbf{W}_F^e \triangleq [\mathbf{W}_\alpha^{eH} \mathbf{G}_\alpha^{eH}]^H$  can be expressed in compact form as follows

$$\mathbf{R}_e^e = \tilde{\mathbf{B}}_F^{eH} \bar{\mathbf{R}}_x^A \tilde{\mathbf{B}}_F^e - \tilde{\mathbf{B}}_F^{eH} \bar{\mathbf{R}}_x^A \bar{\mathbf{H}}_A^T \mathbf{W}_F^e - \mathbf{W}_F^{eH} \bar{\mathbf{H}}_A \bar{\mathbf{R}}_x^A \tilde{\mathbf{B}}_F^e + \mathbf{W}_F^{eH} \mathbf{R}_y^A \mathbf{W}_F^e \quad (2.118)$$

where  $\mathbf{R}_y^A$  is the correlation matrix of the output vector  $\mathcal{E}_0 [\mathbf{T}_{N_f} [\mathbf{y}_k, k]]$ , and where  $\bar{\mathbf{R}}_x^A$ ,  $\bar{\mathbf{R}}_{x^{(r)}}$ ,  $\mathbf{H}_A$ , and  $\bar{\mathbf{H}}_R^{(r)}$  are defined below in (2.119), with  $\bar{\mathbf{R}}_{x^{(rc)}}$ ,  $\bar{\mathbf{R}}_{x^{(c)}}$ ,  $\bar{\mathbf{R}}_{x^{(cc)}}$ , and  $\bar{\mathbf{R}}_{x^{(c)}}$  defined in accordance with  $\bar{\mathbf{R}}_{x^{(r)}}$ , and with  $\bar{\mathbf{H}}_I^{(r)}$ ,  $\bar{\mathbf{H}}_R^{(c)}$ , and  $\bar{\mathbf{H}}_I^{(c)}$  defined accordingly to  $\bar{\mathbf{H}}_R^{(r)}$ .

$$\begin{aligned}
\bar{\mathbf{R}}_x^A &\triangleq \begin{bmatrix} \bar{\mathbf{R}}_{x(r)} & \bar{\mathbf{R}}_{x(rc)} & \bar{\hat{\mathbf{R}}}_{x(rc)} \\ \bar{\mathbf{R}}_{x(rc)}^T & \bar{\mathbf{R}}_{x(cc)} & \bar{\hat{\mathbf{R}}}_{x(cc)} \\ \bar{\hat{\mathbf{R}}}_{x(rc)}^T & \bar{\hat{\mathbf{R}}}_{x(cc)}^T & \bar{\hat{\mathbf{R}}}_{x(c)} \end{bmatrix} & (2.119) \\
\bar{\mathbf{R}}_{x(r)} &\triangleq \begin{cases} \begin{bmatrix} \mathbf{I}_{n_r(N_f+\nu^e)} \\ \mathbf{0}_{n_r(\nu-\nu^e)\times n_r(N_f+\nu^e)} \end{bmatrix}^T \mathbf{R}_{x(r)} \begin{bmatrix} \mathbf{I}_{n_r(N_f+\nu^e)} \\ \mathbf{0}_{n_r(\nu-\nu^e)\times n_r(N_f+\nu^e)} \end{bmatrix} & : \nu \geq \nu^e \\ \begin{bmatrix} \mathbf{R}_{x(r)} & \mathbf{0}_{n_r(N_f+\nu)\times n_r(\nu^e-\nu)} \\ \mathbf{0}_{n_r(\nu^e-\nu)\times n_r(N_f+\nu)} & \mathbf{0}_{n_r(\nu^e-\nu)\times n_r(\nu^e-\nu)} \end{bmatrix} & : \nu^e > \nu \end{cases} \\
\bar{\mathbf{H}}_A &\triangleq \begin{bmatrix} \bar{\mathbf{H}}_R^{(r)} & \bar{\mathbf{H}}_R^{(c)} & -\bar{\mathbf{H}}_I^{(c)} \\ \bar{\mathbf{H}}_I^{(r)} & \bar{\mathbf{H}}_I^{(c)} & \bar{\mathbf{H}}_R^{(c)} \end{bmatrix} \\
\bar{\mathbf{H}}_R^{(r)} &\triangleq \begin{cases} \mathbf{H}_R^{(r)} \begin{bmatrix} \mathbf{I}_{n_r(N_f+\nu^e)} \\ \mathbf{0}_{n_r(\nu-\nu^e)\times n_r(N_f+\nu^e)} \end{bmatrix} & : \nu \geq \nu^e \\ \begin{bmatrix} \mathbf{H}_R^{(r)} & \mathbf{0}_{n_o N_f \times n_r(\nu^e-\nu)} \end{bmatrix} & : \nu^e > \nu \end{cases}
\end{aligned}$$

## 2.10 Numerical results

In this section we present the performance analysis of the proposed equalizers. It will be shown that the adoption of WL filters in the design of the DF-based equalizer allows one to achieve considerable performance gain over the conventional structures based on linear filtering. Section 2.10.1 presents the results proposed in [44] with reference to the Scenario 1: the performances are evaluated in terms of achieved MSE, both in presence and absence of channel mismatching. In Section 2.10.2, some results proposed in [60] are presented with reference to the Scenario 2: the performance analysis is carried out in terms of both achieved MSE and symbol error rate. Since the real-valued representation and the complex-valued one are equivalent when WL-WDF-MMSE equalization is performed in Scenario 1, obviously, we will not refer to any specific structure in Section 2.10.1.

### 2.10.1 Numerical results in Scenario 1

In this section we present the performance analysis of the proposed equalizers. The performance is evaluated in terms of the signal-to-noise ratio at the

decision point defined as

$$\text{SNR} \triangleq \frac{\frac{1}{N_i(N_f+\nu)}\text{trace}(\mathbf{R}_x)}{\frac{1}{N_i}\text{trace}(\mathbf{R}_e^{(opt)})} \quad (2.120)$$

where  $\mathbf{R}_x$  has been defined in (2.36) and  $\mathbf{R}_e^{(opt)}$  is the error-correlation matrix, which, for WL-WDF-MMSE equalizer, is given by (2.88). The averaged SNR (ASNR) in dB is obtained by averaging over 100 independent trials: in each trial, each channel tap  $\mathbf{H}_m$  is randomly generated according to a complex zero-mean uncorrelated Gaussian random process with variance  $\sigma^2 = \frac{1}{\nu}$ , which assures unit-energy FIR filters. Moreover, the channel impulse-responses used in our experiment are FIR filters with taps  $\mathbf{H}_m$  ( $m = 0, \dots, \nu$ ) and  $\nu = 6$ .

Before presenting the performance results based on the mismatching analysis carried out in the previous section, we consider at first a set of experiments in an ideal scenario, namely the channel impulse-response is assumed to be exactly known and the assumptions (used to design the equalizers) of signals and noise spatially and temporally uncorrelated with known power are fully verified. Moreover, assuming unit powers for the input signals and  $n_r$  real-valued constellations, we assume the following input-correlations:

$$\begin{aligned} r_x^{(i,j)}(\ell) &\triangleq E \left[ x_{k+\ell}^{(i)} x_k^{(j)*} \right] = \delta_\ell \delta_{i-j} & (2.121) \\ \hat{r}_x^{(i,j)}(\ell) &\triangleq E \left[ x_{k+\ell}^{(i)} x_k^{(j)} \right] = \begin{cases} \delta_\ell \delta_{i-j} & i = 1, \dots, n_r \\ 0 & i = n_r + 1, \dots, N_i \end{cases} \end{aligned}$$

The noise is assumed complex white WSS Gaussian zero-mean processes with same power  $\sigma_n^2$  and spatially uncorrelated, namely:

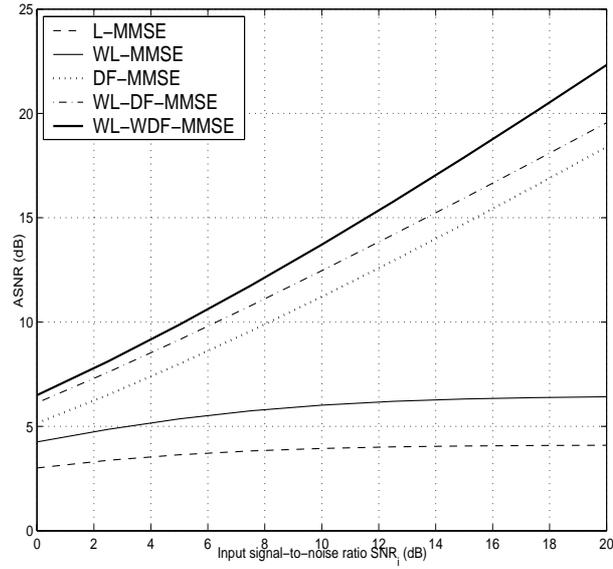
$$\begin{aligned} r_n^{(i,j)}(\ell) &\triangleq E \left[ n_{k+\ell}^{(i)} n_k^{(j)*} \right] = \sigma_n^2 \delta_\ell \delta_{i-j} & (2.122) \\ \hat{r}_n^{(i,j)}(\ell) &\triangleq E \left[ n_{k+\ell}^{(i)} n_k^{(j)} \right] = 0 \end{aligned}$$

According to (2.121) and (2.122), the input signal-to-noise ratio is defined as  $\text{SNR}_i \triangleq \frac{1}{\sigma_n^2}$ .

Unless not specified, in the following simulations, we consider a  $6 \times 6$  MIMO channel,  $n_r = 3$  rotationally variant transmitted sequences,  $N_f = 5$ , and the processing delay  $\Delta$  that optimizes the performance.

### A. Ideal scenario

In Fig. 2.4, the decision-point ASNRs of all the considered MMSE equalizers are plotted versus  $\text{SNR}_i$ . In such a scenario the WL-WDF equalizer outper-



**Figure 2.4:** ASNR of different equalizers versus  $\text{SNR}_i$ .

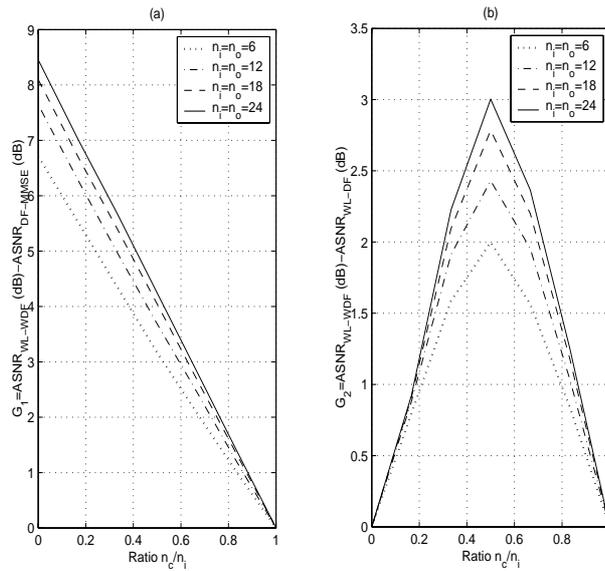
forms all the other equalizers. We have evaluated the performance gains of the WL-WDF equalizer over both DF and WL-DF equalizers defined as follows

$$G_1 \triangleq \text{ASNR}_{\text{WL-WDF}} (\text{dB}) - \text{ASNR}_{\text{DF}} (\text{dB}) \quad (2.123)$$

$$G_2 \triangleq \text{ASNR}_{\text{WL-WDF}} (\text{dB}) - \text{ASNR}_{\text{WL-DF}} (\text{dB}). \quad (2.124)$$

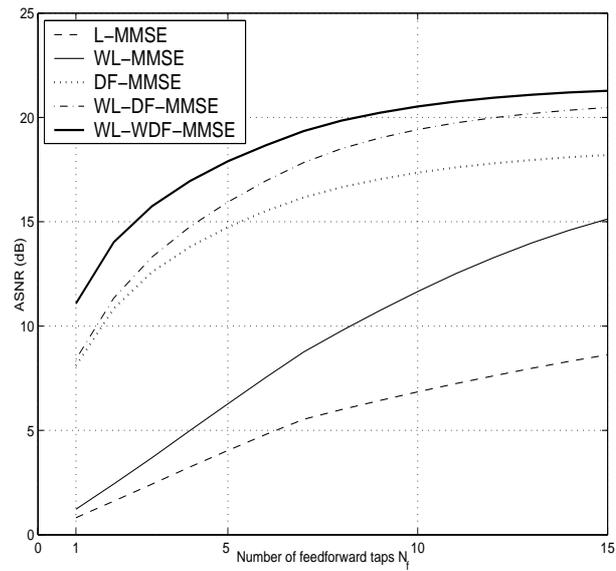
Figs. 2.5 (a)-(b) report  $G_1$  and  $G_2$  versus the ratio  $\frac{n_c}{N_i}$ , being  $n_c$  the number of circularly symmetric transmitted sequences, for different values of  $N_i = N_o$ . Note that  $G_1$  reaches the maximum value for  $\frac{n_c}{N_i} = 0$  and, as expected, the minimum one for  $\frac{n_c}{N_i} = 1$  since, in this case, the widely linear equalizers degenerate into the linear ones. The gain  $G_2$  is null, *vice versa*, when  $\frac{n_c}{N_i} = 0$  and  $\frac{n_c}{N_i} = 1$ : the former condition means that all the components of the transmitted vector are real-valued and so the WL-WDF equalizer degenerates into WL-DF one; the latter condition means that the transmitted vector is rotationally invariant and, therefore, both the WL-WDF and WL-DF equalizers degenerate into the DF equalizer. In the intermediate situation, i.e.,  $\frac{n_c}{N_i} = \frac{1}{2}$ , the WL-WDF equalizer exhibits the maximum performance gain. The results show that the performance gains are approximately independent of  $N_i$ .

Fig. 2.6 shows the behaviors of the decision-point ASNRs of the different

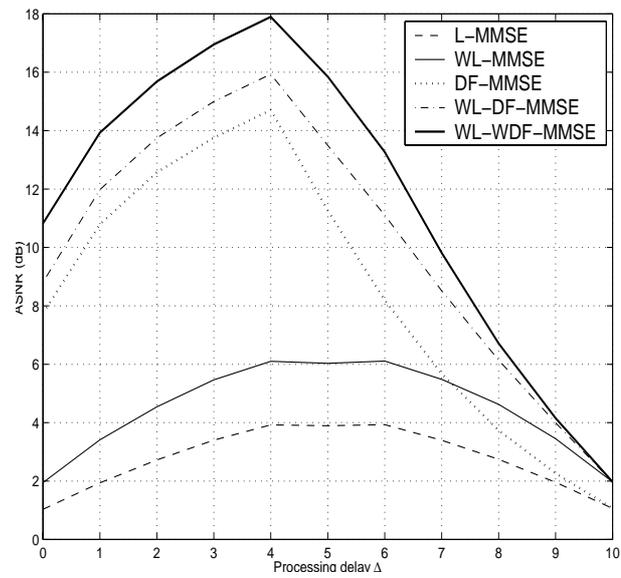


**Figure 2.5:** Performance gain of the WL-WDF-MMSE equalizer over DF-MMSE (a) and WL-DF-MMSE (b) equalizers versus the ratio  $\frac{n_c}{N_i}$  for different values of  $N_i = N_o$ .

equalizers versus  $N_f$  when  $\text{SNR}_i = 15\text{dB}$ . We can notice that the WL-WDF equalizer provides satisfactory performance also when  $N_f$  is small and the WL-DF equalizer outperforms the DF one for larger values of  $N_f$ . As expected, all the decision-feedback equalizers outperform the WL and L ones at the expense of an additional computational complexity. Finally, we investigate the dependence on the processing delay  $0 \leq \Delta \leq (N_f + \nu - 1)$  of the decision-point ASNR. The results of Fig. 2.7 show that the performances of the WL and L equalizers are very sensitive to the variations of  $\Delta$ . Such a behavior is due to the fact that, when  $\Delta$  is too small, the equalizers cannot satisfactorily perform anticausal processing; on the other hand, when  $\Delta$  is too large, the equalizer cannot satisfactorily perform causal processing. Moreover, the performances of all the DF equalizers are more sensitive to variations of  $\Delta$ . The results in Fig. 2.7 agree with those, reported in [61, 62], where it is suggested to use  $\Delta = N_f - 1$  for SISO DF equalizer. Note that, when  $\Delta = N_f + \nu - 1$ , then,  $N_b = 0$ : it follows that the WL-WDF and the WL-DF equalizers degenerate into the WL one, and, similarly, the DF equalizer degenerates into the linear equalizer.



**Figure 2.6:** Effects of the number of feedforward taps  $N_f$ .



**Figure 2.7:** Variation of decision-point ASNR with processing delay  $\Delta$ .

### B. Mismatched scenario

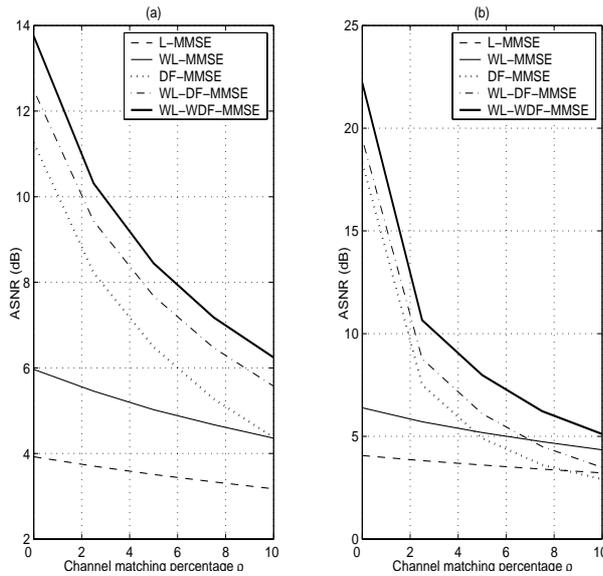
Channel-mismatch effects can manifest themselves in several forms including errors affecting the channel impulse response coefficients and/or errors in determining the channel order. In our numerical experiments, the effects of channel mismatch are modelled by assuming that all the estimated coefficients  $h_k^{(i,j)^e}$  of the channel impulse response are defined as

$$h_k^{(i,j)^e} \triangleq h_k^{(i,j)} + \varepsilon_k^{(i,j)} \quad (2.125)$$

with  $\varepsilon_k^{(i,j)}$  zero-mean Gaussian random variables, uncorrelated each other, with variance:

$$\sigma_\varepsilon^2 \triangleq \frac{E \left[ \left| \varepsilon_k^{(i,j)} \right|^2 \right]}{E \left[ \left| h_k^{(i,j)} \right|^2 \right]} . \quad (2.126)$$

In Fig. 2.8, the decision-point ASNRs of the equalizers under consideration are plotted versus the channel mismatching percentage defined as  $\rho \triangleq 100\sigma_\varepsilon^2$

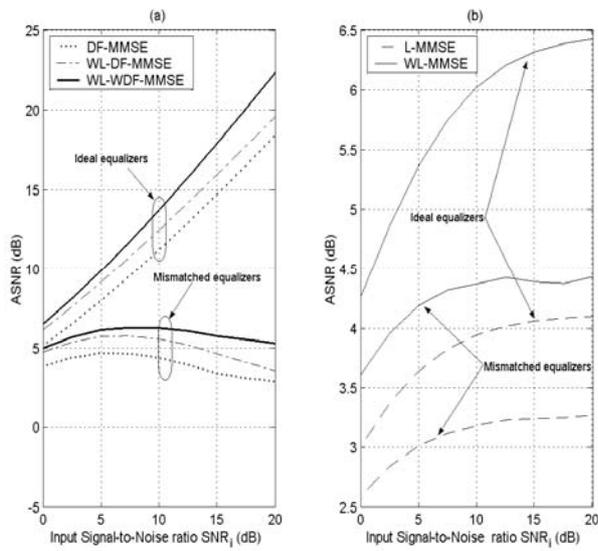


**Figure 2.8:** Effect of channel mismatching percentage on decision-point ASNR.

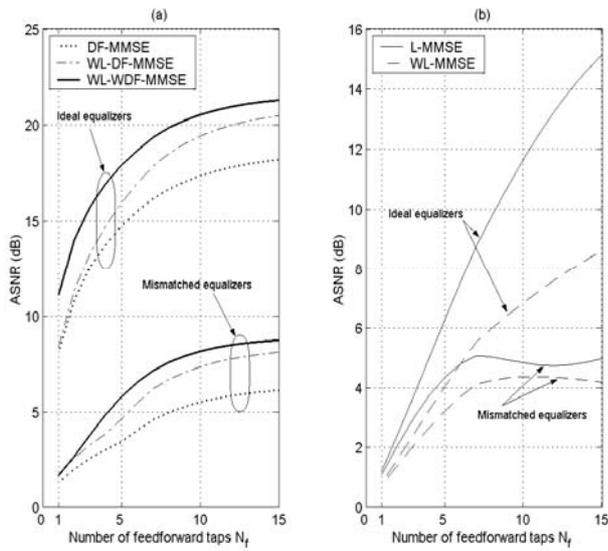
for  $\text{SNR}_i = 10\text{dB}$  (a) and  $\text{SNR}_i = 20\text{dB}$  (b). The results show that, unlike feedforward-based WL and L equalizers, the decision-feedback equalizers are very sensitive to channel mismatch: such a behavior, which agrees with the results derived in [61] for a special case of SISO channel, is due to the incapability of the feedback filter to correctly compensate the interference originated by the past samples. Moreover, DF-based equalizers appear more sensitive to channel mismatch for large values of  $\text{SNR}_i$  where, in the presence of severe mismatching, L equalizer outperforms WL equalizer, especially when  $n_r = N_i$ .

