MULTI-LINE COORDINATED COMMUNICATION FOR BROADBAND ACCESS NETWORKS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Georgios Ginis
June 2002

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John M. Cioffi
(Principal Adviser)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Donald C. Cox

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Andrea J. Goldsmith

Approved for the University Committee on Graduate Studies:

Abstract

THE INTEREST IN DESIGNING high-speed communication systems over metallic wire conductors remains very strong. This interest is exemplified by current efforts to deliver Very-high-speed Digital Subscriber Line (VDSL) systems, Ethernet-in-the-First-Mile (EFM), multi-Gbps Ethernet over copper twisted pairs of short length, and multi-Gbps chip-to-chip serial links over cable or backplane traces. A prevailing characteristic of such systems is that multiple lines emanate from a central node and reach remote terminals. It has been realized that crosstalk among the lines is in several cases the dominant constraining factor for communication.

This thesis examines how *coordination* among signals of different lines during transmission or reception facilitates reliable, high-speed communication. The basic assumption is that coordination is possible only at the central node, since the remote terminals are generally geographically dispersed. A fundamental result of this work is that such coordination in multi-line channels is capable of achieving performance that is very close to that attained in the absence of crosstalk.

The multi-line channel is modeled as a Multiple-Input-Multiple-Output (MIMO) channel, and multicarrier transmission with a cyclic prefix is assumed. It is observed that the matrix representation of the channel corresponding to a specific frequency has a characteristic structure: The diagonal elements are in all scenarios dominant across the columns for upstream transmission, or dominant across the rows for downstream transmission. Considering each remote terminal as an individual user, it is observed that upstream and downstream transmission can be viewed as a multiple access channel and a broadcast channel, respectively.

The Generalized Decision Feedback Equalizer (GDFE) is reviewed, and its interesting connections with known equalization structures such as the Minimum-Mean-Square-Error Decision Feedback Equalizer (MMSE-DFE) are highlighted. The utility of the GDFE for MIMO communication scenarios is brought forward, and its relation with the Vertical Bell-Labs Space Time architecture (V-BLAST) is revealed.

Using the zero-forcing version of the GDFE as a foundation, and incorporating Discrete Multi-Tone (DMT) transmission, techniques for crosstalk mitigation are derived. These involve coordination at the central node, employing successive cancellation at the upstream receivers, and modulo precoding at the downstream transmitters. It is shown that an equivalent approach for computing the parameters of these structures is through the QR factorization of the channel matrix. As a consequence of the diagonally dominant structure of the channel matrices, crosstalk suppression in multi-line channels with these techniques is very close to perfect cancellation.

In the context of coordinated communication, transmission optimization is also treated. The objective is the optimum allocation of transmitted power across frequency and among the users in order to maximize a weighted sum of the user data rates. It is found that in its simplest form the solution degenerates into independent water-filling for each user. Under the additional requirement of power backoff for protection of existing services, the problem can only be solved computationally. However, the characterization of the problem as convex allows for efficient search algorithms to be utilized. Also, the case of joint energy allocation and upstream/downstream frequency selection is considered, where a straightforward solution has a computational cost that grows exponentially with the number of tones. An alternative approach is adopted: A relaxation of the original problem is formed, which is computationally tractable, and still yields a near-optimum solution.

Finally, an adaptive implementation of the upstream coordinated receivers is presented, utilizing decorrelation conditions as the adaptation criteria. The technique is based on the expression of the unitary feedforward matrix operation as the product of Givens rotations terms. After allowing for an adaptation period, the receiver is capable of achieving a satisfactory performance level in the absence of channel knowledge and without requiring a training sequence.

Acknowledgements

I HAVE BEEN EXTRAORDINARILY FORTUNATE in associating with a very remarkable set of people during my doctoral degree program.

I would like to deeply thank my advisor, Prof. John Cioffi, for his guidance and support over the past four years. His insight, intuition and knowledge have never ceased to amaze me.

I wish to thank Prof. Andrea Goldsmith for serving as my associate advisor and as a reader of this dissertation. I am grateful to Prof. Donald Cox for serving as a member of my orals committee and as a reader. I would also like to thank Prof. Anthony Siegman for serving as the chairman of my orals committee. My gratitude extends to Prof. Zeke Bar-Ness for his guidance on the material of Chapter 6.

I would like to thank Wei Yu, Avneesh Agrawal and Vladimir Stojanovich for collaborating and co-authoring papers with me, and Hichan Moon for sharing with me ideas related to Appendix C. My interaction with members and visitors of the research group has been extremely beneficial. Many thanks to Acha, Ana, Ardavan, Atul, Carlos, Christian, Daniel, Joe, John, Joon-Suk, Jose, Kok-Wui, Louise, Rohit, Steve, Susan, Wonjong, Won-Joon, and Zi-Ning, as well as to the new generation, Amal, Dimitris, Ghazi, Jeannie, Jungwon, Kee-Bong, and Taek. I wish to express my appreciation to Carlos and Kee-Bong for their computer administration work. I am also especially grateful to Joice DeBolt for her administration work, as well as for being a model of kindness.

I wish to thank my several friends at Stanford that helped make the five past years very enjoyable. Particularly, I would like to mention Dimitris, Menelaos, Stavros and Yorgos. Certainly, I want to thank Ioanna for her patience and support.

Last, but very importantly, I would like to thank my family for motivating me to reach my dreams.