In all the following experiments we have set the channel mismatching percentage at 10% to assure that the decision-feedback equalizers perform as well as the non decision-feedback ones. Note that the considered mismatching is larger than the value achieved by many channel estimation techniques and, therefore, it represents the worst case in many realistic scenarios. Figs. 2.9 (a)-(b) report the decision-point ASNRs versus  $\text{SNR}_i$ . While the performances of the ideal equalizers increase with  $\text{SNR}_i$ , this is no longer true when the channel impulse response is misestimated. Also in this scenario, the WL and L equalizers appear more tolerant to the channel mismatch (note that, for the sake of representation, a different scale is used in Fig. 2.9(b)).

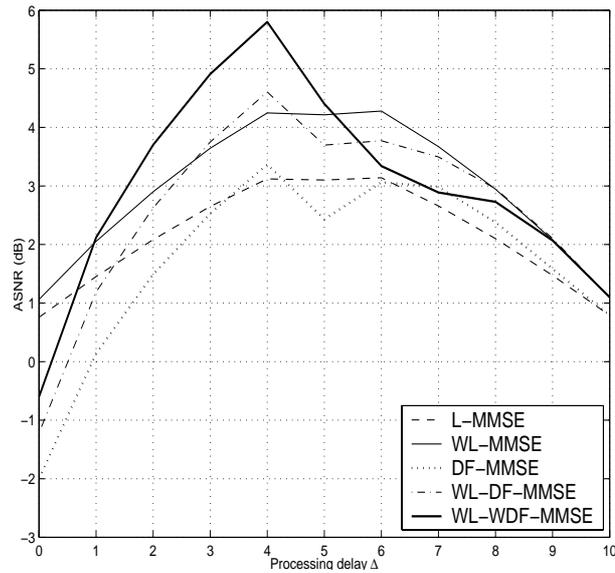
In Figs. 2.10 (a)-(b), the performances of the decision-feedback equalizers and the non decision-feedback ones, respectively, are evaluated for different values of  $N_f$  with  $\text{SNR}_i = 15\text{dB}$ . The results show that, unless a large ASNR offset is present, the behaviors of the equalizers operating in ideal and mismatched scenarios are the same. Finally, in Fig. 2.11, we analyze the behaviors of the equalizers under consideration with respect to the processing delay  $\Delta$  ( $\text{SNR}_i = 15\text{dB}$ ). The results show that the decision-feedback equalizers are more sensitive to the variations of  $\Delta$  with respect the non decision-feedback ones.



**Figure 2.9:** Effect of channel mismatch on decision-point ASNR versus  $SNR_i$ .



**Figure 2.10:** Effect of channel mismatch on decision-point ASNR versus  $N_f$ .



**Figure 2.11:** Effect of channel mismatch on decision-point ASNR versus  $\Delta$ .

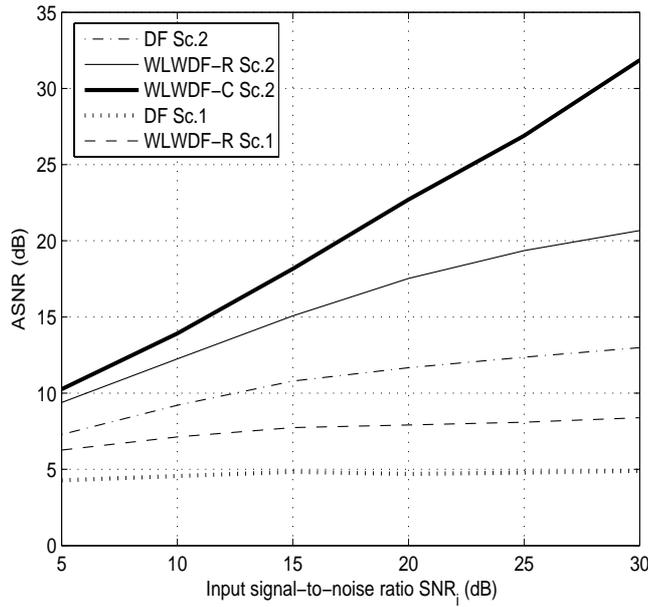
### 2.10.2 Numerical results in Scenario 2

We first consider the case where the optimum decision ordering is adopted and no error propagation is present. Then, we consider the case where the optimum ordering is still available, but the effects of error propagation in DF equalization are taken into account. In order to obtain accurate symbol error rate estimates, the computer simulations have required a large amount of time. In this study, we admit that at least 10 errors would occur for the lowest level of symbol error rate, resulting in a 95% confidence interval [63].

#### A. DF equalization in absence of error propagation

The performances are evaluated in terms of the SNR defined in (2.120) and optimized with respect to  $\Delta$  and  $\mathbf{P}$ . The ASNR in dB is obtained by averaging over 100 independent trials: in each trial, the channel taps  $h_m^{(i,\ell)}$  ( $i = 1, \dots, N_i$ ,  $\ell = 1, \dots, N_o$ ) are randomly generated according to a complex zero-mean white Gaussian random process with unitary variance and uncorrelated with each other. Unless not specified, in the following simulations, we assume:

- $N_i = 4$  input sequences with correlations defined in (2.121), such that  $n_r = 2$ ;



**Figure 2.12:** ASNRs of the equalizers versus  $\text{SNR}_i$  ( $N_o = 2$ ).

- $n_c = N_i - n_r$  complex-valued circularly symmetric input sequences;
- $N_o$  noise with correlation defined in (2.122);
- channel order  $\nu = 1$ ;
- $N_f = 4$  feedforward matrix taps.

Moreover, we denote the WL-WDF-MMSE equalizer obtained by adopting the complex-valued representation with WLWDF-C, while WLWDF-R will denote the WL-WDF-MMSE equalizer obtained by adopting the real-valued representation. In the following, the abbreviations Sc. 1 and Sc. 2 will denote Scenario 1 and Scenario 2, respectively.

In Fig. 2.12, the ASNRs at the decision point of the ideal WL-WDF-MMSE equalizers and the ideal conventional DF-MMSE equalizer are plotted versus  $\text{SNR}_i \triangleq \frac{1}{\sigma_n^2}$  for  $N_o = 2$ . The WLWDF-C significantly outperforms all the other equalizers in this ideal conditions. Then, to analyze the performance dependence on the number of outputs, we have evaluated the performance relative gain  $G_1$  of the WLWDF-C equalizer over the WLWDF-R one,

the performance relative gain  $G_2$  of the WLWDF-R equalizer over the DF-MMSE equalizer in both Scenario 1 and Scenario 2, the performance relative gain  $G_3$  of the WLWDF-R equalizer in Scenario 2 over the same equalizer in Scenario 1, and, finally, the performance relative gain  $G_4$  of the DF equalizer in Scenario 2 over the same equalizer in Scenario 1. More specifically:

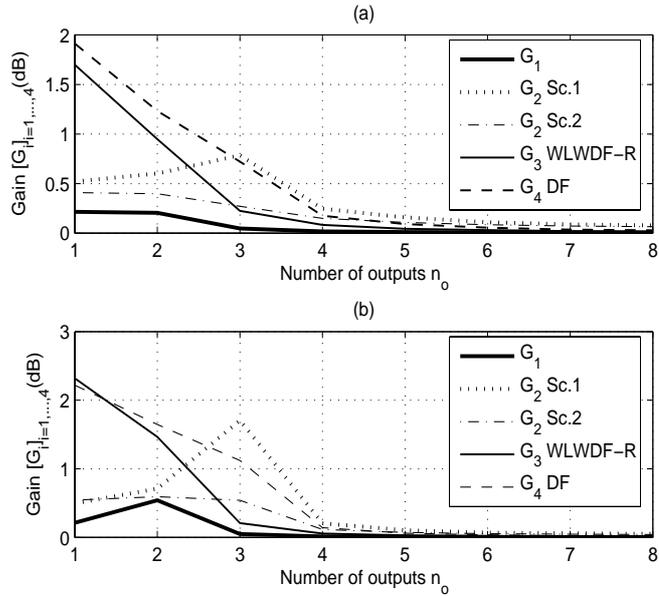
$$G_1 \triangleq \frac{\text{ASNR}_{\text{WLWDF-C}}(\text{dB}) - \text{ASNR}_{\text{WLWDF-R}}(\text{dB})}{\min \{ \text{ASNR}_{\text{WLWDF-C}}(\text{dB}), \text{ASNR}_{\text{WLWDF-R}}(\text{dB}) \}} \quad (2.127)$$

$$G_2 \triangleq \frac{\text{ASNR}_{\text{WLWDF-R}}(\text{dB}) - \text{ASNR}_{\text{DF}}(\text{dB})}{\text{ASNR}_{\text{DF}}} \quad (2.128)$$

$$G_3 \triangleq \frac{\text{ASNR}_{\text{WLWDF-R}}^{(\text{Sc.2})}(\text{dB}) - \text{ASNR}_{\text{WLWDF-R}}^{(\text{Sc.1})}(\text{dB})}{\text{ASNR}_{\text{WLWDF-R}}^{(\text{Sc.1})}(\text{dB})} \quad (2.129)$$

$$G_4 \triangleq \frac{\text{ASNR}_{\text{DF}}^{(\text{Sc.2})}(\text{dB}) - \text{ASNR}_{\text{DF}}^{(\text{Sc.1})}(\text{dB})}{\text{ASNR}_{\text{DF}}^{(\text{Sc.1})}(\text{dB})} \quad (2.130)$$

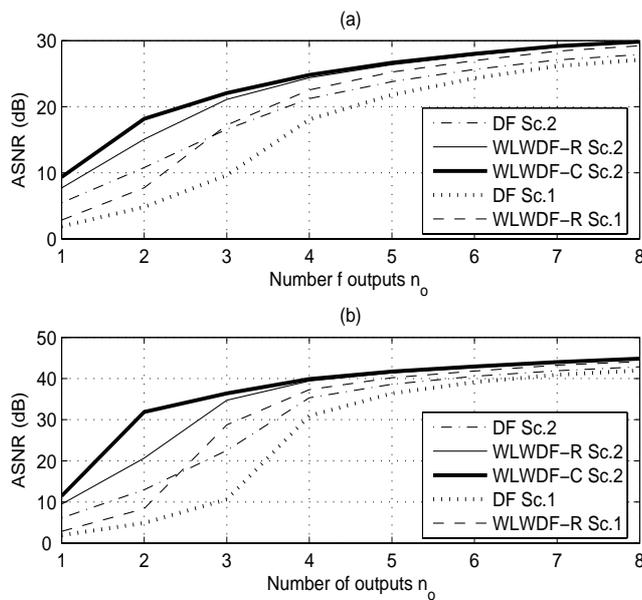
where  $G_2$  will be evaluated for both Scenario 1 and Scenario 2. Fig. 2.13 (a)



**Figure 2.13:** Gains  $[G_i]_{i=1,\dots,4}$  versus the number of outputs  $n_o$  for  $\text{SNR}_i = 15\text{dB}$  (a) and  $\text{SNR}_i = 30\text{dB}$  (b).

and (b) report  $[G_i]_{i=1,\dots,4}$  versus the number of channel outputs  $N_o$  for  $\text{SNR}_i = 15\text{dB}$  (a) and  $\text{SNR}_i = 30\text{dB}$  (b), respectively. Note that the WLWDF-C equalizer reaches its largest gain  $G_1$  over the WLWDF-R equalizer for  $N_o < N_i$ ; moreover,  $G_1$  exhibits the maximum value for  $N_o = 2$  in presence of large  $\text{SNR}_i$ . Similarly, the gain  $G_2$  of the WLWDF-R equalizer over the DF is different from zero for  $N_o < N_i$  both in Scenario 1 and in Scenario 2; more specifically,  $G_2$  exhibits its maximum for  $N_o = 3$  when Scenario 1 is considered, while it is constant for  $N_o < N_i$  when Scenario 2 is considered. The performance improvement gained by the WL processing, for fixed  $N_i$  and  $N_o$ , is due to a better exploitation of the statistical redundancy exhibited by the useful signal component. As expected, similarly to  $G_1$ , also  $G_2$  approaches to zero when  $N_o$  increases. Finally, the gains  $G_3$  and  $G_4$  monotonically decrease with  $N_o$ . The feedback of current decisions (Scenario 2) allows one to achieve a large gain over the equalizer structures of the Scenario 1 when  $N_o$  is lower than  $N_i$  since they guarantee the capability to discriminate the  $N_i$  inputs.

In Fig. 2.14 (a)-(b), a performance analysis of all the considered equalizers



**Figure 2.14:** ASNRs of the equalizers versus  $N_o$  for  $\text{SNR}_i = 15\text{dB}$  (a) and  $\text{SNR}_i = 30\text{dB}$  (b).

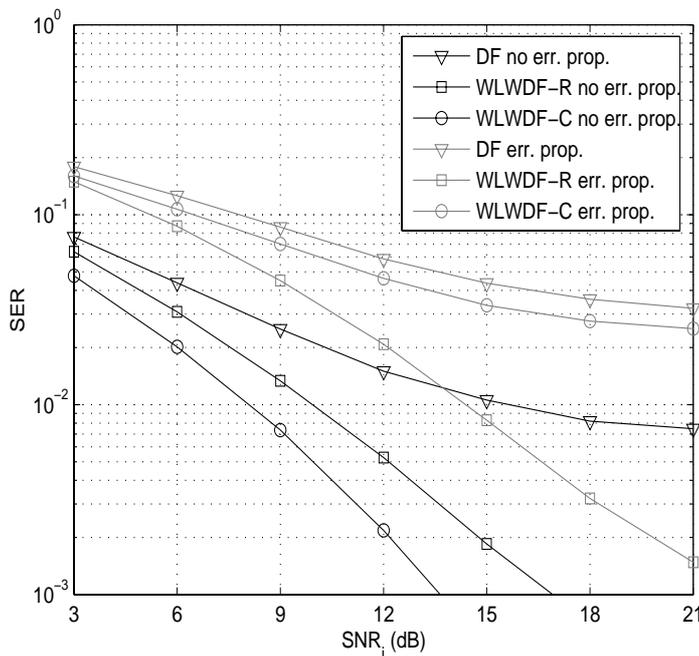
is also presented where the ASNRs are plotted versus  $N_o$  for  $\text{SNR}_i = 15\text{dB}$  and  $\text{SNR}_i = 30\text{dB}$ , respectively. The simulation results show a different sensitivity of the ASNR's curves of the different equalizers to the growth of the number of channel outputs. Specifically, the WL equalizers appear more sensitive with respect to the linear equalizers to the variations of  $N_o$ , especially for  $N_o < N_i$ . It follows that the WL processing allows one to efficiently exploit the diversity gain due to the multiple observations and it provides significant performance improvements when a small number of channel outputs is available at the receiver (typical down-link scenarios). Note the performance equivalence of the different equalizers when  $N_o$  is sufficiently large.

### ***B. DF equalization in presence of error propagation***

In real word scenarios the performances of the DF-based equalizers are also affected by the presence of the error propagation; its negative effects in DF-based MIMO equalization have been studied in [49], where it has been shown that they can be reduced by properly re-ordering the decision vector. In the following we analyze by computer simulations the effects of the error propagation when the WL-WDF-MMSE is employed at the receiver side. More specifically, we single out important differences between the complex-valued representation and the real-valued one when Scenario 2 is considered.

We have considered a  $1 \times 2$  MISO channel model (Fig. 2.15) and a  $2 \times 2$  one (Fig. 2.16). BPSK and 4-QAM modulations are assumed to be employed for the real-valued input sequence and the complex-valued circularly symmetric one (i.e.,  $n_r = 1$  and  $n_c = 1$ ). The number of the feedforward matrix taps is set to  $N_f = 2$  and all the equalizer parameters are chosen according to the MMSE criterion.

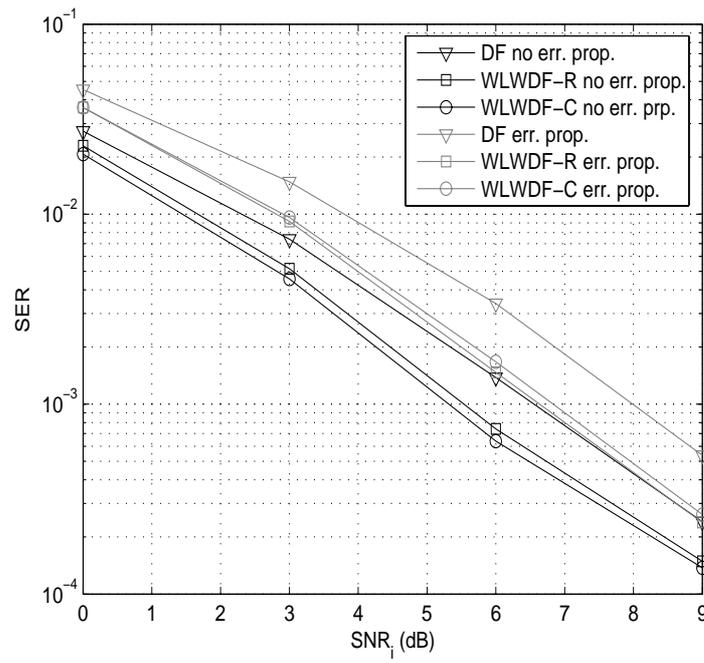
In presence of error propagation we have adopted, as performance measure, the symbol error rate (SER) averaged over the  $N_i$  inputs; moreover, the SER curves have been averaged over 100 independent channel realizations. In Fig. 2.15 and 2.16, the SER of the considered equalizers in Scenario 2 is plotted versus  $\text{SNR}_i$  both in absence and in presence of error propagation: in correspondence of each value of the  $\text{SNR}_i$ , the equalizer performances are optimized (with an exhaustive procedure) over all the possible decision orderings, i.e. over all the  $(N_i + n_c)!$  permutation matrices ( $N_i!$  for DF-MMSE equalizer). The reported SER curves show that the WLWDF-C equalizer, which outperforms all the other equalizers when correct decisions feed the feedback filter (see the black lines), can perform very poorly (see the grey lines) in the presence of error propagation. In such a scenario the WLWDF-R equalizers



**Figure 2.15:** SERs of the equalizers versus SNR;  $2 \times 1$  MISO channel.

outperforms all the other equalizers.

The results of such analysis are dramatically different from those obtained in the previous subsection where the error propagation effects were not taken into account. In fact, the WLWDF-C equalizer is able to utilize the decision over the conjugate version of  $x_{k-\Delta}^{(i)}$  to improve the estimate of  $x_{k-\Delta}^{(i)}$  (and *vice versa*). However, such an improvement holds only when  $\hat{x}_{k-\Delta}^{(i)} = x_{k-\Delta}^{(i)}$ , i.e. when the estimation error is enough small to allow to achieve a correct decision. It follows that the achieved accuracy improvement in estimating  $x_{k-\Delta}^{(i)}$  (correspondent to an increase in the SNR at the decision point) does not reduce the probability of error.



**Figure 2.16:** SERs of the equalizers versus  $SNR_i$ ;  $2 \times 2$  MIMO channel.

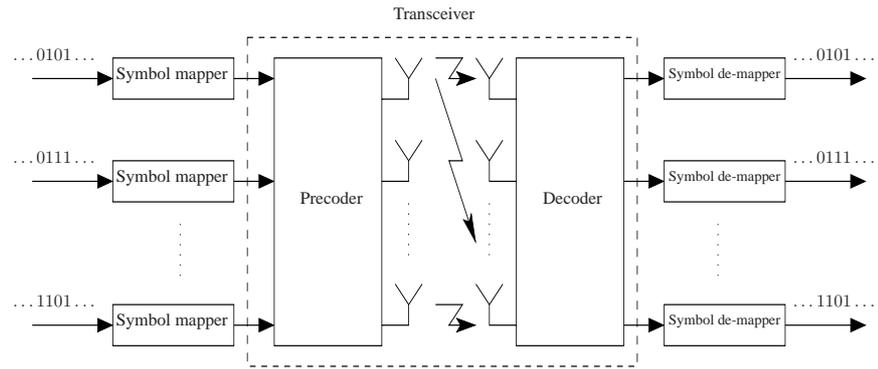
## Chapter 3

# Transceiver design

This chapter considers the design of MIMO communication systems with channel knowledge at both the transmitter and receiver side. By exploiting the CSI, the system can adapt itself to each channel realization to improve the spectral-efficiency and/or communication reliability. As shown in Chapter 1, the optimum transceiver, which maximizes the mutual information between the system input and output, diagonalizes the MIMO channel: ideal Gaussian codes have to be transmitted through the *eigen subchannels* with a water-filling power allocation procedure [18].