Georgios Ginis
June 2002

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The need for the design of higher-speed communication systems appears to be insatiable. Since the ultimate information sources and sinks are humans, one could claim that this need will subside when the end-to-end performance of communication networks approaches the capacity of the human mind to process information [1]. But this argument becomes irrelevant with the emergence of virtual agents that act on behalf of humans, and thus increase the communication needs regardless of the limits of the human mind.

This need is addressed by a variety of technologies. Wireless communication offers the unique benefit of untethered operation, and waveguide (especially optical) communication already provides unparalleled speeds. Still, vast numbers of communication systems currently operate over metallic wire conductors, and there is evidence that the interest in such systems will remain strong in the forthcoming years. Their applications are very varied, and they range from Digital Subscriber Line (DSL) systems, to Local Area Networks (LANs) over copper, to serial links over backplane traces for chip-to-chip communication.

The resilience of metallic wire conductor communication media can be attributed mostly to two factors. First, there is a huge installed base of copper twisted pairs. Both the telephone subscriber loop infrastructure, and the vast majority of in-building wiring are such examples. Second, in several scenarios metallic media are easier to install, and they offer the advantage of not requiring signal conversions (e.g. electrical to optical, or electrical to radio and vice versa), which leads to communication devices of smaller cost and size.

Metallic wires present a very challenging environment for digital transmission. However, advances in signal processing and coding, combined with the exponential improvements in silicon technology, have realized performance leaps that were previously considered impossible. Although many existing systems have achieved an impressive level of sophistication, their design is focused on providing point-to-point communication, and mostly ignores the fact that there is a multiplicity of users. This thesis explores the theoretical gains that can be obtained from addressing the multiuser aspects, and presents techniques capable of realizing these gains.

In the remainder of this chapter, Section 1.1 gives an overview of current metallic wire systems, and provides the motivation for this thesis. Section 1.2 outlines the contents of the following chapters, and Section 1.3 summarizes the research contributions. Section 1.4 presents the notation used in this dissertation.

1.1 Motivation

1.1.1 Communication Systems over Metallic Wire Conductors

DSL systems provide a solution to the problem of *last mile access* by utilizing the existing copper twisted pair infrastructure, which was originally built with the purpose of providing telephone service. The effort to upgrade from copper to fiber will probably not affect DSL adversely in the following years. Such replacement requires huge capital outlays, and the cost increases steeply as one attempts to deploy fiber closer and closer to the customer. Efforts to extend fiber installation from the *exchange* to the *cabinet* are actually beneficial for DSL, since they improve its reach and data rate. The most successful DSL system is currently Asymmetric DSL (ADSL) [2], and a promising next generation system is Very-high-speed DSL (VDSL) [3, 4].

LANs over copper can be characterized as ubiquitous. The IEEE standard 802.3 (Ethernet) [5] has proven to be the most popular and resilient to time. The latest physical layer specification for copper is 1000BASE-T [6], and refers to operation at 1Gbps. There is considerable interest in developing a 10Gbps system over copper. The main benefits of such a system over its fiber counterpart are argued to be the existence of a large installed copper base and its competitiveness in terms of both cost and size of the equipment.

It is interesting to observe the evolution of chip-to-chip data buses into communication systems employing equalization and coding techniques. As processor speeds grow, the need for faster links between chips becomes imperative. Given the limited room for channel re-engineering, a number of existing and proposed systems have adopted multi-level transmission and equalization [7, 8, 9, 10, 11].

The systems described above have common characteristics, besides the mentioned fact that they all use metallic wire conductors as their communication media. Their typical design topology involves a number of lines that emanate from a *central node* and reach *remote terminals.* Additionally, signals transmitted on a line cause interference into receivers of neighboring lines. Such interference is in most settings so severe that it represents a major transmission impairment. In the following, these characteristics are discussed in the context of DSL systems, but several of the observations also hold for the other systems.

1.1.2 DSL Evolution

Figure 1.1 illustrates the current DSL topology [12]. Copper twisted pairs emanate from a Central Office (CO) and terminate at the remote equipment of the customers. The pairs are physically close for a large part of their length, thus signals transmitted on a pair result in crosstalk noise at receivers of the other pairs. Several tele-communication companies (telcos) may provide DSL service on the pairs of a single cable. In the current environment, each telco has physical access to the pairs that it services, and consequently has control over the corresponding modems and network equipment.

Inherently, there are no means for the equipment of independent telcos to *cooperate.* Since signal transmission on pairs owned by one telco may negatively affect

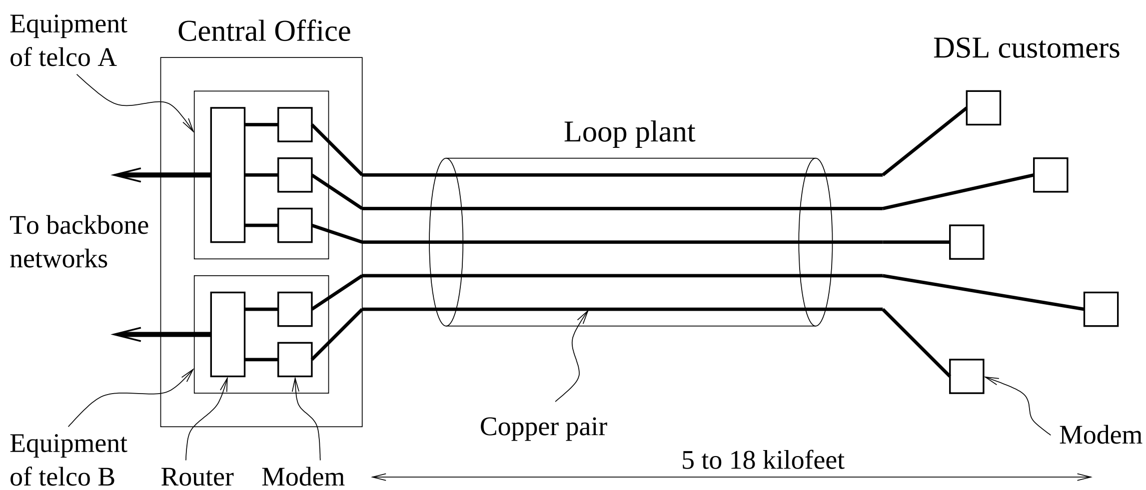

Figure 1.1: Topology of current DSL deployment.

transmission on pairs owned by another telco, rules must be applied to prevent excessive levels of crosstalk. In the absence of cooperation among the telcos, the restrictions are necessarily conservative, in order to ensure that transmission on any pair is not severely degraded.

Figure 1.2 illustrates the future DSL topology, as it will most likely emerge from efforts to replace copper with fiber. An Optical Network Unit (ONU) is introduced between the CO and the customers, and a fiber link connects the CO with the ONU.

The ONU performs multiplexing/demultiplexing operations between the DSL lines and the fiber, and the CO performs the inverse operations. Multiple telcos may still provide service, but the ONU is typically controlled by a single entity. In this case, the telcos do not have physical access to the pairs that they service.

The topology of Figure 1.2 has a number of implications. Since the pairs have shorter lengths (and thus smaller attenuation), higher data rates can be achieved. But most importantly, the modems of the ONU are all controlled by a single entity. Therefore, they can employ *cooperative* strategies for signal transmission or reception. Communication can be treated as point-to-multipoint, or multipoint-to-point, instead of the current simplistic approach of point-to-point links.

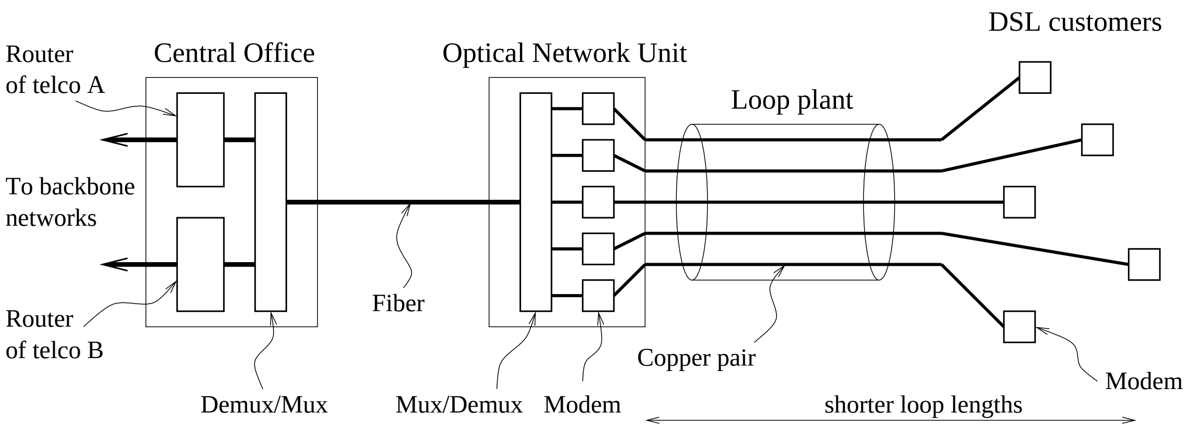

Figure 1.2: Topology of future DSL deployment.

This topology is also pervasive in several LANs and chip-to-chip links. Obtaining a better understanding of communication in such scenarios has been the primary motivation for the research results presented in the following chapters. The next section explains the layout of this dissertation, and highlights its contributions.

1.2 Outline

The layout of the material is such that more abstract and theoretical concepts are presented first, and these are followed by practical techniques. The latter are accompanied by simulation results. Most simulations use DSL as the transmission environment, since DSL systems represent a prime candidate for the adoption of the described techniques. In more detail, the outline of the thesis is as follows:

- Chapter 2 introduces the multi-line communication channel, for which a Multiple-Input-Multiple-Output (MIMO) linear model is adopted. The assumption of multiple tone transmission with a cyclic prefix allows an equivalent representation of the channel as a set of matrices, each corresponding to a specific tone frequency. An important result in this chapter is that these matrices have a distinct property: For upstream, they are characterized as *column-wise diagonally* *dominant*, which means that the magnitude of a diagonal element is always much larger than the magnitude of the other elements in the same column. For downstream, they are characterized as *row-wise diagonally dominant*, where the magnitude of a diagonal element is always much larger than the magnitude of the other elements in the same row. This property is fundamental for multi-line systems, and has significant consequences for the methods presented in the later chapters.

It is also observed that the multi-line channel is a multiuser channel, and specifically it becomes a *vector gaussian multiple access channel* for upstream transmission, and a *vector gaussian broadcast channel* for downstream transmission. For these multiuser channels, information theoretic results are reviewed, and capacity region bounds are defined.