In practice, the system is divided into an inner subsystem and an outer one, and the ideal Gaussian codes are replaced by the symbol constellations of finite dimension. The former performs channel coding in order to gain error correction capabilities, whereas the latter transmits symbols belonging to specific signal constellations. For mathematical tractability, the overall system design problem is treated by separately addressing the two subsystem design problems. Let us concentrate on the second subsystem, i.e., the uncoded one. Such a subsystem can be further divided into two parts: the constellation mapping and the signal processing. The constellation mapping refers to how the data bits are mapped into points of a constellation, whereas the signal processing refers to any additional processing in the form of precoding at the transmitter and equalization at the receiver, that transforms the channel into an equivalent one.

We will focus on the signal processing part, i.e., we consider the optimization of the pair of transformations of blocks of the transmit symbols and receive samples, namely the precoder and the decoder, that operate on the time and space dimensions. In the existing literature, the linear transceiver, i.e.,



**Figure 3.1:** The transceiver architecture.

the one that employs linear precoder and decoder, have been widely studied [23, 25, 27, 64]. In this chapter, we point out that if the symbols (to be transmitted) are known to be circularly variant, then a solution based on the assumption of circular symmetric signals is not optimal. Therefore, the transceiver design is generalized to the case where WL filters, which exhibit significant performance improvements over conventional linear filters in presence of rotationally variant inputs, are utilized as precoder and decoder. It is shown that, in presence rotationally variant transmitted symbols, the proposed transceiver structure outperforms the linear-filtering based one in terms of mean square error and symbol error rate.

### 3.1 System Model

Fig. 3.1 shows the MIMO communication system model already considered in Section 1.8. At the transmitter side, the information bit streams are encoded to generate the information symbol streams that are processed by the precoder. The precoder output is then transmitted over the MIMO channel with  $N_i$  inputs and  $N_o$  outputs. At the receiver side, the channel outputs are processed by the decoder whose outputs are decoded to recover an estimate of the information bit streams. In the following, with reference to such a system model, we assume that specific signal constellations have been selected for all the information symbol streams that must be detected by the receiver, i.e. we focus on the boxed section of the communication system shown in Fig. 3.1.

Under the assumption that channel state information (CSI) is available at

both ends of the link, the precoder and the decoder can be jointly designed, according to a chosen optimization criterion (see [23, 25, 27, 64, 65], and references therein). In previous works, linear filters have been extensively employed as precoder and decoder: the resulting transceiver structure is referred in the following to as linear transceiver (LT). On the other hand, WL filters, which exhibit significant performance improvements over conventional linear filters in presence of circularly variant inputs (as shown in the previous chapter), have been extensively proposed in the literature (e.g., [41, 43, 44, 46]) but with exclusion of the transceiver design. For such a reason, in this chapter, we propose to employ WL filters as precoder as well as decoder (we will refer to as widely linear transceiver (WLT)) in the general case where symbol streams can belong to different symbol alphabets.

Since the WL processing can be performed by adopting the real-valued representation of the transmitted and received vectors (see Chapter 2), as well as the complex-valued one, in the two next subsections we introduce the two system models according to such representations. Although they are equivalent, we will recognize that the real-valued representation allows us to design the WLT by utilizing the LT design method already proposed in the literature.

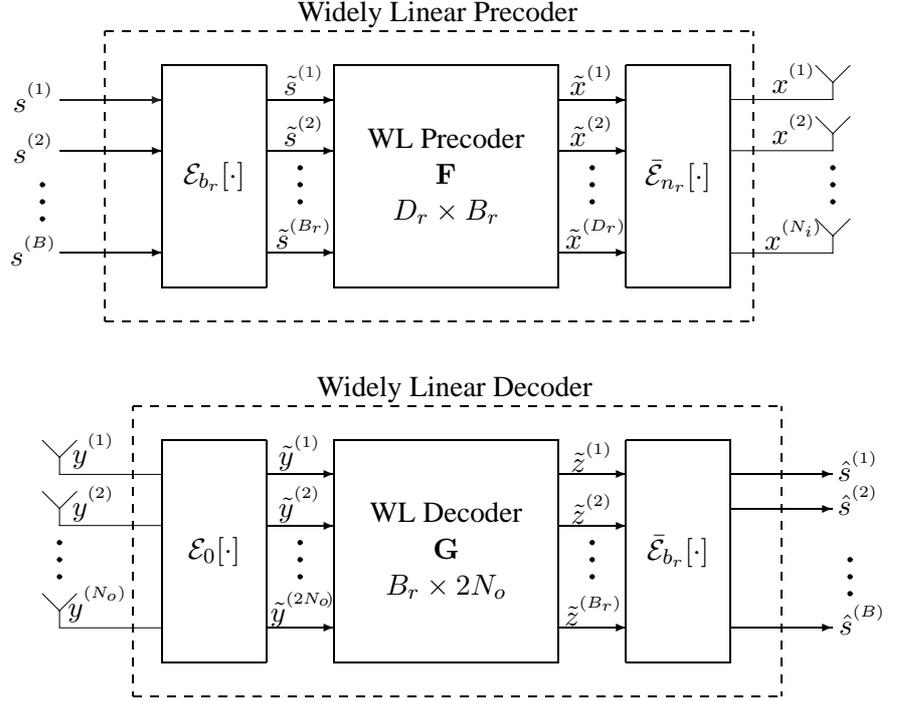
### 3.1.1 Real-valued system model

Consider a baseband equivalent noisy non-dispersive MIMO channel. The input-output relationship, as shown in Section 1.3, is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad , \quad (3.1)$$

where  $\mathbf{H}$  is the channel matrix, and  $\mathbf{n}$  denotes the noise vector assumed to be independent of  $\mathbf{s}$ , and where we have removed the time-index  $k$  for convenience. In this chapter, our focus is on the frequency flat environment since it allows us to provide a performance analysis of the proposed transceiver structures by utilizing the asymptotic results (Theorem 1.2) of the random matrix theory. However, the more general time-dispersive MIMO channel can be analogously treated by resorting to the augmented model of (1.73) (see also [25] for details).

By resorting to the operators (2.16) and (2.18), we can simply define the WLT structure as shown in Fig. 3.2. The symbol vector to be transmitted is denoted with  $\mathbf{s} \triangleq [s^{(1)}, s^{(2)}, \dots, s^{(B)}]^T$ : each symbol  $s^{(i)}$  is drawn from the constellation  $S_i$  ( $i = 1, \dots, B$ ) and, with no loss of generality, we consider both the real-valued constellations (e.g., PAM) and complex-valued constellations (e.g., MPSK with  $M \geq 4$  and QAM). We order the symbol sequences



**Figure 3.2:** Block diagram of the widely linear transceiver.

so that the real-valued constellations have indices  $i \in \{1, \dots, b_r\}$ ; the case  $b_r = 0$  accounts for the absence of real-valued constellations, whereas  $b_r = B$  is the case of all real-valued constellations.

By linearly combining the outputs of the filters that separately process the real and the imaginary parts of  $\mathbf{s}$ , the WL precoder provides the real and the imaginary parts of the channel-input vector  $\mathbf{x}$  of size  $N_i$ , whose first  $n_r$  components are assumed to be real-valued (the case  $n_r = 0$  accounts for the absence of real-valued components, whereas  $n_r = N_i$  is the case of all complex-valued components). Let us note that the number of real-valued components of  $\mathbf{x}$  represents a degree of freedom in the WLT design. The operators (2.17) and (2.18), together with  $\bar{\mathcal{E}}_p[\cdot]$  defined in (2.23), allow us to simply describe the WL precoding by utilizing a linear transformation  $\mathbf{F} \in \mathbb{R}^{D_r \times B_r}$ , with  $B_r \triangleq 2B - b_r$ , and  $D_r \triangleq 2N_i - n_r$ , from the aug-

mented vector  $\tilde{\mathbf{s}} \triangleq [\tilde{s}^{(1)}, \tilde{s}^{(2)}, \dots, \tilde{s}^{(B_r)}]^T = \mathcal{E}_{b_r} [\mathbf{s}]$  to the augmented vector  $\tilde{\mathbf{x}} \triangleq [\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(D_r)}]^T = \mathcal{E}_{n_r} [\mathbf{x}]$ . More specifically, one has:

$$\begin{aligned} \mathbf{x} &\triangleq [x^{(1)} \ x^{(2)} \ \dots \ x^{(N_i)}]^T \\ &= \bar{\mathcal{E}}_{n_r} [\mathbf{F} \mathcal{E}_{b_r} [\mathbf{s}]]. \end{aligned} \quad (3.2)$$

At the receiver side, the output  $\hat{\mathbf{s}}$  of the decoder is obtained by WL processing the received vector  $\mathbf{y}$ , to provide independent estimations of the real and the imaginary parts of  $\mathbf{s}$ . Hence,  $\hat{\mathbf{s}}$  can be written as follows:

$$\hat{\mathbf{s}} \triangleq \bar{\mathcal{E}}_{b_r} [\mathbf{G} \mathcal{E}_0 [\mathbf{y}]] \quad (3.3)$$

where  $\mathbf{G} \in \mathbb{R}^{B_r \times (2N_o)}$  is a linear transformation from the augmented vector  $\tilde{\mathbf{y}} \triangleq [\tilde{y}^{(1)}, \tilde{y}^{(2)}, \dots, \tilde{y}^{(2N_o)}]^T = \mathcal{E}_0 [\mathbf{y}]$  to the augmented vector  $\tilde{\mathbf{z}} \triangleq [\tilde{z}^{(1)}, \tilde{z}^{(2)}, \dots, \tilde{z}^{(B_r)}]^T = \mathcal{E}_{b_r} [\hat{\mathbf{s}}]$ .

Finally, let us note that, since the input-output relation (3.1) can be replaced by the equivalent one

$$\mathcal{E}_0 [\mathbf{y}] = \tilde{\mathcal{E}}_{n_r} [\mathbf{H}] \mathcal{E}_{n_r} [\mathbf{x}] + \mathcal{E}_0 [\mathbf{n}] \quad \leftrightarrow \quad \tilde{\mathbf{y}} \triangleq \tilde{\mathbf{H}} \tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \quad (3.4)$$

we can define the overall WLT system equation as follows:

$$\tilde{\mathbf{z}} = \mathbf{G} \tilde{\mathbf{H}} \tilde{\mathbf{F}} \tilde{\mathbf{s}} + \mathbf{G} \tilde{\mathbf{n}}, \quad (3.5)$$

where  $\tilde{\mathbf{z}}$  is utilized to recover the estimate  $\hat{\mathbf{s}}$  of  $\mathbf{s}$  by means of  $\hat{\mathbf{s}} = \bar{\mathcal{E}}_{b_r} [\tilde{\mathbf{z}}]$ . The system model in (3.5), as it will be shown in Section 3.3, allows us to synthesize the WLT, i.e. the matrix filters  $\mathbf{F}$  and  $\mathbf{G}$ , by utilizing the procedures already proposed in [25, 27] with reference to the LT design.

### 3.1.2 Complex-valued system model

By resorting to the operators (2.17) and (2.19), the WLT structure can be defined according to the complex-valued representation of the transmitted and received vectors. It is simple to verify that the system equation (3.5) is equivalent to the system equation:

$$\tilde{\mathcal{Z}} = \mathcal{G} \tilde{\mathcal{C}}_{n_r} [\mathbf{H}] \mathcal{F} \mathcal{C}_{b_r} [\mathbf{s}] + \mathcal{G} \mathcal{C}_0 [\mathbf{n}], \quad (3.6)$$

where, according to (2.27), the precoder matrix  $\mathcal{F} \in \mathbb{C}^{D_r \times B_r}$  is structured as follows

$$\mathcal{F} = \begin{bmatrix} \mathcal{F}_{1,1} & \mathcal{F}_{1,2} & \mathcal{F}_{1,2}^* \\ \mathcal{F}_{2,1} & \mathcal{F}_{2,2} & \mathcal{F}_{2,3} \\ \mathcal{F}_{2,1}^* & \mathcal{F}_{2,3}^* & \mathcal{F}_{2,2}^* \end{bmatrix} \quad (3.7)$$

with  $\mathcal{F}_{1,1} \in \mathbb{C}^{n_r \times b_r}$ ,  $\mathcal{F}_{1,2} \in \mathbb{C}^{n_r \times (B-b_r)}$ ,  $\mathcal{F}_{2,1} \in \mathbb{C}^{(N_i-n_r) \times b_r}$ , and, finally,  $\mathcal{F}_{2,2}, \mathcal{F}_{2,3} \in \mathbb{C}^{(N_i-n_r) \times (B-b_r)}$ , and the decoder matrix  $\mathcal{G} \in \mathbb{C}^{B_r \times (2N_o)}$  is structured as follows

$$\mathcal{G} = \begin{bmatrix} \mathcal{G}_{1,1} & \mathcal{G}_{1,1}^* \\ \mathcal{G}_{2,1} & \mathcal{G}_{2,2} \\ \mathcal{G}_{2,2}^* & \mathcal{G}_{2,1} \end{bmatrix} \quad (3.8)$$

with  $\mathcal{G}_{1,1} \in \mathbb{C}^{b_r \times (2N_o)}$  and  $\mathcal{G}_{2,1}, \mathcal{G}_{2,2} \in \mathbb{C}^{(B-b_r) \times (2N_o)}$ . The decoder output, i.e., the estimate  $\hat{\mathbf{s}}$  of the transmitted vector  $\mathbf{s}$ , is equal to  $\hat{\mathbf{s}} = \bar{\mathcal{C}}_{b_r}[\mathcal{Z}]$ , with  $\bar{\mathcal{C}}_p[\cdot]$  defined in (2.23).

Since no special structures are assumed for both the precoder and decoder matrices, it is easily verified (see subsection 2.2.2) that the system models in (3.5) and (3.6) are equivalent. However, the adoption of the real-valued model makes the design of WLT simpler with respect to the adoption of the complex-valued one. In fact, any design method that adopts the real-valued representation avoids to take into account for the special block structure of the matrices  $\mathcal{F}$  and  $\mathcal{G}$  in (3.7) and (3.8), respectively, which define the WL filters according to the complex-valued representation. Only in the special case of real-valued information symbols and real-valued channel input, i.e., when  $b_r = B$  and  $n_r = N_i$ , the two models can be utilized without distinction, as it will be shown in the next section.

### 3.2 MMSE linear transceiver design

The overall system equation of a LT is given by (see also (1.85)):

$$\hat{\mathbf{s}}_L = \mathbf{G}_L \mathbf{H} \mathbf{F}_L \mathbf{s} + \mathbf{G}_L \mathbf{n} \quad (3.9)$$

where  $\mathbf{F}_L \in \mathbb{C}^{N_i \times B}$  and  $\mathbf{G}_L \in \mathbb{C}^{B \times N_o}$  denote the linear precoder and decoder, respectively,  $\hat{\mathbf{s}}_L$  is the estimate of  $\mathbf{s}$ , and where  $\mathbf{H}$  and  $\mathbf{n}$  have been defined in (3.1). In this section, we extend the design procedure proposed in [25] to the case where the components  $s^{(\ell)}$  of input vector can exhibit different variances. Such a condition is necessary since it allows us to simply derive the MMSE WLT by applying the proposed LT design procedure to the augmented system model (3.5). In fact, let us note that, although it is reasonable to assume, as in [25], that the outputs of the symbol mappers  $s_\ell$  exhibit the same variance, the components of the augmented vector  $\tilde{\mathbf{s}}$  in (3.5) exhibit the same variance only in the special cases of all real-valued or all complex-valued information symbols (i.e.,  $b_r = B$  or  $b_r = 0$ ).

To this aim, we assume zero-mean uncorrelated components for the symbol vector  $\mathbf{s}$ , i.e.,

$$\mathbf{R}_s \triangleq E[\mathbf{s}\mathbf{s}^H] = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_B^2) \quad (3.10)$$

where  $\sigma_\ell^2$  is the variance of the  $\ell$ th component of  $\mathbf{s}$ . The noise vector  $\mathbf{n}$  is assumed to be zero-mean with correlation matrix  $\mathbf{R}_n \triangleq E[\mathbf{n}\mathbf{n}^H]$ . Finally, we define the error vector  $\mathbf{e} \triangleq \hat{\mathbf{s}}_L - \mathbf{s}$  measured at the output of the decoder  $\mathbf{G}_L$  and its correlation matrix

$$\mathbf{R}_e \triangleq E[\mathbf{e}\mathbf{e}^H] = (\mathbf{G}_L \mathbf{H} \mathbf{F}_L - \mathbf{I}_B) \mathbf{R}_s (\mathbf{G}_L \mathbf{H} \mathbf{F}_L - \mathbf{I}_B)^H + \mathbf{G}_L \mathbf{R}_n \mathbf{G}_L^H \quad (3.11)$$

According to the MMSE criterion, the optimum LT precoder and decoder  $\mathbf{F}_L^{(opt)}$  and  $\mathbf{G}_L^{(opt)}$  are derived by solving the constrained minimization problem:

$$\begin{aligned} & \min_{\mathbf{F}_L, \mathbf{G}_L} \text{trace}(\mathbf{W} \mathbf{R}_e) \\ & \text{subject to: } \text{trace}(\mathbf{F}_L \mathbf{R}_s \mathbf{F}_L^H) = \mathcal{P}_0 \end{aligned} \quad (3.12)$$

where  $\mathbf{W}$  is a diagonal weight matrix of size  $B$ , and where  $\mathcal{P}_0$  denotes the total available transmission power. Note that without any power constraint, minimizing the above cost function leads to the trivial solution corresponding to  $\|\mathbf{G}\| = 0$  and requiring infinite power to be transmitted  $\|\mathbf{F}\| = \infty$ . The weighted MSE cost function in (3.12), as shown in [25], provides a unified framework for designing jointly optimal precoder and decoder according to different criterions<sup>1</sup>. In the particular case of  $\mathbf{W} = \mathbf{I}_B$ , (3.12) reduces to the conventional MMSE design.

The receiver filter matrix  $\mathbf{G}_L^{(opt)}$  that minimizes (3.12) is the WL Wiener filter for any given  $\mathbf{F}_L$  and it is equal to

$$\mathbf{G}_L^{(opt)} = \mathbf{R}_s \mathbf{F}_L^H \mathbf{H}^H (\mathbf{H} \mathbf{F}_L \mathbf{R}_s \mathbf{F}_L^H \mathbf{H}^H + \mathbf{R}_n)^{-1}, \quad (3.13)$$

which, as expected, does not depend on  $\mathbf{W}$ . Using the optimum decoder matrix settings of (3.13) in (3.12), and applying the matrix inversion lemma, we can rewrite the minimization problem (3.12) as follows:

$$\begin{aligned} & \min_{\mathbf{F}_L} \text{trace} \left( \mathbf{W} (\mathbf{R}_s^{-1} + \mathbf{F}_L^H \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} \mathbf{F}_L)^{-1} \right) \\ & \text{subject to: } \text{trace}(\mathbf{F}_L \mathbf{R}_s \mathbf{F}_L^H) = \mathcal{P}_0 \end{aligned} \quad (3.14)$$

<sup>1</sup>For example, by opportunely choosing the weight matrix, the transceiver can be optimized according to the maximum information rate design, as well as a QoS one.