- Chapter 3 reviews the Generalized Decision Feedback Equalizer (GDFE), and re-derives it as a MIMO version of the Minimum-Mean-Square-Error Decision Feedback Equalizer (MMSE-DFE) employed for Inter-Symbol Interference (ISI) channels. A proof of the asymptotic equivalence between the GDFE and the MMSE-DFE is given in Appendix A. Also, the Vertical Bell-Labs Space Time Architecture (V-BLAST) receiver processing is shown to be a special case of the GDFE, and the proof is given in Appendix B. The GDFE serves as the foundation for the coordinated transmission and reception methods of the next chapters.

- Chapter 4 derives methods for removing Far-End Crosstalk (FEXT) by combining Discrete-Multi-Tone (DMT) modulation with the zero-forcing version of the GDFE. For upstream transmission, the structure is comprised of multiple decision feedback crosstalk cancelers at the receiving side, while for downstream transmission, the structure consists of multiple modulo precoders at the transmitting side. Each of the cancelers and precoders operates on a specific tone. A result of considerable interest is that the zero-forcing GDFE parameters can be computed easily by means of the QR factorization. The most significant finding of this chapter is that such zero-forcing coordinated transmission or reception is capable of achieving data rates that lie very close to those that would be attained in a crosstalk-free environment. Additionally, the user data rates are very close to those corresponding to the case of only one user utilizing the entire channel.

The derived methods may be combined with known channel coding methods in a straightforward way. Appendix C describes an algorithm for the computation of parameters defining the *distance spectrum* of codes with trellis descriptions. Tables with the parameter values of a large number of convolutional codes are included.

- Chapter 5 treats the subject of transmission optimization for coordinated communication. The optimization criterion is defined as the maximization of the weighted sum of the user data rates. The optimization parameters are the allocation of energy across frequencies and users, and optionally the allocation of frequency tones for transmission in each direction. Three scenarios are examined: First, it is assumed that energy constraints apply independently for each user, in which case it is found that within certain approximations the solution amounts to water-filling for each user individually. Then, the assumption is that the energy allocation must be constrained, in order to prevent excessive crosstalk into non-coordinated neighboring systems. Although in this case the solution cannot be reduced to water-filling, it is deduced that the problem is convex, and efficient numerical algorithms are employed to arrive at a solution. Finally, it is supposed that both the energy allocation and the allocation of tones to upstream or downstream must be optimized. Although an optimal solution involves a prohibitive computational cost that grows exponentially with the number of tones, a relaxation of the original problem is shown to be convex. This fact allows the derivation of a near-optimum solution with a moderate computational cost.

- Chapter 6 presents a blind receiver training method for upstream coordinated communication. The feedforward matrix of each crosstalk canceller is found in Chapter 4 to be unitary, thus, it can be expressed as the multiplication of

Givens rotation terms. A Givens rotation term is characterized by a single parameter, that can be blindly adapted using an appropriate decorrelation criterion. The feedback matrix coefficients are adapted using similar decorrelation criteria. The algorithm convergence is demonstrated through simulation, and the performance is compared against a receiver with perfect channel knowledge.

1.3 Research Contributions

This dissertation investigates the concept of coordination for multi-line systems at several levels. These levels include the characterization of the multi-user channel, the development of coordinated communication techniques, and the derivation of algorithms for the optimal allocation of the communication resources. The contributions of this work are listed next by chapter:

- In Chapter 2, it is shown that multi-carrier transmission can under certain conditions be extended to the case of multi-line systems. It is indicated that the channel matrices of multi-line systems have a *diagonally dominant* structure, which is fundamental for the results appearing in the later chapters. Although the observation about the matrix structure is quite intuitive, it appears that its consequences for the multi-user channel have not been previously realized.

- In Chapter 3, the GDFE is re-derived using an approach that differs from that of previous developments. Additionally, two new results concerning the relationship of the GDFE with the MMSE-DFE for ISI channels and with V-BLAST are presented. The detailed proofs of these results are given in Appendices A and B.

- In Chapter 4, crosstalk cancellation methods are derived for both upstream and downstream communication by combining the GDFE with DMT modulation. The upstream cancellation scheme constitutes a novel application of successive interference cancellation for multi-line systems. The downstream scheme is a new development extending the concept of precoding to multi-user systems.

It is shown that the performance of these schemes is very close to that of a channel with no crosstalk. Additionally, each user's performance is shown to approximate the performance that would be achieved if no other user utilized the channel.

- In Chapter 5, it is shown that the task of optimum energy allocation for the crosstalk cancellation methods can be decomposed into independent water-fillings for each of the users. When energy allocation is combined with the need for power back-off, it is shown that the resulting optimization problem is convex and that a solution can be obtained numerically. In the case of joint energy allocation and frequency planning, a technique yielding a near-optimum solution is applied. This technique has previously been employed in different problems, such as subcarrier allocation for multiuser Orthogonal-Frequency-Division-Multiplexing (OFDM).

- In Chapter 6, a method for blind receiver training of the upstream coordinated system is developed. This method extends the previously known *bootstrap* technique, so that multi-user decision feedback is included.

1.4 Notation

In this dissertation, the following conventions hold:

- Vectors are denoted by lower-case bold letters.
- Matrices are denoted by upper-case letters.
- A^T is the transpose of matrix A.
- A^* is the conjugate-transpose of matrix A.
- $A^{-T} = (A^{-1})^T$ is the transpose of the inverse of matrix A.
- $A^{-*} = (A^{-1})^*$ is the conjugate-transpose of the inverse of matrix A.
- $E(\mathbf{x})$ is the expectation of the random vector $\mathbf{x}$.
- $R_{\mathbf{xx}} = E(\mathbf{xx}^*)$ is the covariance matrix of the random vector $\mathbf{x}$.
- $R_{\mathbf{xy}} = E(\mathbf{xy}^*)$ is the cross-covariance matrix of the random vectors $\mathbf{x}$ and $\mathbf{y}$.
- $|A|$ is the determinant of matrix A.
- I_k is the $k \times k$ identity matrix.
- $O_{k\times n}$ is the $k \times n$ matrix with all elements being 0.
- $\mathrm{diag}(A_1, \ldots, A_K)$ is the diagonal (or block diagonal) matrix formed by $A_1, \ldots, A_K$.
- $\mathrm{diagonal}(A)$ is the diagonal matrix formed by the diagonal elements of A.
- $\Re(x)$ is the real part of x.
- $\Im(x)$ is the imaginary part of x.
- $\lfloor x \rfloor$ is the largest integer smaller or equal to x.
- $< x >$ is the time average of x.

Chapter 2

Multiuser Channels with Multiple Lines

COMMUNICATION CHANNELS with multiple lines emanating from a central node and reaching remote terminals arise in a wide variety of systems, such as DSL, LANs and chip-to-chip links. Although there is typically strong coupling among the lines, the current design approach treats each line as a point-to-point connection. This chapter adopts the more general view of considering the multi-line channel as a multi-user system.

In Section 2.1, the transmission environment is presented, the various effects are described, and channel models are introduced. The principle of multi-carrier modulation is then extended to multiuser systems. It is seen that the multiuser channels have distinct properties that have major importance for the techniques that are derived in the later chapters. Although these properties are fairly intuitive, their significance for multiuser transmission is a novel observation. Section 2.2 illuminates the multiuser aspects of the channel, reviews information theoretic results, and expresses capacity region bounds. In later chapters, these bounds are compared against the achievable performance of the proposed techniques.

2.1 Multi-line Channels

2.1.1 Crosstalk

The topology of a multi-line system with a central node is shown in Figure 2.1. Depending on the application, the central node may be called a Central Office (CO), a Remote Terminal (RT), an Optical Network Unit (ONU), or a Hub. The geographically dispersed remote users are connected with the central node through metallic wire conductors. Transmission of information from the remote users to the central node is referred to as *upstream* communication, while transmission of information in the opposite direction is referred to as *downstream* communication. Although the communication media are not shared among users, electromagnetic coupling among the lines implies that signal transmission on one line causes interfering signals known as *crosstalk* on neighboring lines.

Crosstalk is a dominant impairment in multi-line systems, and its impact becomes more severe as higher frequencies are utilized for communication. Far-End Crosstalk (FEXT) describes the coupled signals that originate from the end opposite of the affected receiver, while Near-End Crosstalk (NEXT) describes the coupled signals that originate from the same end as the affected receiver. Figure 2.1 illustrates both NEXT and FEXT signals induced by upstream transmission of user 1. A number of systems suppress NEXT by "separating" downstream and upstream transmission either through Time Division Duplex (TDD) or by Frequency Division Duplex (FDD) [13].

2.1.2 Channel Model

The channel model for the topology of Figure 2.1 is now presented. It is assumed that either FDD or TDD is employed, so that NEXT is eliminated. It is also assumed that the transceivers, both at the central node and at the remote equipment, have exactly the same sampling period. This assumption is reasonable for the transceivers at the central node, because their clock signal typically originates from a common generator. It also holds for the transceivers at the remote equipment, because the

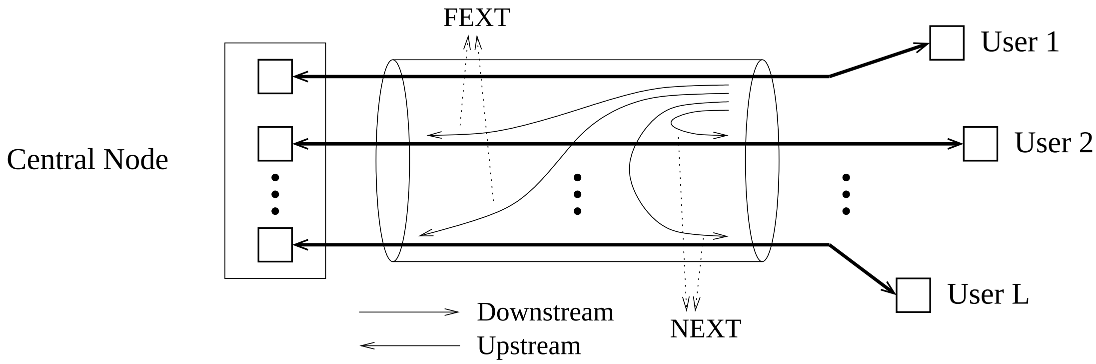

Figure 2.1: Crosstalk effects with transmission over multiple lines.

clock signal in that case can be derived from the received signals.