Such an optimization problem can be solved by resorting to the eigenvalue decomposition (EVD)

$$\mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} = [\mathbf{U} \ \bar{\mathbf{U}}] \begin{bmatrix} \mathbf{\Gamma} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Gamma}} \end{bmatrix} \begin{bmatrix} \mathbf{U}^H \\ \bar{\mathbf{U}}^H \end{bmatrix} \quad (3.15)$$

where  $\mathbf{U}$  and  $\bar{\mathbf{U}}$  are orthogonal matrices of size  $N_i \times B$  and  $N_i \times (N_i - B)$ , respectively,  $\mathbf{\Gamma} \triangleq \text{diag}(\gamma_1, \dots, \gamma_B)$  is the diagonal matrix containing the  $B$  nonzero eigenvalues (arranged in decreasing order), and  $\bar{\mathbf{\Gamma}}$  is a diagonal matrix containing the zero eigenvalues. In defining (3.15) we have implicitly assumed that  $B = \text{rank}(\mathbf{H})$  as in [25]. The case  $B < \text{rank}(\mathbf{H})$ , as it will be shown, does not require to define a different mathematical framework, while the case  $B > \text{rank}(\mathbf{H})$  is not considered here<sup>2</sup>. By following the same procedure described in [25, 27, 66], it can be verified that the optimum matrix filter  $\mathbf{F}_L^{(opt)}$  is equal to:

$$\mathbf{F}_L^{(opt)} = \mathbf{U} \mathbf{\Theta} \quad (3.16)$$

where  $\mathbf{\Theta}$  is a diagonal matrix of size  $B$  with diagonal entries

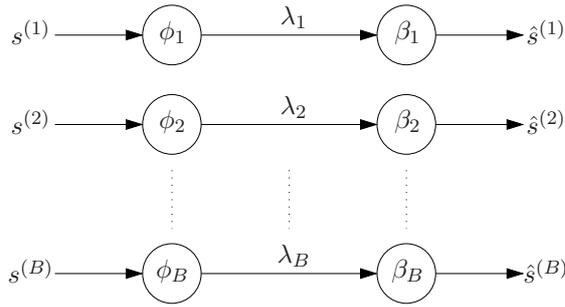
$$\Theta_\ell = \left( \frac{\mathcal{P}_0 + \sum_{i=1}^{\bar{B}} \gamma_i^{-1}}{\sum_{i=1}^{\bar{B}} w_\ell^{1/2} \gamma_i^{-1/2} \sigma_\ell} \cdot \frac{w_\ell^{1/2}}{\gamma_\ell^{1/2} \sigma_\ell} - \frac{1}{\gamma_\ell \sigma_\ell^2} \right)_+ \quad (3.17)$$

and where  $w_\ell$  are the entries of  $\mathbf{W}$ ,  $(a)_+ \triangleq \max(a, 0)$ , and  $\bar{B} \leq B$  is such that  $|\Theta_\ell|^2 > 0$  for  $\ell \in [1, \dots, \bar{B}]$  and  $|\Theta_\ell|^2 = 0$  for all others  $\ell$ . The variance of the  $\ell$ th component of the error vector  $\mathbf{e}$  corresponding to  $\mathbf{F}_L^{(opt)}$  and  $\mathbf{G}_L^{(opt)}$  is given by (3.11) and it can be shown to be equal to

$$\begin{aligned} \mathbf{R}_e(\ell, \ell) &= E[|e^{(\ell)}|^2] && \ell = 1, \dots, \bar{B} \\ &= w_\ell^{1/2} \gamma_\ell^{-1/2} \sigma_\ell \mu^{1/2} \end{aligned} \quad (3.18)$$

with  $\mu^{1/2} \triangleq \frac{\sum_{i=1}^{\bar{B}} w_\ell^{1/2} \gamma_i^{-1/2} \sigma_\ell}{\mathcal{P}_0 + \sum_{i=1}^{\bar{B}} \gamma_i^{-1}}$ . Note that the optimum precoder and decoder diagonalize both the channel  $\mathbf{H}$  and the error correlation matrix  $\mathbf{R}_e$ , giving rise to  $\bar{B}$  *eigen subchannels* (or equivalently *spatial models*) with channel gain  $\gamma_\ell$  (see [25], Lemma 2), as shown in Fig. 3.3.

<sup>2</sup>Since the transmitter knows the channel, it is reasonable that it will not transmit more than  $\text{rank}(\mathbf{H})$  independent data streams; however, very recently, it has been considered the possibility of transmitting  $B > \text{rank}(\mathbf{H})$  data streams by resorting to a DF-based decoder [36].



**Figure 3.3:** Decomposition of the MIMO channel into *eigen subchannels*: the gains  $\beta_\ell$  are optimized according to the MMSE criterion.

*Remark :* The MMSE LT has been derived by assuming uncorrelated components for the symbol vector  $\mathbf{s}$ . If  $\mathbf{R}_s$  is not diagonal, a prewhitening operation can be performed over the symbol vector  $\mathbf{s}$  before precoding, and the corresponding inverse operation can be performed after decoding.

### 3.3 MMSE Widely Linear transceiver design

In this section, we derive the MMSE WLT, i.e., the optimum matrices  $\mathbf{F}^{(opt)}$  and  $\mathbf{G}^{(opt)}$  in (3.5), by utilizing the method proposed in the previous section for the LT design.

Unlike LT optimization, which is based only on the knowledge of the input and noise correlation matrices  $\mathbf{R}_s$  and  $\mathbf{R}_n$ , the WLT optimization requires the knowledge of the augmented correlation matrices  $\mathbf{R}_{\tilde{\mathbf{s}}} \triangleq E[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T]$  and  $\mathbf{R}_{\tilde{\mathbf{n}}} \triangleq E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T]$ . In other words, it requires the knowledge of both the correlation matrix  $\mathbf{R}_s$  ( $\mathbf{R}_n$ ) and the pseudo-correlation matrix  $\mathbf{R}_{ss^*} \triangleq E[\mathbf{s}\mathbf{s}^T]$  ( $\mathbf{R}_{nn^*} \triangleq E[\mathbf{n}\mathbf{n}^T]$ ). We assume zero-mean uncorrelated components for the symbol vector  $\tilde{\mathbf{s}}$ , i.e.,

$$\mathbf{R}_{\tilde{\mathbf{s}}} = \text{diag}(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \dots, \tilde{\sigma}_{B_r}^2) \quad (3.19)$$

with  $\tilde{\sigma}_\ell^2$  denoting the variance of the  $\ell$ th component of  $\tilde{\mathbf{s}}$ . Note that, if the complex-valued components of  $\mathbf{s}$  are circularly symmetric, then  $\tilde{\sigma}_{b_r+\ell}^2 = \tilde{\sigma}_{B+\ell}^2$  with  $\ell = 1, \dots, (B - b_r)$ , i.e. the real and the imaginary parts exhibit the same variance.

According to (3.5), the error vector measured at the output of the WLT decoder  $\mathbf{G}$  is defined as  $\tilde{\mathbf{e}} \triangleq \tilde{\mathbf{z}} - \tilde{\mathbf{s}}$  and its correlation matrix is given by

$$\mathbf{R}_{\tilde{\mathbf{e}}} \triangleq E [\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T] = \left( \mathbf{G}\tilde{\mathbf{H}}\mathbf{F} - \mathbf{I}_{B_r} \right) \mathbf{R}_{\tilde{\mathbf{s}}} \left( \mathbf{G}\tilde{\mathbf{H}}\mathbf{F} - \mathbf{I}_{B_r} \right)^T + \mathbf{G}\mathbf{R}_{\tilde{\mathbf{n}}}\mathbf{G}^T. \quad (3.20)$$

The MMSE WLT precoder and decoder  $\mathbf{F}^{(opt)}$  and  $\mathbf{G}^{(opt)}$  are derived by solving the following constrained minimization problem:

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{G}} \text{trace} \left( \tilde{\mathbf{W}} \mathbf{R}_{\tilde{\mathbf{e}}} \right) \\ \text{subject to: } \text{trace} \left( \mathbf{F}\mathbf{R}_{\tilde{\mathbf{s}}}\mathbf{F}^T \right) = \mathcal{P}_0 \end{aligned} \quad (3.21)$$

where  $\tilde{\mathbf{W}}$  is a diagonal weight matrix of size  $B_r$ , and where  $\mathcal{P}_0$  is defined in (3.12). To this aim, define the EVD

$$\tilde{\mathbf{H}}^T \mathbf{R}_{\tilde{\mathbf{n}}}^{-1} \tilde{\mathbf{H}} = [\mathbf{V} \ \bar{\mathbf{V}}] \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Lambda}} \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \\ \bar{\mathbf{V}}^T \end{bmatrix} \quad (3.22)$$

where  $\mathbf{V}$  and  $\bar{\mathbf{V}}$  are orthogonal matrices of size  $D_r \times B_r$  and  $D_r \times (D_r - B_r)$ , respectively,  $\mathbf{\Lambda} \triangleq \text{diag}(\lambda_1, \dots, \lambda_{B_r})$  is the diagonal matrix containing the  $B_r$  nonzero eigenvalues (arranged in decreasing order), and  $\bar{\mathbf{\Lambda}}$  is a diagonal matrix containing the zero eigenvalues. In defining (3.22) we have implicitly assumed that  $B_r = \text{rank}(\tilde{\mathbf{H}})$  as in (3.15), and that  $n_r \geq B_r - N_i$ , so that all the  $B_r$  components of  $\tilde{\mathbf{s}}$  can be transmitted.

By utilizing the procedure proposed in the previous section for the MMSE LT design, it is straightforwardly verified that  $\mathbf{F}^{(opt)}$  and  $\mathbf{G}^{(opt)}$  are:

$$\mathbf{G}^{(opt)} = \mathbf{R}_{\tilde{\mathbf{s}}}\mathbf{F}^T \tilde{\mathbf{H}}^T \left( \tilde{\mathbf{H}}\mathbf{F}\mathbf{R}_{\tilde{\mathbf{s}}}\mathbf{F}^T \tilde{\mathbf{H}}^T + \mathbf{R}_{\tilde{\mathbf{n}}} \right)^{-1} \quad (3.23)$$

$$\mathbf{F}^{(opt)} = \mathbf{V}\mathbf{\Phi} \quad (3.24)$$

where  $\mathbf{\Phi}$  is a diagonal matrix of size  $B_r$  with diagonal entries

$$\Phi_\ell = \left( \frac{\mathcal{P}_0 + \sum_{i=1}^{\bar{B}_r} \lambda_i^{-1}}{\sum_{i=1}^{\bar{B}_r} \tilde{w}_i^{1/2} \lambda_i^{-1/2} \tilde{\sigma}_i} \cdot \frac{\tilde{w}_\ell^{1/2}}{\lambda_\ell^{1/2} \tilde{\sigma}_\ell} - \frac{1}{\lambda_\ell \tilde{\sigma}_\ell^2} \right)_+^{1/2} \quad (3.25)$$

and where  $\tilde{w}_\ell$  are the entries of  $\tilde{\mathbf{W}}$ , and where  $\bar{B}_r \leq B_r$  is such that  $|\Phi_\ell|^2 > 0$  for  $\ell \in [1, \dots, \bar{B}_r]$  and  $|\Phi_\ell|^2 = 0$  for all others  $\ell$ .

Analogously, the variance of the  $\ell$ th component of the error vector  $\tilde{\mathbf{e}}$ , corresponding to  $\mathbf{F}^{(opt)}$  and  $\mathbf{G}^{(opt)}$ , is given by:

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{e}}}^{(opt)}(\ell, \ell) &= E[\tilde{e}_\ell] & \ell = 1, \dots, \bar{B}_r \\ &= \tilde{w}^{1/2} \lambda_\ell^{-1/2} \tilde{\sigma}_\ell \tilde{\mu}^{1/2} \end{aligned} \quad (3.26)$$

$$\text{with } \tilde{\mu}^{1/2} \triangleq \frac{\sum_{i=1}^{\bar{B}_r} \tilde{w}_i^{1/2} \lambda_i^{-1/2} \tilde{\sigma}_i}{\mathcal{P}_0 + \sum_{i=1}^{\bar{B}_r} \lambda_i^{-1}}.$$

Let us note that the cost function in (3.14) depends on the parameter  $n_r$ . More specifically, the sizes of both the channel matrix  $\tilde{\mathbf{H}}$  and of the precoder matrix  $\mathbf{F}$  depend on such a parameter and, hence, since there is no functional dependence of the cost function on  $n_r$ , the optimization (over  $n_r$ ) has to be carried out by an exhaustive procedure. ■

*Remark 1 :* We point out that in presence of possibly correlated complex-valued circularly variant symbol vector  $\mathbf{s}$ , the prewhitening operation, which provides the uncorrelated input vector to the precoder, has to be performed by utilizing a WL filter operating on the augmented vector  $\tilde{\mathbf{s}}$  instead of a linear one operating on  $\mathbf{s}$ . In fact, the adoption of a linear filter does not ensure that the real and the imaginary parts of the complex-valued components of the output of the prewhitening filter will be uncorrelated as assumed in (3.19). ■

*Remark 2 :* Let us consider the complex-valued system model (3.6) in the special case where  $b_r = B$  and  $n_r = N_i$ . From (3.7), one has that the precoder matrix  $\mathcal{F}$  degenerates into the nonstructured real-valued matrix  $\mathcal{F}_{1,1}$  of size  $N_i \times B$ . Moreover, define the matrices  $\mathcal{H} \triangleq \tilde{\mathcal{C}}_{N_i}[\mathbf{H}]$  and  $\mathcal{R}_{\tilde{\mathbf{n}}} \triangleq E[\mathcal{C}_0[\mathbf{n}]\mathcal{C}_0[\mathbf{n}]^H]$ . The optimum precoder and decoder  $\mathcal{F}^{(opt)}$  and  $\mathcal{G}^{(opt)}$  can be derived by utilizing the same procedure proposed in this section with reference to the real-valued representation. However, since it can be easily verified that

$$\begin{aligned} \mathcal{H}^H \mathcal{R}_{\tilde{\mathbf{n}}}^{-1} \mathcal{H} &= \tilde{\mathbf{H}}^H \mathbf{R}_{\tilde{\mathbf{n}}}^{-1} \tilde{\mathbf{H}} \\ &= \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T, \end{aligned} \quad (3.27)$$

the complex-valued representation leads an optimum WLT which is equivalent to the one obtained according to the real-valued notation. Specifically, one has

that

$$\mathcal{F}^{(opt)} = \mathbf{V}\Phi \quad (3.28)$$

and the optimum  $\mathcal{G}^{(opt)}$  is obtained from  $\mathbf{G}$  by means of (2.8).

### 3.4 MSE Analysis

In this section, we compare the performances of the MMSE WLT and MMSE LT in the special scenarios of real-valued information symbols  $s^{(i)}$  (i.e.,  $b_r = B$ ), and in the special one of complex-valued circularly symmetric information symbols  $s^{(i)}$  (i.e.,  $b_r = 0$ ). Since the MSE achieved by the MMSE LT and by MMSE WLT are functions of the eigenvalues  $\gamma_\ell$  in (3.15) and  $\lambda_\ell$  in (3.22), respectively, the performance comparison will be carried out by studying the relationship between  $\gamma_\ell$  and  $\lambda_\ell$ . We will show that the MMSE WLT outperforms the MMSE LT, whereas for the case where  $B$  complex-valued circularly symmetric information symbols have to be transmitted (i.e.,  $b_r = 0$ ), the MMSE WLT degenerates into the MMSE LT. Moreover, our analysis points out the incapability by the MMSE LT of transmitting the  $B$  information symbols when  $N_i$  and  $N_o$  are sufficiently large and  $\frac{N_i}{N_o} \approx 1$ .

The following subsections 3.4.1 and 3.4.2 present, with reference to the case of real-valued information symbols ( $b_r = B$ ), two structures for the design of the MMSE-WLT: the former assumes that the transmitted symbols  $x^{(\ell)}$  are real-valued (i.e.,  $n_r = N_i$ ), whereas the latter assumes that they are complex-valued (i.e.,  $n_r = 0$ ). Finally, in subsection 3.4.3, we consider the case where  $B$  complex-valued circularly symmetric information symbols are transmitted by utilizing  $N_i$  complex-valued channel inputs. Although the case  $0 < b_r < B$  is not studied in the sequel, we remark that, due to the non circularly variant nature of the desired vector, the MMSE WLT is expected to outperform the MMSE LT.

With no loss of generality, in the following, we assume unit-variance information symbols, i.e., we assume  $\tilde{\sigma}_i^2 = 1$  for  $i = 1, \dots, b_r$ , and  $\tilde{\sigma}_i^2 = 1/2$  when  $i = b_r + 1, \dots, B_r$ . Moreover, to clearly show the effects of WL processing on the *eigen subchannel* decomposition, we assume uncorrelated complex-valued circularly symmetric noise components with variance  $\sigma_n^2$ , i.e.,  $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_{N_o}$  and  $\mathbf{R}_{nn^*} = \mathbf{0}$ , or, equivalently,  $\mathbf{R}_{\tilde{n}} = \frac{\sigma_n^2}{2} \mathbf{I}_{2N_o}$ .

### 3.4.1 Real-valued information symbols and real-valued transmitted symbols ( $br = B, n_r = N_i$ )

The MSE analysis provided in this subsection is first carried out by considering two scenarios for the matrix  $\mathbf{H}$ :

1.  $\mathbf{H}$  is deterministic;
2.  $\mathbf{H}$  is a random matrix whose entries are i.i.d. random variables, and such that  $N_i, N_o \rightarrow \infty$  with a finite ratio  $\frac{N_i}{N_o}$ .

#### Deterministic channel

Accounting for the assumption of uncorrelated complex-valued circularly symmetric noise components, it can be verified that

$$2 \cdot \tilde{\mathbf{H}}^T \tilde{\mathbf{H}} = \mathbf{H}^H \mathbf{H} + \mathbf{H}^T \mathbf{H}^* \quad . \quad (3.29)$$

Before proceeding any further, let us remind here the Weyl's theorem [55, p. 396], on the perturbation matrix theory:

**Theorem 3.1** *If  $\mathbf{A}$  and  $\mathbf{A} + \mathbf{E}$  are  $n \times n$  Hermitian matrices, then*

$$\lambda_k(\mathbf{A}) + \lambda_n(\mathbf{E}) \leq \lambda_k(\mathbf{A} + \mathbf{E}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{E}) \quad k = 1, \dots, n$$

where  $\lambda_k(\mathbf{A})$  denote the  $k$ th largest eigenvalue of  $\mathbf{A}$ .

By applying the Weyl's theorem to (3.29), the following bounds hold:

$$\lambda_\ell(\mathbf{H}^H \mathbf{H}) + \lambda_{N_i}(\mathbf{H}^H \mathbf{H}) \leq \lambda_\ell(2 \cdot \tilde{\mathbf{H}}^H \tilde{\mathbf{H}}) \leq \lambda_\ell(\mathbf{H}^H \mathbf{H}) + \lambda_1(\mathbf{H}^H \mathbf{H}) \quad (3.30)$$

where we have taken into account for  $\lambda_\ell(\mathbf{H}^H \mathbf{H}) = \lambda_\ell(\mathbf{H}^T \mathbf{H}^*)$ . From (3.30) one has that  $\gamma_\ell + \gamma_{N_i} \leq \lambda_\ell \leq \gamma_\ell + \gamma_1$ , i.e., the MMSE WLT decomposes the MIMO channel into  $B$  eigen subchannels with gains larger than those of the MMSE LT eigen subchannels.

Let us now consider the achieved MMSEs in (3.18) and (3.26). We assume that the same number of information symbols  $\bar{B} = \bar{B}_r = B$  are transmitted by the MMSE LT and the MMSE WLT. However, the same results apply to the case where  $\bar{B} = \bar{B}_r < B$ .

The  $(\ell, \ell)$  entry of the error correlation matrices  $\mathbf{R}_e^{(opt)}$  and  $\mathbf{R}_\varepsilon^{(opt)}$  can be both expressed as follows:

$$\mathbf{R}_e^{(opt)}(\ell, \ell) = f_\ell(\gamma_1, \dots, \gamma_B) \quad (3.31)$$

$$\mathbf{R}_\varepsilon^{(opt)}(\ell, \ell) = f_\ell(\lambda_1, \dots, \lambda_B) \quad (3.32)$$

where

$$f_\ell(t_1, \dots, t_B) \triangleq \frac{t_\ell^{-1} + t_\ell^{-1/2} \sum_{n \neq \ell} t_n^{-1/2}}{\mathcal{P}_0 + t_\ell^{-1} + \sum_{n \neq \ell} t_n^{-1}} \quad (3.33)$$

with  $t_\ell > 0$  ( $\ell = 1, \dots, B$ ) such that

$$\mathcal{P}_0 + \sum_{n \neq \ell} t_n^{-1} \geq t_\ell^{-1/2} \sum_{n \neq \ell} t_n^{-1/2} \quad \forall \ell$$

to ensure that the error variance does not exceed the input variance. It can be shown that  $f_\ell(t_1, \dots, t_B)$  monotonically decreases with  $t_\ell$ , and hence, being  $\gamma_\ell \leq \lambda_\ell$ , one has

$$\mathbf{R}_{\tilde{\mathbf{e}}}^{(opt)}(\ell, \ell) \leq \mathbf{R}_{\mathbf{e}}^{(opt)}(\ell, \ell) \quad , \quad (3.34)$$

i.e., the MMSE WLT outperforms the MMSE LT for any of the  $B$  desired symbols.

### Random channel with large $N_i$ and $N_o$

We consider a channel matrix  $\mathbf{H}$  whose entries  $h^{(\ell, i)}$  are modeled as complex-valued circularly symmetric independent zero-mean i.i.d. random variables with variance  $\frac{1}{N_o}$ . According to above assumptions, the entries of the augmented channel matrix  $\tilde{\mathbf{H}}$  are zero-mean i.i.d. random variables with the same variance  $\frac{1}{2N_o}$ . Moreover, we assume that the number of channel inputs and outputs grows up to infinity ( $N_i, N_o \rightarrow \infty$ ).