The sampled output for a specific user for either upstream or downstream transmission depends on the present and past input symbols of both the intended user and the other crosstalking users. A block of N output samples for user i satisfies

$$\begin{bmatrix} y_{i,N-1} \\ y_{i,N-2} \\ \vdots \\ y_{i,0} \end{bmatrix} = H^c_{i,1} \cdot \begin{bmatrix} x_{1,N-1} \\ x_{1,N-2} \\ \vdots \\ x_{1,0} \\ \vdots \\ x_{1,-\nu} \end{bmatrix} + \cdots + H^c_{i,i} \cdot \begin{bmatrix} x_{i,N-1} \\ x_{i,N-2} \\ \vdots \\ x_{i,0} \\ \vdots \\ x_{i,-\nu} \end{bmatrix} + \cdots + H^c_{i,L} \cdot \begin{bmatrix} x_{L,N-1} \\ x_{L,N-2} \\ \vdots \\ x_{L,0} \\ \vdots \\ x_{L,-\nu} \end{bmatrix} + \begin{bmatrix} n_{i,N-1} \\ n_{i,N-2} \\ \vdots \\ n_{i,0} \end{bmatrix}, \tag{2.1}$$

where L is the number of users, and ν represents the maximum memory of the transfer and crosstalk coupling functions expressed in number of samples. Also, $H^c_{i,1}, \ldots, H^c_{i,L}$ are convolution matrices derived from the line transfer function and the crosstalk coupling functions, $y_{i,j}$ is the j-th output sample of receiver i, $n_{i,j}$ is the j-th noise sample of receiver i, and $x_{i,j}$ is the j-th symbol of user i. The $N \times (N+\nu)$ convolution matrix $H^c_{i,j}$ is defined as

$$H^c_{i,j} = \begin{bmatrix} h_{i,j}(0) & h_{i,j}(1) & \cdots & h_{i,j}(\nu) & 0 & \cdots & 0 \\ 0 & h_{i,j}(0) & h_{i,j}(1) & \cdots & h_{i,j}(\nu) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{i,j}(0) & h_{i,j}(1) & \cdots & h_{i,j}(\nu) \end{bmatrix}, \tag{2.2}$$

where $h_{i,j}(n)$, $n = 0, \ldots, \nu$ is the impulse response of length $\nu + 1$ for a signal transmitted on line j and received on line i. Equation (2.1) is expressed more compactly as

$$\mathbf{y}_i = H^c_{i,1}\mathbf{x}^p_1 + \cdots + H^c_{i,i}\mathbf{x}^p_i + \cdots + H^c_{i,L}\mathbf{x}^p_L + \mathbf{n}_i \ , \tag{2.3}$$

where $\mathbf{y}_i = [y_{i,N-1} \ y_{i,N-2} \ \ldots \ y_{i,0}]^T$, $\mathbf{x}^p_k = [x_{k,N-1} \ x_{k,N-2} \ \ldots \ x_{k,0} \ \ldots \ x_{k,-\nu}]^T$, and $\mathbf{n}_i = [n_{i,N-1} \ n_{i,N-2} \ \ldots \ n_{i,0}]^T$. The noise samples represent the superposition of several noise sources such as crosstalk from *alien* systems, radio frequency ingress, impulse noise, and background noise. In the following, impulsive noise sources are ignored, and the emphasis is placed on crosstalk, which typically represents the dominant noise source and is commonly approximated as Gaussian [14]. Although crosstalk is generally non-white, the existence of noise-whitening filters is assumed at the receivers. Therefore, $\mathbf{n}_i$ is considered to be white and Gaussian, and $R_{\mathbf{n}_i\mathbf{n}_i} = I$.

2.1.3 Channel Representation with Multiple Tone Transmission

The following two conditions are necessary in order to have independent transmission within each tone, and they are assumed hereafter:

- All users employ block transmission with a cyclic prefix (CP) of length ν. The operation of cyclic prefix insertion is shown in Figure 2.2, and more details are given in [15]. The mathematical justification is shown later, but an intuitive explanation for this requirement is that the cyclic prefix makes the channel impulse response appear periodic within a block. Thus, the channel impulse response may be expressed as a sum of discrete tones. Given an input consisting

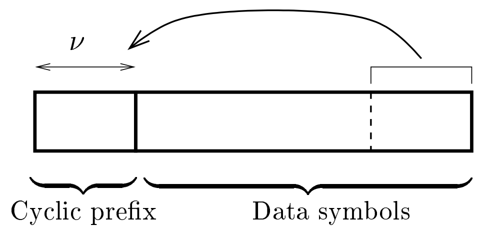

Figure 2.2: Cyclic prefix insertion.

of independent tones, the channel output also becomes a sum of independent tones.

- Block transmission and reception at the central node is synchronized as illustrated in the timing diagram of Figure 2.3. Such synchronization makes it feasible to process the blocks of all users jointly. As seen in the diagram, synchronized block transmission is relatively straightforward to achieve, but synchronized block reception requires that the block boundaries for upstream transmission be aligned in such a way, so that the blocks of all users arrive simultaneously at the central node. This assumption is realistic, given that the central node is controlled by a single entity.