By utilizing the Marčenko-Pastur law (Theorem 1.2), we can determine the asymptotic (i.e.,  $N_i, N_o \rightarrow \infty$ ) probability density function (pdf) of the eigenvalues  $\gamma_\ell$  and  $\lambda_\ell$  versus  $\beta \triangleq \frac{N_i}{N_o}$ :

$$p_{\gamma_\ell}(\alpha) = \sigma_n^2 \left[ \left(1 - \frac{1}{\beta}\right)_+ \delta(\alpha) + \frac{\sqrt{(\alpha - a_L)_+ (b_L - \alpha)_+}}{2\pi\beta\alpha} \right] \quad (3.35)$$

$$p_{\lambda_\ell}(\alpha) = \frac{\sigma_n^2}{2} \left[ \left(1 - \frac{2}{\beta}\right)_+ \delta(\alpha) + \frac{\sqrt{(\alpha - a_{WL})_+ (b_{WL} - \alpha)_+}}{\pi\beta\alpha} \right] \quad (3.36)$$

$$a_L \triangleq \frac{1}{\sigma_n^2} (1 - \sqrt{\beta})^2 \quad b_L \triangleq \frac{1}{\sigma_n^2} (1 + \sqrt{\beta})^2 \quad (3.37)$$

$$a_{WL} \triangleq \frac{2}{\sigma_n^2} \left(1 - \sqrt{\frac{\beta}{2}}\right)^2 \quad b_{WL} \triangleq \frac{2}{\sigma_n^2} \left(1 + \sqrt{\frac{\beta}{2}}\right)^2 \quad , \quad (3.38)$$

where  $\alpha \in [a_L, b_L]$  in (3.35) and  $\alpha \in [a_{WL}, b_{WL}]$  in (3.36). According to (3.35) and (3.36), the mean values of  $\gamma_\ell$  and  $\lambda_\ell$  are

$$E[\gamma_\ell] = \min \left\{ 1, \frac{1}{\beta} \right\} \cdot \frac{1}{\sigma_n^2} \quad (3.39)$$

$$E[\lambda_\ell] = \min \left\{ 1, \frac{2}{\beta} \right\} \cdot \frac{2}{\sigma_n^2} \quad , \quad (3.40)$$

respectively, i.e., the mean value of  $\lambda_\ell$  is at least twice the mean value of  $\gamma_\ell$ . Moreover, let us note that, in the spacial case of Gaussian channel entries ( $h^{(\ell,i)} \sim \mathcal{N}(0, 1/N_o)$ ), the maximum (minimum nonnull) eigenvalues of the matrices  $\frac{1}{\sigma_n^2} \mathbf{H}^H \mathbf{H}$  and  $\frac{2}{\sigma_n^2} \tilde{\mathbf{H}}^H \tilde{\mathbf{H}}$  converge almost surely to  $b_L$  and  $b_{WL}$ , respectively, ( $a_L$  and  $a_{WL}$  respectively) [67]. Since both the performances of the MMSE WLT and of the MMSE LT are governed by the weakest subchannels, the lower limits  $a_L$  and  $a_{WL}$  allow to state a performance comparison between the two structures in the considered asymptotic scenario.

When  $\beta \lll 1$ , from (3.35) and (3.38), the random variables  $\gamma_\ell$  and  $\lambda_\ell$  can be reasonably approximated to their mean values, more specifically,  $\gamma_\ell \cong \frac{1}{\sigma_n^2}$  and  $\lambda_\ell \cong \frac{2}{\sigma_n^2}$ . By substituting such values in (3.33), and assuming  $\frac{P_0}{\sigma_n^2} = \xi B$  ( $\xi \in \mathbb{R}^+$ ), one has

$$\lim_{\beta \rightarrow 0} \frac{\mathbf{R}_{\tilde{e}}^{(opt)}(\ell, \ell)}{\mathbf{R}_e^{(opt)}(\ell, \ell)} = \frac{1}{2} \cdot \frac{\xi + 1}{\xi + \frac{1}{2}} \quad . \quad (3.41)$$

From (3.41), it follows that the performance gain provided by the MMSE WLT over the MMSE LT over each subchannel approaches to 3dB when  $\xi \ggg 1$ , i.e., when the ratio of the transmitted power to the noise variance at each channel output is larger than  $B$ .

When  $\beta \rightarrow 1$ , the probability of having null  $\gamma_\ell$  increases since  $a_L$  achieves the minimum value for  $\beta = 1$ ; moreover,  $\gamma_B \rightarrow 0$  almost surely when  $h^{(\ell,i)} \sim \mathcal{N}(0, 1/N_o)$ . Let us note that, as  $\gamma_\ell \rightarrow 0$ , one has

$$\lim_{\gamma_\ell \rightarrow 0} \mathbf{R}_e^{(opt)}(\ell, \ell) = 1 \quad (3.42)$$

$$\lim_{\gamma_\ell \rightarrow 0} \mu^{1/2} = 0 \quad . \quad (3.43)$$

From (3.42), it follows that the MMSE LT performs poorly over the weakest *eigen subchannels*. Moreover, from (3.43), it can happen that the  $B$  symbols  $s^{(\ell)}$  are partially transmitted. On the other hand, since  $a_{WL} = (\sqrt{2} - 1)^2 / \sigma_n^2$  for  $\beta = 1$ , the MMSE WLT, as  $\sigma_n^2 \rightarrow 0$ , improves the estimate of all the  $B$

symbols. Moreover, by improving the *eigen subchannel* gains, the WL processing allows us to achieve a multiplexing gain, i.e., it reduces the probability of discarding the weakest subchannels.

Finally, note that the impulsive terms in (3.35) and (3.36) have nonnull amplitudes for  $\beta > 1$  and for  $\beta > 2$ , respectively, and they account for the  $N_i - N_o$  zero eigenvalues in (3.15) and the  $N_i - 2N_o$  zero eigenvalues in (3.22). In such a scenario, the comparison between the smallest values  $a_L$  and  $a_{WL}$  of the range of the eigenvalues  $\gamma_\ell$  and  $\lambda_\ell$ , respectively, does not allow us to carry out the performance comparison between the two transceivers. In fact, since  $N_i > N_o$ , one has that  $a_L$  still represents a quality index of the transmission over the weakest subchannels, whereas  $a_{WL}$  does not. For the sake of clarity, let us consider the numerical example  $\beta = 2$  ( $N_i = 2N_o$ ). The impulsive term  $\frac{1}{2}\delta(\alpha)$  in (3.35) accounts for the  $N_o$  zero eigenvalues, while  $a_L$  and  $b_L$  define the interval where the nonnull eigenvalues lie. On the other hand, the impulsive term in (3.36) has null amplitude, and  $a_{WL} = 0$ . Being  $B = N_o$ , the eigenvalues  $(\lambda_{N_o+1}, \dots, \lambda_{2N_o})$  are the channel gains of the unutilized *eigen subchannels* and, therefore,  $a_{WL}$  does not provide any information about the transmission over the weakest *eigen subchannel*.

■

*Remark :* It is important to note that the assumption of real-valued transmitted symbols  $x^{(\ell)}$  limits the search for the optimum MMSE WLT precoder  $\mathbf{F}$  in the subset of the real-valued matrices, whereas the optimum  $\mathbf{F}_L$  is complex-valued. Nevertheless we have shown in this section that such a restriction does not determine any loss when the MMSE is adopted as performance measure. On the other hand, if the MMSE WLT and the MMSE LT are compared in terms of symbol error rate, the above restriction can lead to a performance loss when special scenarios are considered (see Section 3.5).

### 3.4.2 Real-valued information symbols and complex-valued transmitted symbols ( $b_r = B, n_r = 0$ )

In this scenario, we have to analyze separately the case  $B = 2K$  and the case  $B = 2K - 1$  (with  $K$  positive integer). We will show that, regardless the value of  $B$  (even or odd): a) the MMSE WLT outperforms the MMSE LT b) the MMSE WLT precoder can be implemented with a linear filter (rather than WL) that processes a complex-valued input vector whose components belong to a different constellation set.

$B = 2K$  : By assuming (for the sake of simplicity)  $\bar{B} = B$ , i.e.,  $\mu^{1/2} \leq \sqrt{\gamma_B}$  and accounting for (3.51), one has that

$$\begin{aligned} \tilde{\mu}^{1/2} &= \frac{\sum_{i=1}^{B_r} \lambda_i^{-1/2}}{\mathcal{P}_0 + \sum_{i=1}^{B_r} \lambda_i^{-1}} \\ &= \sqrt{2} \frac{\sum_{i=1}^K \gamma_i^{-1/2}}{\mathcal{P}_0 + \sum_{i=1}^K \gamma_i^{-1}} . \end{aligned} \quad (3.44)$$

By observing that  $\mu^{1/2} \geq \frac{\sum_{i=1}^K \gamma_i^{-1/2}}{\mathcal{P}_0 + \sum_{i=1}^K \gamma_i^{-1}}$ , it can be shown that  $\tilde{\mu}^{1/2} \leq \sqrt{\lambda_B} = \sqrt{2\gamma_K}$ , i.e., the MMSE WLT utilizes all the  $\bar{B}_r = B$  subchannels to transmit the information vector. Straightforwardly, by utilizing (3.44), it can be verified that

$$\begin{aligned} \mathbf{R}_{\tilde{e}}(2\ell - 1, 2\ell - 1) &= \mathbf{R}_{\tilde{e}}(2\ell, 2\ell) \\ &= \gamma_\ell^{-1/2} \frac{\sum_{i=1}^K \gamma_i^{-1/2}}{\mathcal{P}_0 + \sum_{i=1}^K \gamma_i^{-1}} \quad \ell = 1, \dots, K \end{aligned} \quad (3.45)$$

and, consequently,  $\mathbf{R}_{\tilde{e}}(n, n) \leq \mathbf{R}_e(n, n)$  ( $n = 1, \dots, B$ ), namely the MMSE WLT outperforms the MMSE LT.

Let us show now that the precoder of the MMSE WLT is equivalent to the precoder of the MMSE LT transmitting  $B/2$  complex-valued information symbols  $s^{(\ell)}$  of a QAM constellation. In such a case, the matrix  $\Theta$  in the expression (3.18) of the optimum precoder  $\mathbf{F}_L^{(opt)} = \mathbf{U}\Theta$  is such that

$$|\Theta_\ell|^2 = \begin{cases} \left( \frac{\gamma_\ell^{-1/2}}{\mu^{1/2}} - \frac{1}{\gamma_\ell} \right)_+ & \ell = 1, \dots, B/2 \\ 0 & \ell = B/2 + 1, \dots, B \end{cases} \quad (3.46)$$

with  $\mu^{1/2} = \frac{\tilde{\mu}^{1/2}}{\sqrt{2}}$ . Consequently, from (3.44), it follows that the optimum matrix entries  $\Phi_\ell$  in the expression (3.25) of the MMSE WLT precoder are given by:

$$\Phi_{2\ell-1} = \Phi_{2\ell} = \frac{1}{\sqrt{2}} \cdot \Theta_\ell \quad \ell = 1, \dots, B/2 . \quad (3.47)$$

Accounting for (3.47) and (3.51), it can be verified that

$$\mathbf{F}^{(opt)} = \frac{1}{\sqrt{2}} \cdot \tilde{\mathcal{E}}_0 \left[ \mathbf{F}_L^{(opt)} \right] , \quad (3.48)$$

and the output  $\bar{\mathcal{E}}_0[\tilde{\mathbf{x}}]$  of the optimum WL precoder is equal to

$$\begin{aligned}\bar{\mathcal{E}}_0[\tilde{\mathbf{x}}] &= \frac{1}{\sqrt{2}} \cdot \mathbf{F}_L^{(opt)} \bar{\mathcal{E}}_0[\mathbf{s}] \\ &\triangleq \mathbf{F}_L^{(opt)} \mathbf{s}_{QAM} \quad ,\end{aligned}\tag{3.49}$$

where the  $\ell$ th component of the vector  $\mathbf{s}_{QAM}$  is equal to  $s_{QAM}^{(\ell)} \triangleq \frac{1}{\sqrt{2}}(s_{QAM}^{(\ell)} + js_{QAM}^{(\ell+B/2)})$  ( $\ell = 1, \dots, B/2$ ), i.e., its in-phase and quadrature components are modulated by the two independent (of each other) symbols  $s_\ell$  and  $s_{\ell+B/2}$ , and normalized so that  $s_{QAM}^{(\ell)}$  has unit variance<sup>3</sup>. From (3.49), it follows that the output of MMSE WLT precoder can be obtained by linearly precoding the vector  $\mathbf{s}_{QAM}$  with the optimum MMSE LT precoder designed to transmit  $B/2$  information symbols.

Note that, since  $\mathbf{s}_{QAM}$  is circularly symmetric, the output of the optimum WL precoder, and so the channel output, are jointly circularly symmetric<sup>4</sup>. However, since the desired vector is actually not circularly symmetric, the optimum decoder  $\mathbf{G}^{(opt)}$  is still a WL filter. Let us also note that, the output vector  $\hat{\mathbf{s}}$  (scaled by  $\sqrt{2}$ ) of  $\mathbf{G}^{(opt)}$  can be equivalently obtained by utilizing the optimum linear decoder  $\mathbf{G}_L^{(opt)}$  corresponding to  $\mathbf{F}_L^{(opt)}$  in (3.48), and applying the non linear operator  $\mathcal{E}_0[\cdot]$  to its output vector  $\hat{\mathbf{s}}_{QAM}$ .

Finally, we compare the performances of the MMSE WLT proposed in this subsection (say  $\mathbf{R}_\varepsilon^{[B]}(n, n)$  the achieved MSE) with the performances of the MMSE WLT proposed in subsection 3.4.1 (say  $\mathbf{R}_\varepsilon^{[A]}(n, n)$  the achieved MSE) with reference to the special case where  $N_i, N_o \rightarrow \infty$  and  $\beta \rightarrow 0$ . In such a scenario, it has been shown that  $\gamma_\ell$  and  $\lambda_\ell$  can be approximated with the asymptotic values  $\frac{1}{\sigma_n^2}$  and  $\frac{2}{\sigma_n^2}$ , respectively. Accounting for (3.45), it is easy verified that as  $\sigma_n^2 \rightarrow 0$ ,

$$\begin{aligned}\mathbf{R}_\varepsilon^{[B]}(n, n) &\rightarrow \frac{\sigma_n^2 B}{2\mathcal{P}_0} \\ \mathbf{R}_\varepsilon^{[A]}(n, n) &\rightarrow \frac{\sigma_n^2 B}{2\mathcal{P}_0} \quad ,\end{aligned}\tag{3.50}$$

i.e., the two MMSE WLT structures performs equivalently. The limits in (3.50) imply also that the MMSE WLT which transmits  $B$  symbols belonging

<sup>3</sup>Note that, if  $\mathbf{s}^{(\ell)}$  belongs to the M-ASK constellation, then  $s_{QAM}^{(\ell)}$  belongs to the M<sup>2</sup>-QAM constellation.

<sup>4</sup>If the input vector is circularly symmetric and the channel is linear, then it can be shown that the input and output of the channel are jointly circularly symmetric.

to a real-valued alphabet (such as M-ASK) achieves the same performances of an MMSE LT which transmits  $B/2$  symbols  $s_{QAM}^{(\ell)}$  belonging to a denser alphabet. Unfortunately, the proposed framework does not allow us to provide a complete performance comparison between the two structures.

*B odd* : Following the same guidelines reported in the previous paragraph, it can be verified that when  $B$  is odd

- if  $\bar{B} = B$ , then  $\bar{B}_r = B$ , i.e., the MMSE WLT and the MMSE LT utilize the same number of subchannels to transmit the information symbols;
- $\mathbf{R}_{\bar{e}}(n, n) \leq \mathbf{R}_e(n, n)$  ( $n = 1, \dots, B$ ), i.e., the MMSE WLT outperforms the MMSE LT;
- the MMSE WLT precoder can be implemented with a linear filter that processes a new input vector of size  $\frac{B-1}{2}$  whose first  $K-1$  components are equal to  $s_\ell + js_{\ell+K-1}$  ( $\ell = 1, \dots, K-1$ ), and the  $K$ th one is the real-valued component  $s_{2K-1}$ . However, in such a case, the MMSE WLT can not be equivalently performed by an MMSE LT;
- as  $N_i, N_o \rightarrow \infty$  and  $\beta \rightarrow 0$ , the MMSE WLT proposed in this subsection and the MMSE WLT in subsection 3.4.1 perform equivalently.

### 3.4.3 Complex-valued information symbols and complex-valued transmitted symbols ( $b_r = 0, n_r = 0$ )

In this subsection, we show that the MMSE WLT degenerates into the MMSE LT when the information symbols  $s^{(\ell)}$  are complex-valued circularly symmetric ( $b_r = 0$ ) and the channel inputs  $x^{(\ell)}$  are assumed to be complex-valued ( $n_r = 0$ ). For the sake of simplicity, let us assume, with no loss of generality,  $B = N_i = \text{rank}(\mathbf{H})$  in (3.15).

By computing the EVD of the matrix  $\mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}$  as in (3.15), after some matrix manipulations, it can be verified that

$$\frac{2}{\sigma_n^2} \tilde{\mathbf{H}}^T \tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{U}_{\Re} & -\mathbf{U}_{\Im} \\ \mathbf{U}_{\Im} & \mathbf{U}_{\Re} \end{bmatrix} \begin{bmatrix} 2\Gamma & \mathbf{0} \\ \mathbf{0} & 2\Gamma \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\Re} & -\mathbf{U}_{\Im} \\ \mathbf{U}_{\Im} & \mathbf{U}_{\Re} \end{bmatrix}^T \quad (3.51)$$

where  $\mathbf{U}_{\Re} \triangleq \Re\{\mathbf{U}\}$ , and  $\mathbf{U}_{\Im} \triangleq \Im\{\mathbf{U}\}$ . Since  $\lambda_{2\ell-1} = \lambda_{2\ell} = 2\gamma_\ell$  for  $\ell = 1, \dots, B$ , and accounting for the assumption of circularly symmetric information symbols ( $\tilde{\sigma}_{2\ell}^2 = \tilde{\sigma}_{2\ell-1}^2 = \sigma_\ell^2/2$ , for  $\ell = 1, \dots, B$ ), one has that  $\bar{B}_r = 2\bar{B}$ . In fact, from (3.15) it follows that the number  $\bar{B}$  of subchannels

utilized by the MMSE LT is such that  $\mu^{1/2} \leq \rho_{\bar{B}}$  with  $\rho_\ell \triangleq w_\ell^{1/2} \gamma_\ell^{1/2} \sigma_\ell$  and ordered in decreasing manner:  $\rho_1 \geq \dots \geq \rho_B$ . Analogously, from (3.22),  $\bar{B}_r$  is such that  $\tilde{\mu}^{1/2} \leq \tilde{\rho}_{\bar{B}_r}$  with  $\tilde{\rho}_\ell \triangleq \tilde{w}_\ell^{1/2} \tilde{\lambda}_\ell^{1/2} \tilde{\sigma}_\ell$  (and ordered in decreasing manner). Being  $\tilde{\rho}_{2\ell} = \tilde{\rho}_{2\ell-1} = \rho_\ell$  ( $\ell = 1, \dots, B$ ), one has that

$$\mu^{1/2} \leq \rho_{\bar{B}} \Rightarrow \tilde{\mu}^{1/2} \leq \tilde{\rho}_{2\bar{B}} \quad , \quad (3.52)$$

i.e.,  $\bar{B}_r = 2\bar{B}$ . Consequently, from (3.18) and (3.25), it is easily verified that  $\Theta_\ell = \Phi_{2\ell-1} = \Phi_{2\ell}$ , with  $\ell = 1, \dots, \bar{B}$  and

$$\mathbf{F}^{(opt)} = \tilde{\mathcal{E}}_0 \left[ \mathbf{F}_L^{(opt)} \right] \quad . \quad (3.53)$$

From such a relation it holds that

$$\begin{aligned} \mathbf{x} &= \bar{\mathcal{E}}_0 \left[ \mathbf{F}^{(opt)} \tilde{\mathbf{s}} \right] \\ &= \mathbf{F}_L^{(opt)} \mathbf{s} \quad , \end{aligned} \quad (3.54)$$

i.e., the optimum WL precoder degenerates into the optimum linear precoder. Moreover, being the desired vector and the received one jointly circularly symmetric, the optimum WL decoder degenerates into the optimum linear one [44] and, consequently, the MMSE WLT degenerates into the MMSE LT.

### 3.5 SER analysis

In this section, we provide a performance comparison in terms of the symbol error rate (SER) between the MMSE LT and the MMSE WLT. The MSE analysis carried out in the previous section showed that the WLT is optimum in the MMSE sense. However, let us note that when the SER is adopted as performance measure, only the real part of the estimates of the real-valued symbols  $s_\ell$  ( $\ell = 1, \dots, b_r$ ) is relevant for the decision rule. For such a reason, the MSE improvement provided by the MMSE WLT is not fully translated into a SER improvement.