Taking the above into account, (2.3) becomes

$$\mathbf{y}_i = H_{i,1}\mathbf{x}_1 + \cdots + H_{i,i}\mathbf{x}_i + \cdots + H_{i,L}\mathbf{x}_L + \mathbf{n}_i \ , \tag{2.4}$$

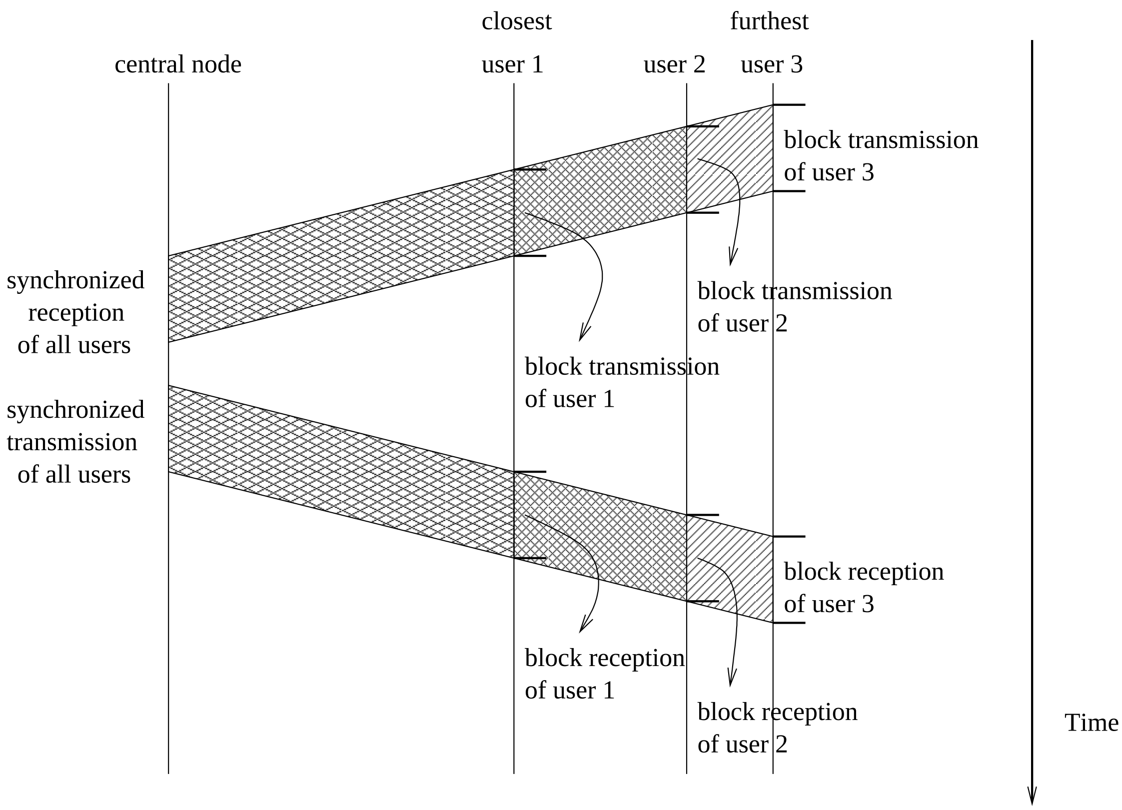

Figure 2.3: Timing diagram for synchronized block transmission and reception at the central node.

where $\mathbf{x}_k = [x_{k,N-1}\ x_{k,N-2}\ \ldots\ x_{k,0}]^T$, and $H_{i,j}$ is a circulant matrix defined by

$$H_{i,j} = \begin{bmatrix} h_{i,j}(0) & h_{i,j}(1) & \cdots & h_{i,j}(\nu) & 0 & \cdots & 0 \\ 0 & h_{i,j}(0) & h_{i,j}(1) & \cdots & h_{i,j}(\nu) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{i,j}(0) & h_{i,j}(1) & \cdots & h_{i,j}(\nu) \\ h_{i,j}(\nu) & 0 & \cdots & 0 & h_{i,j}(0) & \cdots & h_{i,j}(\nu-1) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h_{i,j}(1) & \cdots & h_{i,j}(\nu) & 0 & \cdots & 0 & h_{i,j}(0) \end{bmatrix} . \tag{2.5}$$

Combining the L users, (2.4) becomes

$$\mathbf{y} = H\mathbf{x} + \mathbf{n} , \tag{2.6}$$

where $\mathbf{y} = \left[\mathbf{y}_1^T\, \mathbf{y}_2^T\, \ldots\, \mathbf{y}_L^T\right]^T$, $\mathbf{x} = \left[\mathbf{x}_1^T\, \mathbf{x}_2^T\, \ldots\, \mathbf{x}_L^T\right]^T$, $\mathbf{n} = \left[\mathbf{n}_1^T\, \mathbf{n}_2^T\, \ldots\, \mathbf{n}_L^T\right]^T$, and H is a matrix whose (i, j) block is $H_{i,j}$. The noise covariance matrix is $R_{\mathbf{nn}} = I$, implying that the noise of different receivers is uncorrelated. This is an approximation, which may potentially incur some loss in practical systems.