To gain some insight into the problem, let us consider, as in subsection 3.4.1, the scenario where  $b_r = B$  unit variance real-valued symbols  $s^{(\ell)}$  are transmitted and where the channel inputs  $x^{(\ell)}$  are real-valued. Let us note that, although the desired vector is real-valued, the output of the MMSE LT is complex-valued and, hence, the variance of imaginary part of the components of  $\mathbf{e}_L$  (the error vector measured at the output of the conventional MMSE LT) does not affect the SER. It follows that the performance comparison in terms

of SER between the MMSE LT and the MMSE WLT has to be carried out by comparing the variances of the components of  $\mathbf{e}_{L,R} \triangleq \Re\{\mathbf{e}_L\}$  with those of the components of  $\tilde{\mathbf{e}}$  (the error vector measured at the output of the MMSE WLT).

To this aim, it can be verified (use *Lemma 2* and *Theorem 1* in [25]) that:

$$\mathbf{R}_{e,R} \triangleq E[\mathbf{e}_{L,R}\mathbf{e}_{L,R}^H] = \frac{1}{2}\mu^{1/2}\mathbf{\Gamma}^{-1/2}(\mathbf{I}_B + \mu^{1/2}\mathbf{\Gamma}^{-1/2}) \quad (3.55)$$

$$\mathbf{R}_{e,I} \triangleq E[\mathbf{e}_{L,I}\mathbf{e}_{L,I}^H] = \frac{1}{2}\mu^{1/2}\mathbf{\Gamma}^{-1/2}(\mathbf{I}_B - \mu^{1/2}\mathbf{\Gamma}^{-1/2}) \quad (3.56)$$

where  $\mathbf{e}_{L,I} \triangleq \Im\{\mathbf{e}_L\}$ . Unfortunately, the comparison between  $R_{\tilde{\mathbf{e}}}(\ell, \ell)$  and  $\mathbf{R}_{e,R}(\ell, \ell)$  is not simple to be carried in the general case. For such a reason, we consider the two extreme cases:

- the MMSE LT performs poorly, i.e., when  $\mathbf{R}_e(\ell, \ell) = \mu^{1/2}\gamma_\ell^{-1/2} \approx 1$ ;
- the MMSE LT performs well, i.e., when  $\mathbf{R}_e(\ell, \ell) = \mu^{1/2}\gamma_\ell^{-1/2} \lll 1$ .

According to the former, the variance  $R_e(\ell, \ell)$  of  $e^{(\ell)}$  is governed by the variance  $\mathbf{R}_{e,R}(\ell, \ell)$  of its real part. In such a case, it is reasonable to expect that the MSE improvements provided by the MMSE WLT over the MMSE LT are fully translated into SER improvements.

On the other hand, according to the latter case, one has

$$\begin{aligned} \mathbf{R}_{e,R} &\approx \mathbf{R}_{e,I} \\ &\approx \mu^{1/2}(4\mathbf{\Gamma})^{-1/2} \quad , \end{aligned} \quad (3.57)$$

i.e., the error at the output of the MMSE LT becomes circularly symmetric and hence, the variance of the components of  $\mathbf{e}_{L,R}$  is halved with respect to the variance of the components of  $\mathbf{e}$ . It is straightforwardly verified that, with reference to the  $\ell$ th symbol, the MMSE LT achieves a lower SER with respect to the MMSE WLT when the following inequality is verified:

$$\frac{\mathbf{R}_{e,R}(\ell, \ell)}{\mathbf{R}_{\tilde{\mathbf{e}}}(\ell, \ell)} \leq 1 \Rightarrow \frac{\tilde{\mu}\gamma_\ell}{\mu\lambda_\ell} \geq 0.25 \quad . \quad (3.58)$$

The approach based on the random channel model proposed in subsection 3.4.1 allows us to foresee the scenarios where (3.58) occurs:

$\beta \ggg 1$  by means of (3.37) and (3.38), the strong approximations  $\frac{\gamma_1}{\lambda_1} \rightarrow 1$  and  $\frac{\gamma_B}{\lambda_B} \rightarrow 1$  can be adopted; hence,  $\tilde{\mu} \approx \mu$  and (3.58) holds true;

$B \ll \text{rank}(\mathbf{H}), \sigma_n^2 \rightarrow 0$  the MIMO channel is decomposed into a large number ( $\text{rank}(\mathbf{H})$ ) of *eigen subchannels* with respect to the number of transmitted symbols ( $B$ ); since  $\frac{\gamma_\ell}{\lambda_\ell} \geq 0.5$  (by means of (3.37) and (3.38)), it can happen that (3.58) is verified for the strongest *eigen subchannels*.

Finally, when  $\beta \rightarrow 0$ , one has that  $\lambda_\ell \approx 2\gamma_\ell$ . In such a case it is easily verified that:

$$\frac{\mathbf{R}_{e,R}(\ell, \ell)}{\mathbf{R}_{\tilde{e}}(\ell, \ell)} \leq 1 \Rightarrow \frac{\gamma_\ell}{\lambda_\ell} \geq 0.5 \quad , \quad (3.59)$$

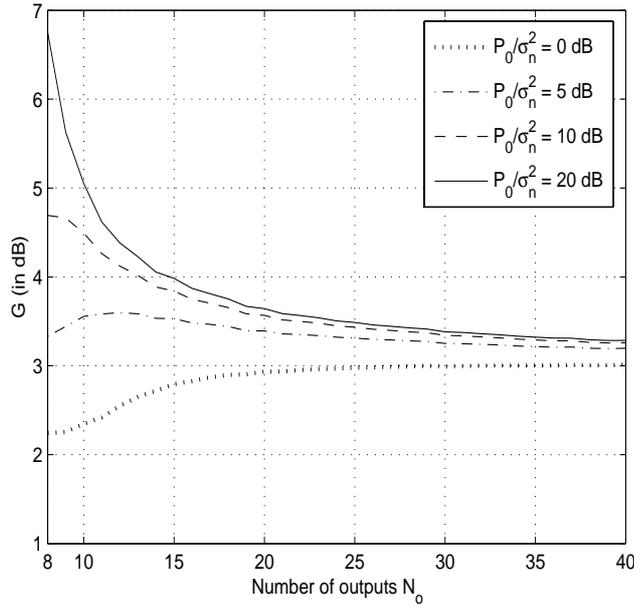
i.e., the MMSE LT can not outperform the MMSE WLT in terms of SER.

### 3.6 Numerical results

In this section, we present a numerical performance analysis of the proposed transceiver structure. In our experiments, the MIMO channel matrix entries  $h^{(\ell,i)}$  are randomly generated according to a complex-valued circularly symmetric zero-mean uncorrelated Gaussian process with variance 2 (i.e.,  $E[(\Re\{h^{(\ell,i)}\})^2] = E[(\Im\{h^{(\ell,i)}\})^2] = 1$ ). The noise components at the output of the MIMO channel are uncorrelated complex-valued circularly symmetric with the same variance  $\sigma_n^2$ , i.e.,  $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_{N_o}$  and  $\mathbf{R}_{nn^*} = \mathbf{0}$ . In each experiment, the channel matrix is known at both the transmitter and the receiver, that are jointly designed according to the MMSE criterion and accounting for the available power  $\mathcal{P}_0 = 1$ .

In a first set of experiments, the (achieved) MSE and the probability of transmitting all the  $B$  information symbols are adopted as performance measure to compare the MMSE WLT and the MMSE LT. The performance parameter curves have been averaged over 500 independent channel realizations. Afterwards, we consider a set of experiments where the performances are evaluated in terms of symbol error rate  $\text{Prob}(\hat{s}_\ell \neq s_\ell)$  (with a slight abuse of notation, we have denoted with  $\hat{s}_\ell$  the output of the decision device). The symbol error rate (SER) curves have been averaged over 100 independent channel realizations by Monte Carlo simulations; moreover, in this study, we admit that at least 10 errors would occur for the lowest level of symbol error rate, resulting in a 95% confidence interval [63].

The performance comparison between the MMSE WLT and the conventional MMSE LT has been carried out by accounting only for that experiments where the two transceivers can transmit the same number of information symbols.



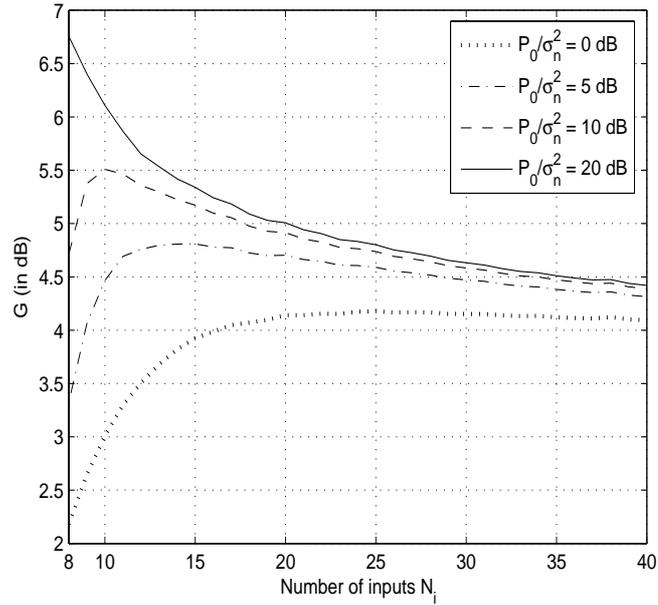
**Figure 3.4:** Average gain  $G$  in dB versus  $N_o \geq 8$  for  $\text{SNR}_i = 0\text{dB}, 5\text{dB}, 10\text{dB}, 20\text{dB}$  ( $B = N_i = 8$ ).

### MSE results

Consider a MIMO system over which  $B = b_r = 8$  independent real-valued information symbol streams have to be transmitted. Accounting for (3.26) and (3.18), let us define (with reference to the  $\ell$ th information symbol) the MSE gain

$$g_\ell \triangleq \frac{\mathbf{R}_e^{(opt)}(\ell, \ell)}{\mathbf{R}_{\tilde{e}}^{(opt)}(\ell, \ell)} \quad \ell = 1, \dots, B \quad (3.60)$$

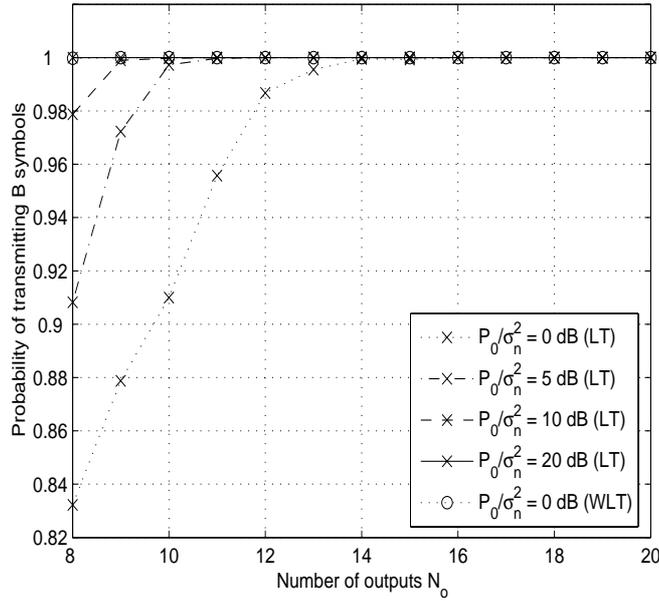
provided by the MMSE WLT over the MMSE LT. In Fig. 3.4, a MIMO channel with  $N_i = 8$  inputs has been considered, and the gain  $g_\ell$  (in dB) averaged over the  $B$  information symbols, i.e.,  $G \triangleq (1/B) \cdot \sum_{\ell=1}^B g_\ell$ , has been plotted versus  $N_o \geq 8$  for different values of the ratio of the total transmit power  $\mathcal{P}_0$  to the noise variance  $\sigma_n^2$  (say,  $\text{SNR}_i$ ). In such a scenario, the MMSE WLT outperforms the MMSE LT. The WL processing allows us to achieve considerable



**Figure 3.5:** Average gain  $G$  in dB versus  $N_i \geq 8$  for  $\text{SNR}_i = 0\text{dB}, 5\text{dB}, 10\text{dB}, 20\text{dB}$  ( $B = N_o = 8$ ).

performance gain when  $N_o$  is comparable with  $N_i$ . In fact, as shown in Section 3.4, when  $\beta \rightarrow 1$  the adoption of the MMSE LT increases the probability of having weak *eigen subchannels*. On the other hand, as expected from (3.41), the gain  $G$  approaches to 3dB when  $\beta \rightarrow 0$ . Analogously, in Fig. 3.5, a MIMO channel with  $N_o = 8$  outputs has been considered, and the gain  $G$  has been plotted versus  $N_i$  for different values of  $\text{SNR}_i$ . As expected from (3.30), the MMSE WLT outperforms the MMSE LT. However, the gain  $G$  decreases when  $N_i$  increases, i.e., when  $\beta \rightarrow \infty$ . In both the considered scenarios, when the number of the outputs and the number of inputs increase, the performance gain provided by the exploitation of the spatial redundancy becomes relevant with respect to the gain provided by the exploitation of the statistical redundancy by means of the WL processing. For such a reason,  $G$  was expected to be a decreasing function as  $\beta \rightarrow 0$  or  $\beta \rightarrow \infty$ .

In the next experiment, we have evaluated the capability of the considered transceivers to transmit all the  $B = 8$  information symbols over a  $N_o \times 8$



**Figure 3.6:** Probability of allocating all the  $B$  information symbols versus  $N_o$ , for  $\text{SNR}_i = 0\text{dB}, 5\text{dB}, 10\text{dB}, 20\text{dB}$ .

MIMO channel, with  $N_o \geq 8$ . To this end, let us define the probabilities:  $P_L \triangleq \text{Prob}(\bar{B} = B)$  and  $P_{WL} \triangleq \text{Prob}(\bar{B}_r = B)$ , where  $\bar{B}$  and  $\bar{B}_r$  denotes the number of symbols allocated by the LT precoder and by the WLT precoder, respectively.

Fig. 3.6 shows the behavior of  $P_L$  and  $P_{WL}$  versus  $N_i$ , for different values of  $\text{SNR}_i$ . The MMSE LT appears more sensitive (with respect to the MMSE WLT) to the variation of both  $\text{SNR}_i$  and  $N_o$ . More specifically, the results in Fig. 3.6 confirm the efforts of MMSE LT in allocating all the  $B$  symbols in low- $\text{SNR}_i$  environments when  $\beta \approx 1$ . On the other hand, the MMSE WLT is able to transmit the  $B$  information symbols ever since  $\text{SNR}_i = 0\text{dB}$ .

Since the MMSE WLT structure depends on the parameter  $n_r$ , let us define (with reference to the  $\ell$ th information symbol) the MSE gain

$$q_\ell^{(k)} \triangleq \frac{[\mathbf{R}_{\bar{e}}^{(opt)}(\ell, \ell)]_{n_r=N_i}}{[\mathbf{R}_{\bar{e}}^{(opt)}(\ell, \ell)]_{n_r=k}} \quad \begin{array}{l} \ell = 1, \dots, B \\ k = 0, \dots, N_i - 1 \end{array} \quad (3.61)$$

provided by the MMSE WLT synthesized by setting  $n_r = k$  over the MMSE WLT synthesized by setting  $n_r = N_i$ . The gain  $q_\ell^{(k)}$  (in dB) averaged over the  $B$  information symbols is reported in Table 3.1 for different values of  $N_o$ , and assuming  $\left(\frac{\mathcal{P}_0}{\sigma_n^2}\right) = 10\text{dB}$ .

$n_r \setminus N_o$	$N_o = 8$	$N_o = 16$	$N_o = 24$	$N_o = 32$
$n_r = 0$	3.671	2.495	2.036	1.728
$n_r = 1$	3.402	2.299	1.850	1.563
$n_r = 2$	3.113	2.067	1.656	1.417
$n_r = 3$	2.802	1.8407	1.449	1.243
$n_r = 4$	2.393	1.564	1.244	1.057
$n_r = 5$	2.034	1.283	1.002	0.868
$n_r = 6$	1.517	0.975	0.752	0.639
$n_r = 7$	0.888	0.561	0.415	0.355

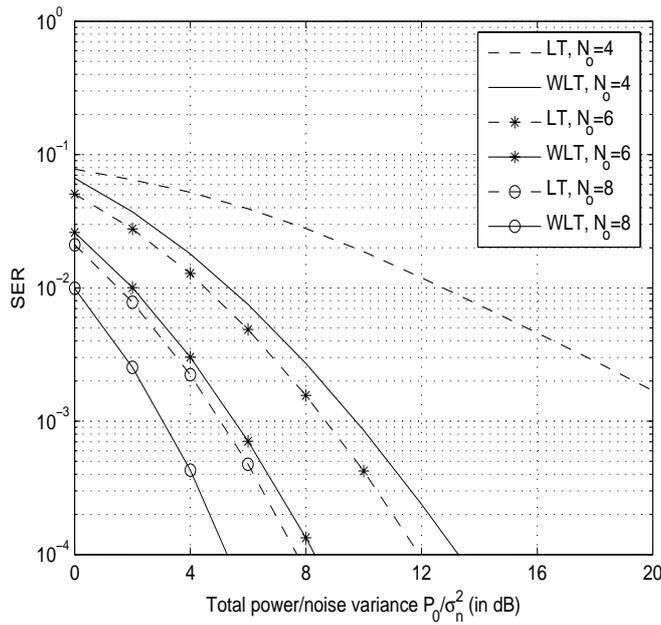
**Table 3.1:** Average  $q_\ell^{(k)}$  (in dB) versus  $0 \leq n_r < N_i$  and  $N_o = 8, 16, 24, 32$  ( $\mathcal{P}_0/\sigma_n^2 = 10\text{dB}$ )

The results show that the MSE achieved by the MMSE WLT reduces when smaller values of  $n_r$  are considered, i.e., when the transmitted vector  $\mathbf{x}$  is complex-valued. Let us point out that, as expected from (3.50), the average gain  $q_\ell^{(0)}$  converges, as  $\beta \rightarrow 0$ , to 0dB; however, the experiments have showed a very slow convergence.

### SER results

In the following experiments we consider a MIMO channel over which  $B = b_r = 4$  independent BPSK symbols have to be transmitted. In Fig. 3.7, a  $4 \times N_o$  MIMO channel have been considered and the SERs achieved by the MMSE LT and the MMSE WLT have been plotted versus  $\text{SNR}_i$ , as the number of outputs increases ( $N_o = 4, 6, 8$ ). The MMSE WLT outperforms the MMSE LT and it provides a considerable gain over the MMSE LT when square MIMO channels ( $\beta = 1$ ) are considered. As  $N_o$  increases, the performance gain provided by the MMSE WLT reduces due to the increasing spatial redundancy, as also confirmed by the curves in Fig. 3.4.

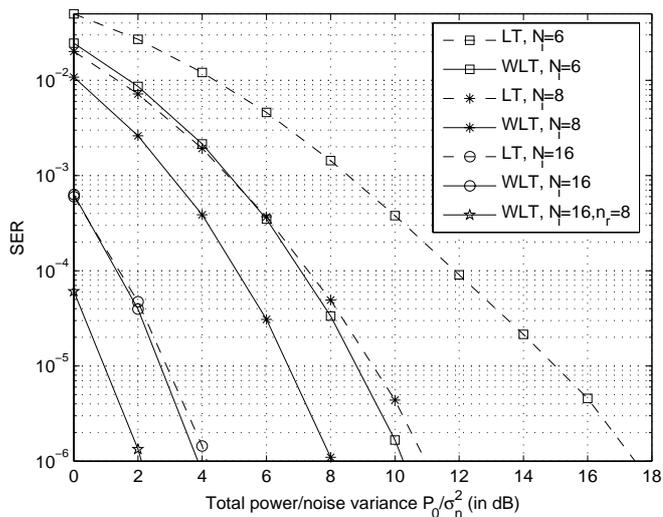
In Fig. 3.8, we have considered an  $N_i \times 4$  MIMO channel. The SER has been plotted versus  $\text{SNR}_i$ , for different values of  $N_i = 4, 6, 8, 16$ . As



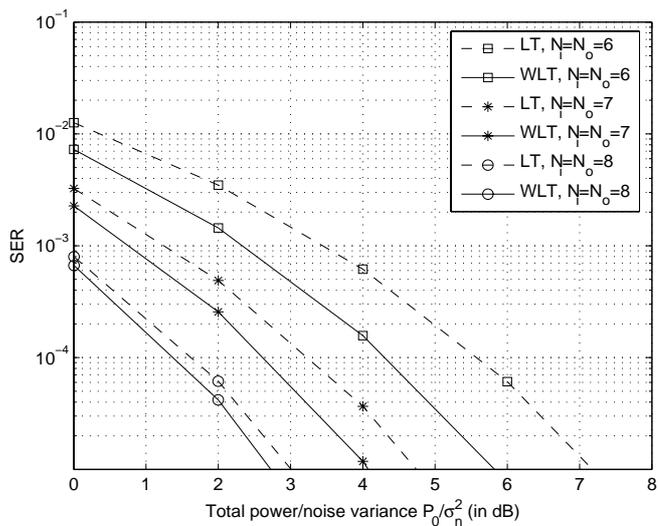
**Figure 3.7:** Symbol error rates of the considered transceivers versus  $\text{SNR}_i$ , for different values of  $N_0$ .

the number of inputs increases, the performance gain provided by the MMSE WLT reduces due to the increasing spatial redundancy, and, when  $N_i = 16$ , the two transceivers perform equivalently. The performance analysis showed that the MMSE LT slightly outperforms the MMSE WLT when  $N_i = 32$ , as expected from the SER analysis carried out in Section 3.5. For completeness, in Fig. 3.8, the SER achieved by the MMSE WLT with  $n_r = 8$  has been plotted: as expected from the MSE analysis in Table 3.1, the performances of MMSE WLT improve when  $n_r$  reduces.

Finally, in Fig 3.9, we have considered the scenario where the information symbols has to be transmitted over an  $N \times N$  MIMO channel. The SER has been plotted versus  $\text{SNR}_i$ , for different values of  $N = 6, 7, 8$ . The performance gain provided by the MMSE WLT over the MMSE LT reduces as  $N$  increases, i.e., when  $B$  becomes much smaller than  $\text{rank}(\mathbf{H})$ . For the largest value of  $N$ , as expected from the SER analysis carried out in Section 3.5, the MMSE LT outperforms the MMSE WLT.



**Figure 3.8:** Symbol error rates of the considered transceivers versus  $\text{SNR}_i$ , for different values of  $N_i$ .



**Figure 3.9:** Symbol error rate of the considered transceiver over a square MIMO channel of size  $N = 6, 7, 8$ .

# Conclusion

In this thesis, the exploitation of widely linear filtering in both MIMO channel equalization and transceiver design has been proposed.

It has been shown that the widely-linear/widely linear decision-feedback MMSE equalizer, without requiring a significant increase in computational complexity, allows one to achieve considerable performance improvements over the DF-MMSE equalizer based on the linear filtering when circularly variant signals are transmitted. The mismatch analysis has also confirmed that the DF-based structures are more sensitive to channel mismatch. More specifically, the performance analysis has shown that the proposed DF-based equalizers, which resort to the widely linear processing, perform satisfactorily, provided that the mismatch channel percentage is limited to roughly 10%, which is, on the other hand, a severe level of mismatching.

Since the widely-linear/widely linear decision-feedback equalizer synthesis can be based on both the real-valued signal representation and the complex-valued one, the issue concerning the choice between the two representations has been addressed. We have recognized that they lead to two nonequivalent structures when the receiver exploits not only past decisions, but also the available current decisions (in Chapter 2 referred to as Scenario 2). The performance comparison between such structures has shown that real-valued representation-based structure outperforms the complex-valued representation-based one since it is more tolerant to the effects of decision errors in the feedback filter. Moreover, with reference to such a receiver strategy, the issue of decision ordering, widely studied in DF equalization over non-dispersive channels, has been addressed with reference to dispersive environments. Specifically, it has been shown that the DF-based equalizers can be expressed as a two-stage equalizer: the former is the optimum time-dispersive equalizer which exploits only past decisions (in Chapter 2 referred to as Scenario 1), whereas the latter performs a non-dispersive equalization exploiting current decisions and, moreover, performs the optimization over the decision

ordering. Such a framework allows one to extend any suboptimum ordering method proposed for non-dispersive environment to the dispersive one.

In the last part of the thesis, the exploitation of the WL filtering has been proposed in the transceiver design when channel state information is available at both the transmitter and the receiver side. According to the MMSE criterion, it has been shown that the transceiver outperforms the linear transceiver when circularly variant symbols are transmitted. Nevertheless, in the spacial case where the symbols to be transmitted are real-valued and the precoder is constrained to be real-valued, the MSE improvements provided by the WL transceiver are not always translated into symbol error rate improvements: specifically, it has been shown that if the number of channel inputs is much larger than the number of channel outputs, or if the number of *eigen subchannels* (in which the MIMO channel can be decomposed) is much larger than the number of symbols to be transmitted, then the linear transceiver can outperform the WL transceiver in terms of symbol error rate. Finally, it has been shown that the optimum WL transceiver degenerates into the linear one when complex-valued circularly symmetric information symbols have to be transmitted.

# Appendix A

## A.1 (Non)-circular random variables and vectors

When dealing with narrowband signals, the linear bandpass single-input single-output channel can be represented by an equivalent  $2 \times 2$  MIMO channel whose inputs and outputs are the in-phase and the quadrature components of the transmitted signal (channel input) and the received one (channel output), respectively. If the in-phase and the quadrature components of the (additive) noise are independent of each other, then a memory-less channel reduces to a couple of independent SISO channels over which the in-phase and the quadrature components of the signal are transmitted independently and do not interfere with each other; in other words, the matrix describing the  $2 \times 2$  equivalent MIMO channel is diagonal. On the other hand, channel memory causes interference between the in-phase and the quadrature components of the received signal, yielding to two non independent SISO transmissions. To simplify the mathematical formulation of the problem, the base-band channels are described in terms of complex-valued signals and complex-valued channel impulse responses [3]. For this reason, the MIMO systems in 1.1 have been introduced according to such a complex-valued representation.

The theory on the complex-valued random variables and vectors is obtained by simply extending the theory of the real-valued ones. For example, the correlation matrix of the complex-valued  $N \times 1$  random vector  $\mathbf{z}$  is defined as  $\mathbf{R}_{zz} \triangleq E[\mathbf{z}\mathbf{z}^H]$ , i.e., it is obtained by simply substituting in  $E[\mathbf{z}\mathbf{z}^T]$ , the correlation matrix of a real-valued vector, the transpose operator with the Hermitian one. Nevertheless, it can happen that the vector  $\mathbf{z}$  and  $\mathbf{z}^*$  are correlated with each other: it follows that the only  $\mathbf{R}_{zz}$  is not sufficient to describe the statistical properties of the random vector  $\mathbf{z}$ , but it is necessary to take into account for the nonzero correlation matrix  $\mathbf{R}_{zz^*} \triangleq E[\mathbf{z}\mathbf{z}^T]$ , namely the

*pseudo-correlation matrix*. A simple example is represented by the real-valued random vectors (or, equivalently, random variables), whose pseudo-correlation matrix is obviously nonzero and such that  $\mathbf{R}_{zz^*} = \mathbf{R}_{zz}$ . Random vectors (and variables), like the real-valued ones, having nonnull pseudo-correlation are said to be rotationally variant. However, the matrix  $\mathbf{R}_{zz^*}$  is rarely introduced in the signal processing literature, and the main reason lies in the fact that it is implicitly or explicitly assumed to be zero.

In the following subsections, we briefly introduce some basic concepts about the (non)circular symmetry of the complex-valued random vectors (see also [39, 68]).

### A.1.1 Definition of circularity

Let us first consider the case of a random variable  $z = z_c + jz_s$ , where  $z_c$  and  $z_s$  denote the real and imaginary parts of  $z$ , respectively; for simplicity, we assume  $z$  having zero-mean.

**Definition A.1** *The complex-valued random variable  $z$  is circularly symmetric if for any  $\alpha$ ,  $z$  and  $z_\alpha \triangleq ze^{j\alpha}$  have the same probability distribution function (p.d.f.).*

The invariance of the p.d.f. to the rotation  $\alpha$  explains the definition of circular symmetry. By defining  $A \triangleq |z|$  and  $\phi \triangleq \angle z$ , the circularity of  $z$  is characterized by the following factorization of the p.d.f.:

$$f_z(\zeta) = f_{z_c, z_s}(\zeta_c, \zeta_s) = f_{A, \Phi}(a, \phi) = \frac{1}{2\pi} f_A(a) \quad , \quad (\text{A.1})$$

i.e., the amplitude  $A$  with arbitrary  $f_A(a)$  is independent of the phase  $\phi$ , which is uniformly distributed in  $[0, 2\pi]$ .

Consider now the complex-valued random vector  $\mathbf{z}$  whose components are the zero-mean random variables  $z^{(k)}$  with  $k = 1, \dots, N$ .

**Definition A.2** *The random vector  $\mathbf{z}$  is marginally circularly symmetric if its scalar components  $z^{(k)}$  are circularly symmetric random variables.*

Accordingly, the p.d.f. of each component  $z^{(k)}$  is given by (A.1) where  $f_A(a)$  can now depend on  $k$ .

**Definition A.3** *The random vector  $\mathbf{z}$  is weakly circularly symmetric if its p.d.f. is equal to the p.d.f. of  $\mathbf{z}e^{j\alpha}$  for any  $\alpha$ .*

Let us denote with  $f_{\mathbf{z}}(\xi) = f_{\mathbf{A}, \Phi}(\mathbf{a}, \phi)$  the p.d.f. of  $\mathbf{z}$  as a function of the random amplitude vector  $\mathbf{A} = [A_1 \ A_2 \ \dots \ A_N]$  and of the random phase

vector  $\Phi = [\phi_1 \ \phi_2 \ \dots \ \phi_N]$ . It can be easily verified that the following implication hold when  $\mathbf{z}$  is weakly circularly symmetric:

$$\begin{aligned} f_{\mathbf{A},\Phi}(\mathbf{a},\phi) &= f_{\mathbf{A},\Phi}(\mathbf{a},\phi_1 + \alpha, \phi_2 + \alpha, \dots, \phi_N + \alpha) \quad \forall \alpha \\ &\Downarrow \\ f_{\mathbf{A},\Phi}(\mathbf{a},\phi) &= f_{\mathbf{A},\Phi}(\mathbf{a},\phi_2 - \phi_1, \phi_2 - \phi_1, \dots, \phi_N - \phi_1) \end{aligned} \quad (\text{A.2})$$

i.e., the p.d.f. of a weakly circularly symmetric random vector depends on only  $N - 1$  random phases.

**Definition A.4** *The random vector  $\mathbf{z}$  is strongly circularly symmetric if its p.d.f. is equal to the p.d.f. of the random vector  $[z^{(1)}e^{j\alpha_1} \ z^{(2)}e^{j\alpha_2} \ \dots \ z^{(N)}e^{j\alpha_N}]^T$ , for any  $\alpha_1, \alpha_2, \dots, \alpha_N$ .*

In such a case, it can be verified that the p.d.f. of  $\mathbf{z}$  is equal to:

$$f_{\mathbf{A},\Phi}(\mathbf{a},\phi) = \left(\frac{1}{2\pi}\right)^N f_{\mathbf{A}}(\mathbf{a}) \quad , \quad (\text{A.3})$$

i.e., the phases  $\phi_k$  are i.i.d random variables uniformly distributed in  $[0, 2\pi]$  and are independent of  $\mathbf{A}$ . Note that when the components of  $\mathbf{z}$  are independent of each other,  $\mathbf{z}$  is said to be *totally circularly symmetric* since  $f_{\mathbf{A}}(\mathbf{a})$  can be factorized in  $N$  functions.

### A.1.2 Second order analysis of a complex-valued random vector

Consider the zero-mean complex-valued random vectors  $\mathbf{z}_1 = \mathbf{z}_{1,c} + j\mathbf{z}_{1,s}$  and  $\mathbf{z}_2 = \mathbf{z}_{2,c} + j\mathbf{z}_{2,s}$  of size<sup>1</sup>  $N$ . To specify the correlation matrix  $\mathbf{R}_{\mathbf{z}_1\mathbf{z}_2} \triangleq E[\mathbf{z}_1\mathbf{z}_2^H]$  of the two random vectors  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , we need to specify the following (real-valued) correlation matrices:

$$\begin{bmatrix} E[\mathbf{z}_{1,c}\mathbf{z}_{2,c}^T] & E[\mathbf{z}_{1,c}\mathbf{z}_{2,s}^T] \\ E[\mathbf{z}_{1,s}\mathbf{z}_{2,c}^T] & E[\mathbf{z}_{1,s}\mathbf{z}_{2,s}^T] \end{bmatrix} \quad . \quad (\text{A.4})$$

However, it can happen that the statistical properties of the considered vectors can not be fully described by the only  $\mathbf{R}_{\mathbf{z}_1\mathbf{z}_2}$ . As a matter of fact, we can introduce the pseudo-correlation matrix

$$\mathbf{R}_{\mathbf{z}_1\mathbf{z}_2^*} \triangleq E[\mathbf{z}_1\mathbf{z}_2^T] \quad (\text{A.5})$$

---

<sup>1</sup>For simplicity, in the following we assume that  $\mathbf{z}_1$  and  $\mathbf{z}_2$  have the same dimension.

where the Hermitian transpose operator in the definition of  $\mathbf{R}_{z_1 z_2}$  has been replaced by the transpose operator. Let us note that the knowledge of the matrices in (A.4) is equivalent to the knowledge of both the correlation and the pseudo-correlation matrices. In fact, one has:

$$\begin{aligned} E[\mathbf{z}_{1,c} \mathbf{z}_{2,c}^T] &= \frac{1}{2} \Re \{ \mathbf{R}_{z_1 z_2} + \mathbf{R}_{z_1 z_2^*} \} \\ E[\mathbf{z}_{1,s} \mathbf{z}_{2,s}^T] &= \frac{1}{2} \Re \{ \mathbf{R}_{z_1 z_2} - \mathbf{R}_{z_1 z_2^*} \} \\ E[\mathbf{z}_{1,s} \mathbf{z}_{2,c}^T] &= \frac{1}{2} \Im \{ \mathbf{R}_{z_1 z_2} + \mathbf{R}_{z_1 z_2^*} \} \\ E[\mathbf{z}_{1,c} \mathbf{z}_{2,s}^T] &= \frac{1}{2} \Im \{ \mathbf{R}_{z_1 z_2} - \mathbf{R}_{z_1 z_2^*} \} . \end{aligned} \quad (\text{A.6})$$

The vectors  $\mathbf{z}_1$  and  $\mathbf{z}_2$  are uncorrelated if and only if the four matrices in (A.4) vanish or, equivalently, if and only if both the correlation and the pseudo-correlation matrices vanish.

The case  $\mathbf{z} = \mathbf{z}_1 = \mathbf{z}_2$  allows one to define the autocorrelation matrix  $\mathbf{R}_{zz} \triangleq E[\mathbf{z}\mathbf{z}^H]$  and the pseudo-correlation matrix  $\mathbf{R}_{zz^*} \triangleq E[\mathbf{z}\mathbf{z}^T]$  of the vector  $\mathbf{z}$ . Moreover, given the correlation matrices

$$\mathbf{R}_{z_c z_c} \triangleq E[\mathbf{z}_c \mathbf{z}_c^T] \quad \mathbf{R}_{z_s z_s} \triangleq E[\mathbf{z}_s \mathbf{z}_s^T] \quad \mathbf{R}_{z_c z_s} \triangleq E[\mathbf{z}_c \mathbf{z}_s^T] , \quad (\text{A.7})$$

it is easily verified that

$$\mathbf{R}_{zz} = \mathbf{R}_{z_c z_c} + \mathbf{R}_{z_s z_s} + J[\mathbf{R}_{z_c z_s}^T - \mathbf{R}_{z_c z_s}] \quad (\text{A.8})$$

$$\mathbf{R}_{zz^*} = \mathbf{R}_{z_c z_c} - \mathbf{R}_{z_s z_s} + J[\mathbf{R}_{z_c z_s}^T + \mathbf{R}_{z_c z_s}] . \quad (\text{A.9})$$

According to Definition A.3, if  $\mathbf{z}$  is weakly circularly symmetric, then  $\mathbf{z}$  and  $\mathbf{z}e^{j\alpha}$  have the same p.d.f. and, consequently, one has that

$$\mathbf{R}_{zz^*} (1 - e^{j2\alpha}) = \mathbf{0} \quad \forall \alpha \quad (\text{A.10})$$

implying that the pseudo-correlation matrix  $\mathbf{R}_{zz^*}$  vanishes when  $\mathbf{z}$  exhibits weak circularity. The condition  $\mathbf{R}_{zz^*} = \mathbf{0}$  is equivalent to

$$\mathbf{R}_{z_c z_c} = \mathbf{R}_{z_s z_s} \quad \text{and} \quad \mathbf{R}_{z_c z_s} = -\mathbf{R}_{z_c z_s}^T \quad (\text{A.11})$$

or, in other words, the real and the imaginary parts of  $\mathbf{z}$  have the same correlation matrix, and their cross correlation is skew-symmetric. Note that the skew-symmetry of  $\mathbf{R}_{z_c z_s}$  implies that  $\mathbf{R}_{z_c z_s}$  has a zero main diagonal, which

means that the real and imaginary part of each component of  $\mathbf{z}^{(k)}$  are uncorrelated. The vanishing of  $\mathbf{R}_{z_c z_s}$  does not, however, imply that the real part of  $\mathbf{z}^{(k)}$  and the imaginary part of  $\mathbf{z}^{(\ell)}$  are uncorrelated for  $\ell \neq k$ .

Based on the above argumentations, we give the following definitions of circularly symmetric and of jointly circularly symmetric random vectors:

**Definition A.5** *The complex-valued  $N \times 1$  random vector  $\mathbf{z}$  is said to be circularly symmetric if the pseudo-correlation matrix  $\mathbf{R}_{zz^*}$  is zero. Conversely, a nonzero pseudo-correlation matrix characterizes the rotationally variant vectors.*

**Definition A.6** *The complex-valued  $N \times 1$  random vectors  $\mathbf{z}_1$  and  $\mathbf{z}_2$  are jointly circularly symmetric if the stacked vector  $[\mathbf{z}_1 \ \mathbf{z}_2]^T$  is circularly symmetric.*

Finally, by omitting the demonstration which can be found in [40], it can be shown that:

**Proposition A.1** *Let  $\mathbf{R}_{zz}$  be a complex-valued Hermitian positive definite matrix, and let  $\mathbf{R}_{zz^*}$  be a complex-valued symmetric matrix; then  $\mathbf{R}_{zz^*}$  is a pseudo-correlation matrix of  $\mathbf{z}$  if and only if  $\mathbf{Q} \triangleq \mathbf{R}_{zz}^* - \mathbf{R}_{zz^*}^H \mathbf{R}_{zz}^{-1} \mathbf{R}_{zz^*}$  is non negative definite.*



## Appendix B

In this appendix, we propose an algorithm to derive the DF-MMSE equalizers in (2.37) and (2.40) in the case where the input and noise sequences are spatially and temporally uncorrelated. By exploiting the diagonal structure of the input and noise correlation matrices, the proposed algorithm exhibits computational complexity still grows quadratically with the system parameters.

To this aim, in absence of decision errors, rewrite the DF equalizer output in (2.34) as follows:

$$\begin{aligned}
 \hat{\mathbf{x}}_{k-\Delta} &= \mathbf{W}^H \mathbf{T}_{N_f} [\mathbf{y}_k, k] - \mathbf{B}^H \mathbf{T}_{N_b+1} [\mathbf{x}_k, k - \Delta] + \mathbf{x}_{k-\Delta} & (\text{B.1}) \\
 &= \mathbf{W}^H \tilde{\mathbf{T}}[\mathbf{H}_k, N_f, \Delta + 1] \cdot \mathbf{T}_{\Delta+1} [\mathbf{x}_k, k] \\
 &\quad + \mathbf{W}^H \tilde{\mathbf{T}}[\mathbf{H}_{k+\Delta+1}, N_f, N_b] \cdot \mathbf{T}_{N_b} [\mathbf{x}_k, k - \Delta - 1] & \clubsuit \\
 &\quad + \mathbf{W}^H \mathbf{T}_{N_f} [\mathbf{n}_k, k] - \mathbf{B}^H \mathbf{T}_{N_b+1} [\mathbf{x}_k, k - \Delta] + \mathbf{x}_{k-\Delta}
 \end{aligned}$$

When Scenario 1 is considered, the DF-MMSE equalizer is constituted by the feedback filter that removes the ISI introduced by the post-cursor  $\clubsuit$ , i.e.  $\mathbf{B}^{(opt)H} = \begin{bmatrix} \mathbf{I}_{N_i} & \mathbf{W}^{(opt)H} \tilde{\mathbf{T}}[\mathbf{H}_{k+\Delta+1}, N_f, N_b] \end{bmatrix}$ , and by the MMSE feedforward filter which processes the observation vector

$$\mathbf{r} \triangleq \tilde{\mathbf{T}}[\mathbf{H}_k, N_f, \Delta + 1] \cdot \mathbf{T}_{\Delta+1} [\mathbf{x}_k, k] + \mathbf{T}_{N_f} [\mathbf{n}_k, k] \quad (\text{B.2})$$

and provides the estimate of  $\mathbf{x}_{k-\Delta}$ . The computation of  $\mathbf{W}^{(opt)}$  would generally requires  $O(N_o^3 N_f^3)$  operations since it mainly depends on the computation of the inverse of the correlation matrix  $\mathbf{R}_r \triangleq [\mathbf{r} \mathbf{r}^H]$ . However, such an inversion can be performed efficiently with  $N_i(\Delta + 1)$  iterations, each one requiring  $O(N_o^2 N_f^2)$  operations.

According to the assumptions, one has:

$$\mathbf{R}_x(m) = E [\mathbf{x}_k \mathbf{x}_{k-m}^H] = \mathbf{\Sigma}_x \delta(m) \quad (\text{B.3})$$

$$\mathbf{R}_n(m) = E [\mathbf{n}_k \mathbf{n}_{k-m}^H] = \mathbf{\Sigma}_n \delta(m) \quad (\text{B.4})$$

with  $\mathbf{\Sigma}_x = \text{diag}(\sigma_{x,1}^2, \dots, \sigma_{x,N_i}^2)$  and  $\mathbf{\Sigma}_n = \text{diag}(\sigma_{n,1}^2, \dots, \sigma_{n,N_o}^2)$ . Define the matrices

$$\mathbf{R}_{x_\Delta} \triangleq \tilde{\mathbf{T}}[\mathbf{R}_x(m), \Delta + 1, \Delta + 1] \quad (\text{B.5})$$

$$\mathbf{R}_n \triangleq \tilde{\mathbf{T}}[\mathbf{R}_n(m), N_f, N_f] \quad (\text{B.6})$$

$$\mathbf{H}_\Delta \triangleq \tilde{\mathbf{T}}[\mathbf{H}_k, N_f, \Delta + 1] \quad (\text{B.7})$$

$$\mathbf{H}_{(0)} \triangleq \tilde{\mathbf{T}}[\mathbf{H}_{k+\Delta}, N_f, 1] \quad (\text{B.8})$$

and the correlation matrix

$$\mathbf{R}_r^{(k)} \triangleq \mathbf{R}_n + \sum_{\ell=1}^k \mathbf{H}_\Delta(:, \ell) \mathbf{R}_{x_\Delta}(\ell, \ell) \mathbf{H}_\Delta(:, \ell)^H \quad (\text{B.9})$$

such that  $\mathbf{R}_r = \mathbf{R}_r^{(N_i(\Delta+1))}$ . The inverse of  $\mathbf{R}_r$  can be obtained after  $N_i(\Delta+1)$  steps by utilizing the following recursive algorithm (in the sequel referred to as Algorithm 1) proposed in [69, 70] based on the Sherman-Morrison formula:

step 0

$$k = 0$$

$$\left[ \mathbf{R}_r^{(0)} \right]^{-1} = \mathbf{R}_n^{-1}$$

step k

$$k \leftarrow k + 1$$

$$\begin{aligned} \left[ \mathbf{R}_r^{(k)} \right]^{-1} \leftarrow & \left[ \mathbf{R}_r^{(k-1)} \right]^{-1} - \left[ \mathbf{R}_r^{(k-1)} \right]^{-1} \mathbf{H}_\Delta(:, k) \left[ \mathbf{H}_\Delta(:, k)^H \left[ \mathbf{R}_r^{(k-1)} \right]^{-1} \right. \\ & \left. \cdot \mathbf{H}_\Delta(:, k) + \mathbf{R}_{x_\Delta}^{-1}(k, k) \right]^{-1} \cdot \mathbf{H}_\Delta(:, k+1)^H \left[ \mathbf{R}_r^{(k-1)} \right]^{-1} \end{aligned}$$

end

where the inversion of the matrices  $\mathbf{R}_{x_\Delta}$  and  $\mathbf{R}_n$  can be performed in  $O(N_i(\Delta + 1))$  and  $O(N_o N_f)$  operations, and where each step requires  $O((N_o N_f)^2)$  operations. As discussed in [70], such a procedure is useful when the structure of  $\mathbf{R}_n$  is “nice” in the sense that the effort involved in evaluating

the inverse of the right hand-side of (B.9) is small relative to the effort of inverting a general square matrix of size  $(N_o N_f)$ .

When Scenario 2 is considered, the estimation of the  $q$ th component of the desired vector  $\mathbf{x}_{k-\Delta}$  is performed by processing the difference of the observation vector  $\mathbf{r}_k$  and the interference due to the first  $q-1$  components of  $\mathbf{x}_{k-\Delta}$ . More specifically, let us denote with  $\mathbf{w}^{(i)}$  the  $i$ th column of  $\mathbf{W}$  and let us define the vector  $\mathbf{x}_{k-\Delta}[q] \triangleq [\mathbf{0}_{1 \times q} \quad \mathbf{x}_{k-\Delta}(q+1 : N_i, 1)^T]^T$  with  $0 \leq q \leq N_i - 1$ . The optimum  $\mathbf{w}^{(opt)}(q+1)$  is the MMSE filter that processes the observation vector

$$\mathbf{r}[q] \triangleq \tilde{\mathbf{T}}[\mathbf{H}_k, N_f, \Delta] \mathbf{T}_\Delta[\mathbf{x}_k, k] + \tilde{\mathbf{H}}_{(0)} \mathbf{x}_{k-\Delta}[q] + \mathbf{T}_{N_f}[\mathbf{n}_k, k] \quad (\text{B.10})$$

to estimate the desired symbol  $x_{k-\Delta}^{(q+1)}$ . From (B.10), one has that

$$\mathbf{r}[q+1] = \mathbf{r}[q] - \mathbf{H}_{(0)}(1 : N_o N_f, q+1) x_{k-\Delta}^{(q+1)} \quad (\text{B.11})$$

and

$$\begin{aligned} \mathbf{R}_r[q+1] &\triangleq E[\mathbf{r}[q+1] \mathbf{r}^H[q+1]] \\ &= \mathbf{R}_r[q] - \mathbf{H}_{(0)}(:, q+1) \sigma_{x, q+1}^2 \mathbf{H}_{(0)}(:, q+1)^H \end{aligned} \quad (\text{B.12})$$

The derivation of the optimum feedforward filter column  $\mathbf{w}^{(opt)}(q+1)$  follows from the inversion of the correlation matrix  $\mathbf{R}_r[q]$ , which generally requires  $O((N_o N_f)^3)$  operations. However, it is simple to verify that, given  $\mathbf{R}_r^{-1}[q-1]$ , the inversion of the matrix  $\mathbf{R}_r[q]$  can be efficiently performed in  $O((N_o N_f)^2)$  operations by exploiting the Sherman-Morrison formula and by inverting the right hand-side of (B.13) rather than the left hand-side one. It follows that, for Scenario 2,  $\mathbf{W}_2^{(opt)}$  is computed with the following algorithm (in the sequel referred to as Algorithm 2) after  $N_i$  steps:

step 0

$$k = 0$$

$$\mathbf{R}_r[0] \leftarrow \mathbf{R}_r$$

$$\mathbf{w}_2^{(opt)}(1) \leftarrow [\mathbf{R}_r[0]]^{-1} \mathbf{H}_{(0)}(:, 1) \sigma_{x,1}^2$$

step k

$$k \leftarrow k + 1$$

$$\begin{aligned}
[\mathbf{R}_r[k]]^{-1} &\leftarrow [\mathbf{R}_r[k-1]]^{-1} - [\mathbf{R}_r[k-1]]^{-1} \mathbf{H}_{(0)}(:, k) \\
&\quad \cdot \left( \mathbf{H}_{(0)}(:, k)^H [\mathbf{R}_r[k-1]]^{-1} \mathbf{H}_{(0)}(:, k) - \frac{1}{\sigma_{x,k}^2} \right)^{-1} \\
&\quad \cdot \mathbf{H}_{(0)}(:, k)^H [\mathbf{R}_r[k-1]]^{-1}
\end{aligned}$$

$$\mathbf{W}_2^{(opt)}(k+1) \leftarrow [\mathbf{R}_r[k]]^{-1} \mathbf{H}_{(0)}(:, k+1) \sigma_{x,k+1}^2$$

end

where the inverse of  $\mathbf{R}_r[0]$  represents the output of Algorithm 1. Finally, the optimum feedback filter  $\mathbf{B}_2^{(opt)}$  given by

$$\mathbf{B}_2^{(opt)H} = \left[ \mathbf{I}_{N_i} + \mathcal{L}(\mathbf{W}_2^H) \mathbf{H}_{(0)} \quad \mathbf{W}_2^H \tilde{\mathbf{T}}[\mathbf{H}_{k+\Delta+1}, N_f, N_b] \right], \quad (\text{B.13})$$

where  $\mathcal{L}(\cdot)$  is the strictly lower triangular part of the matrix argument, and its computation exhibits an  $O(N_i^2 N_o N_f N_b)$  complexity. We must point out that the computational complexity of the proposed algorithm is in general larger than the one of the algorithm proposed in [9] except for the very special case where  $N_o \ll N_i$  and  $N_f^3 \ll \nu^2$ .

On the other hand, the proposed algorithm is compatible with the V-BLAST ordering algorithm and, unlike [52], preserves a computational complexity which grows quadratically with the system parameters. In such a case, the Algorithm 2 needs to be slightly modified by introducing in `step k` the operations that:

- a) calculate the  $N_i$  decision-point signal-to-noise ratio;
- b) chose the index  $i_k$  related maximum decision-point signal-to-noise ratio to define the V-BLAST ordering  $\{i_1, \dots, i_{N_i}\}$ .

# Bibliography

- [1] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *Europ. Trans. Telecomm.*, 10:585–595, Nov./Dec. 1999.
- [2] P. A. Bello. Characterization of randomly time-variant linear channels. *IEEE Trans. on Commun. Syst.*, 11:360–393, Dec. 1963.
- [3] J.G. Proakis and D.G. Manolakis. *Digital Signal Processing*. Macmillan publishing company, 1988.
- [4] S. Benedetto and E. Biglieri. *Principles of digital transmission*. Kluwer Academic, 1999.
- [5] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj. Outdoor MIMO wireless channels: models and performance prediction. *IEEE Trans. on Commun.*, 50(12):1926–1934, Dec. 2002.
- [6] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Khan. Fading correlation and its effects on the capacity of the multielement antenna systems. *IEEE Trans. on Commun.*, 48:502–513, Mar. 2000.
- [7] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen. A stochastic MIMO radio channel model with experimental validation. *IEEE Journal on Selected Areas in Communications*, 20(6):1211–1226, Aug. 2002.
- [8] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela. Keyholes, correlations, and capacities of multielement transmit and receive antennas. *IEEE Trans. on Wireless Commun.*, 1(2):361–368, Apr. 2002.
- [9] N. Al-Dhahir and Ali H. Sayed. The finite-length multi-input multi-output MMSE DFE. *IEEE Trans. on Signal Processing*, 48(10):2921–2936, Oct. 2000.

- 
- [10] G.J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-elements antennas. *Bell Labs Tech. Journal*, 1(2):41–59, Aug. 1996.
- [11] W. C. Jakes. *Microwave mobile communications*. Wiley, 1974.
- [12] R. Lupas and S. Verdú. Linear multiuser detectors for synchronous code-division multiple-access channels. *IEEE Trans. on Information Theory*, 35:123–136, Jan. 1989.
- [13] A. Duel-Hallen. Multiuser detection for CDMA systems. *IEEE Personal Communication*, pages 46–58, Apr. 1995.
- [14] A. C. McCormick and E. A. Al-Susa. Multicarrier CDMA for future generation mobile communication. *Electronics & Communication Engineering Journal*, pages 52–60, Apr. 2002.
- [15] G. L. Stuber, J. R. Barry, S. W. McLaughlin, Y. (G.) Li, M. A. Ingram, and T. G. Pratt. Broadband MIMO-OFDM wireless communications. *Proceedings of IEEE*, 92(2):271–294, Feb. 2004.
- [16] A. Duel-Hallen. Equalizers for multiple input/multiple output channels and PAM systems with cyclostationarity input sequences. *IEEE Trans. on Communications*, 10:630–639, Apr. 1992.
- [17] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time block code from orthogonal designs. *IEEE Trans. Information Theory*, 45:1456–1467, July 1999.
- [18] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley, 1991.
- [19] A. Edelman. *Eigenvalues and condition numbers of random matrices*. PhD thesis, Dept. of Math., MIT, 1989.
- [20] A. Grant. Rayleigh fading multiple antenna channels. *EURASIP Journal on Appl. Signal Processing*, pages 316–329, 2002.
- [21] V. A. Marčenko and L. A. Pastur. Distributions of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, 1:457–483, 1967.
- [22] J. B. Andersen. Array gain and capacity for known random channels with multiple element arrays at both ends. *IEEE Journal on Selected Areas in Communications*, 18:2172–2178, Nov. 2000.

- 
- [23] J. Jang and S. Roy. On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems. *IEEE Trans. on Commun.*, 42:3221–3231, Dec. 1994.
- [24] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela. V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel. In *International Symposium on Signals, Systems, and Electronics, ISSSE 98*, pages 295–300, 29 Sept.-2 Oct. 1998.
- [25] H. Sampath, P. Stoica, and A. Paulraj. Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion. *IEEE Trans. on Communications*, 49(12):2198–2207, Dec. 2001.
- [26] B. Hassibi and B. M. Hochwald. High-rate codes that are linear in space and time. *IEEE Trans. on Information Theory*, 48(7):1804–1824, July 2002.
- [27] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath. Optimal designs for space-time linear precoders and decoders. *IEEE Trans. on Signal Processing*, 59(5):1051–1064, May 2002.
- [28] S. Catreux, L. J. Greenstein, and V. Erceg. Some results and insights on the performance gains of MIMO systems. *IEEE Journal on Selected Areas in Communications*, 21(5):839–847, June 2003.
- [29] S. M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, Oct. 1998.
- [30] N. Al-Dhahir. A new high-rate differential space-time block coding scheme. *IEEE Trans. Commun. Lett.*, 7(11):540–542, Nov. 2003.
- [31] S. Verdu. *Multiuser detection*. Cambridge Univ. Press, 1998.
- [32] O. Simeone, Y. Bar-Ness, and U. Spagnolini. Linear and nonlinear pre-equalization/equalization for MIMO systems with long-term channel state information at the transmitter. *IEEE Trans. on Wireless Communications*, 3(2):373–378, Mar. 2004.
- [33] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber. Precoding in multiantenna and multiuser communications. *IEEE Trans. on Wireless Commun.*, 3(4):1305–1316, July 2004.

- 
- [34] A. Scaglione, S. Barbarossa, and G. B. Giannakis. Filterbank transceivers optimizing information rate in block transmissions over dispersive channels. *IEEE Trans. on Information Theory*, 45(3):1019–1032, Apr. 1999.
- [35] P. Stoica, Y. Jang, and J. Li. On MIMO channel capacity: an intuitive discussion. *IEEE Signal Processing Magazine*, 22(3):83–84, May 2005.
- [36] Y. Jang, J. Li, and W. Hager. Uniform channel decomposition for MIMO communications. *IEEE Trans. on Signal Processing*, 53(11):4283–4294, Nov. 2005.
- [37] Y. Jiang, J. Li, and W. W. Hager. Joint transceiver design for MIMO communications using geometric mean decomposition. *IEEE Trans. on Signal Processing*, 53(10):3791–3803, Oct. 2005.
- [38] W. M. Brown and R. B. Crane. Conjugate linear filtering. *IEEE Trans. Inform. Theory*, 15:462–465, July 1969.
- [39] B. Picinbono. On circularity. *IEEE Trans. on Signal Processing*, 42(12):3473–3482, Dec. 1994.
- [40] B. Picinbono. Second-order complex random vectors and normal distributions. *IEEE Trans. on Signal Processing*, 44:2637–2640, Oct. 1996.
- [41] G. Gelli, L. Paura, and A.R.P. Ragozini. Blind widely linear multiuser detection. *IEEE Communications Letters*, 4(6):187–189, June 2000.
- [42] D. Mattera, L. Paura, and F. Sterle. Widely linear MMSE equaliser for MIMO linear time-dispersive channel. *Electronics Letters*, 39(20):1481–1482, Oct. 2003.
- [43] W.H. Gerstacker, R. Schober, and A. Lampe. Receivers with widely linear processing for frequency-selective channels. *IEEE Trans. on Communications*, 51(9):1512–1523, Sept. 2003.
- [44] D. Mattera, L. Paura, and F. Sterle. Widely linear decision-feedback equalizer for time-dispersive linear MIMO channels. *IEEE Trans. on Signal Processing*, 53(7):2525–2536, July 2005.
- [45] B. Picinbono and P. Chevalier. Widely linear estimation with complex data. *IEEE Trans. on Signal Processing*, 43(8):2030–2033, Aug. 1995.

- 
- [46] S. Buzzi, M. Lops, and A.M. Tulino. A new family of MMSE multiuser receivers for interference suppression in DS/CDMA systems employing BPSK modulation. *IEEE Trans. on Communications*, 49(1), Jan. 2001.
- [47] T. Kailath and A. H. Sayed. Displacement structure: Theory and applications. *SIAM Review*, 37(6):297–386, Sept. 1995.
- [48] T. Kailath and A. H. Sayed. *Fast reliable algorithms for matrices with structure*. PA: SIAM, 1999.
- [49] M. K. Varanasi. Decision feedback multiuser detection: a systematic approach. *IEEE Trans. on Information Theory*, 45(1):219–240, Jan. 1999.
- [50] D. Wubben, R. Bohnke, J. Rinas, V. Kuhn, and K.D. Kammeyer. Efficient algorithm for decoding layered space-time codes. *IEE Electron. Lett.*, 37(22):1348–1350, Oct. 2001.
- [51] J. Luo, K. R. Pattipati, P. K. Willett, and F. Hasegawa. Optimal user ordering and time labeling for ideal decision feedback detection in asynchronous CDMA. *IEEE Trans. on Communications*, 51(11):1754–1757, Nov. 2003.
- [52] X. Zhu and R. D. Murch. Layered space-time equalization for wireless MIMO systems. *IEEE Trans. on Wireless Comm.*, 2(6):1189–1203, Nov. 2003.
- [53] A. Duel-Hallen. A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels. *IEEE Trans. on Communications*, 43:285–290, Feb.-Apr. 1995.
- [54] G. Woodward, R. Ratasuk, M. L. Honig, and P. B. Rapajic. Minimum mean-squared error multiuser decision-feedback detectors for DS-CDMA. *IEEE Trans. on Communications*, 50(12):2104–2111, Dec. 2002.
- [55] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 1989. II Edn.
- [56] F. D. Neeser and J. L. Massey. Proper complex random processes with application to information theory. *IEEE Trans. on Information Theory*, 39(4):1293–1302, July 1993.

- 
- [57] D. Mattera, L. Paura, and F. Sterle. Widely linear MMSE equalizer for MIMO linear time-dispersive channel. *Proceedings of the ISPA 03*, 2:1176–1180, Sep. 2003.
- [58] D. D. Falconer and G. J. Foschini. Theory of minimum mean-square error QAM systems employing decision feedback equalization. *Bell Syst. Tech. Journal*, 52(10):1821–1849, Dec. 1973.
- [59] R. F. H. Fischer and C. Windpassinger. Real versus complex-valued equalisation in V-BLAST systems. *IEE Electronics Letters*, 39(5):470–471, Mar. 2003.
- [60] D. Mattera, L. Paura, and F. Sterle. Two nonequivalent structures for widely-linear decision-feedback MMSE equalization over dispersive MIMO channels. *working in progress*.
- [61] N. Al-Dahir and J.M. Cioffi. Mismatched finite-complexity MMSE decision feedback equalizers. *IEEE Trans. on Signal Processing*, 45(4):935–944, Apr. 1997.
- [62] N. Al-Dahir and J.M. Cioffi. MMSE decision-feedback equalizers: Finite length results. *IEEE Trans. on Information Theory*, 41(4):961–975, July 1995.
- [63] R. M. Buehrer, N. S. Correal-Mendoza, and B. D. Woerner. A simulation comparison of multiuser receivers for cellular CDMA. *IEEE Trans. on Veh. Technol.*, 49(4):1065–1085, July 2000.
- [64] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas. Joint tx-rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization. *IEEE Trans. on Signal Processing*, 51(9):2381–2401, Sept. 2003.
- [65] G. G. Raleigh and J. M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Trans. on Commun.*, 46(3):357–366, Mar. 1998.
- [66] A. Scaglione, G. B. Giannakis, and S. Barbarossa. Redundant filterbank precoders and equalizers part I: Unification and optimal design. *IEEE Trans. on Signal Processing*, 47(7):1988–2006, July 1999.
- [67] H. Dette. Strong approximation of the eigenvalues of large dimensional Wishart matrices by roots of generalized Laguerre polynomials. *Journal of Approximation Theory*, 118:290–304, 2002.

- [68] P.J. Schreier and L.L. Scharf. Second-order analysis of improper complex random vectors and processes. *IEEE Trans. on Signal Processing*, 51(3):714–725, Mar. 2003.
- [69] K. S. Miller. On the inverse of the sum of matrices. *SIAM Math. Magazine*, 54(2):67–72, Mar. 1981.
- [70] W. W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31(2):221–239, June 1989.