By making use of the Discrete Fourier Transform (DFT) modulation principle presented in [16, 15, 17], a frequency-domain representation of the channel can be obtained. An Inverse Discrete Fourier Transform (IDFT) operation is performed on each transmitted data block (prior to appending the CP), and a DFT operation is performed on each received data block (after discarding the CP). Thus,

$$\begin{aligned} \tilde{\mathbf{y}} &= Q_{mDFT} H Q_{mIDFT} \tilde{\mathbf{x}} + Q_{mDFT}\mathbf{n} \\ &= \Lambda \tilde{\mathbf{x}} + \tilde{\mathbf{n}} , \end{aligned} \tag{2.7}$$

where

$$\tilde{\mathbf{y}} = \left[\tilde{\mathbf{y}}_1^T \, \tilde{\mathbf{y}}_2^T \, \ldots \, \tilde{\mathbf{y}}_L^T\right]^T = Q_{mDFT}\mathbf{y} \ , \tag{2.8}$$

$$\mathbf{x} = Q_{mIDFT}\left[\tilde{\mathbf{x}}_1^T \, \tilde{\mathbf{x}}_2^T \, \ldots \, \tilde{\mathbf{x}}_L^T\right]^T = Q_{mIDFT}\tilde{\mathbf{x}} \ , \tag{2.9}$$

$$\tilde{\mathbf{n}} = \left[\tilde{\mathbf{n}}_1'^T \, \tilde{\mathbf{n}}_2'^T \, \ldots \, \tilde{\mathbf{n}}_L'^T\right]^T = Q_{mDFT}\mathbf{n} \ , \tag{2.10}$$

$$Q_{mIDFT} = Q_{mDFT}^* = \operatorname{diag}\left(\overbrace{Q_{IDFT}, Q_{IDFT}, \ldots, Q_{IDFT}}^{L}\right) \ . \tag{2.11}$$

The $N \times N$ matrix Q_{IDFT} is such that, when right-multiplied with a vector, produces its N-point IDFT, and $Q_{DFT} = Q_{IDFT}^*$ is the corresponding DFT matrix. Then, $\tilde{\mathbf{y}}_k$, $\tilde{\mathbf{x}}_k$, $\tilde{\mathbf{n}}_k$ each contain N samples corresponding to user k, $R_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}} = I$, and Λ is a matrix whose (i, j) block is the diagonal matrix $\Lambda_{i,j} = Q_{DFT}H_{i,j}Q_{IDFT}$. The diagonal elements of $\Lambda_{i,j}$ equal the DFT values of $h_{i,j}(n)$, $n = 0, \ldots, \nu$.

Therefore, (2.7) gives a channel description, where the samples are stacked in groups corresponding to users, and each of the groups contains samples corresponding to tones. It is desirable to reorganize these samples, so that they are stacked in groups corresponding to tones, and each of these groups contains samples corresponding to different users. To this end, an $NL \times NL$ permutation matrix P is defined, composed of the $N \times N$ blocks $P_{i,j}$, and $i, j = 1, \ldots, L$. All elements of $P_{i,j}$ equal 0, except for a 1 at position (j, i). Some inspection reveals that, when matrix P is right-multiplied with a vector of size NL, its elements are essentially reordered from L groups of N components into N groups of L components. Also, note that $P^{-1} = P^* = P$. Applying this reordering operation to both the transmitter and the receiver samples, yields

$$\begin{aligned}
\mathbf{z}' &= P\Lambda PP\tilde{\mathbf{x}} + P\tilde{\mathbf{n}} \\
&= P\Lambda P\mathbf{v}' + \mathbf{n}_z' \\
&= H_F\mathbf{v}' + \mathbf{n}_z' \ ,
\end{aligned} \tag{2.12}$$

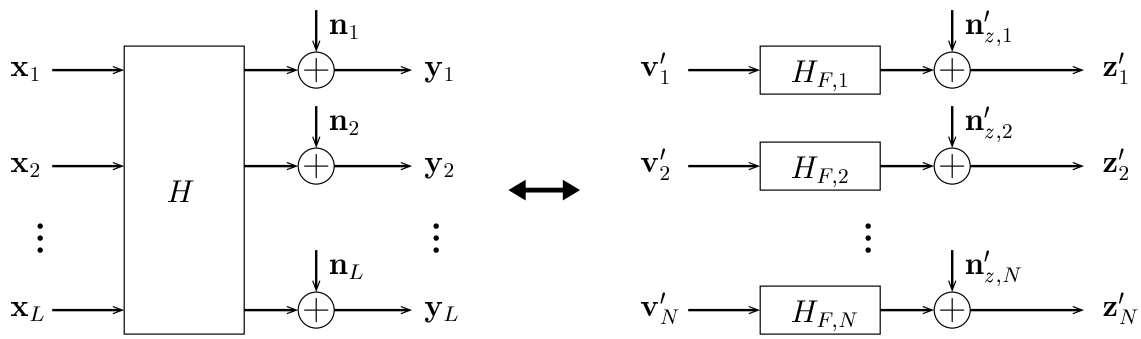

Figure 2.4: Decomposition of the MIMO frequency selective channel into multiple tones.

where

$$\mathbf{z}' = \left[\mathbf{z}_1'^T\, \mathbf{z}_2'^T\, \ldots\, \mathbf{z}_N'^T\right]^T = P\tilde{\mathbf{y}} \ , \tag{2.13}$$

$$\mathbf{v}' = \left[\mathbf{v}_1'^T\, \mathbf{v}_2'^T\, \ldots\, \mathbf{v}_N'^T\right]^T = P\tilde{\mathbf{x}} \ , \tag{2.14}$$

$$\mathbf{n}_z' = \left[\mathbf{n}_{z,1}'^T\, \mathbf{n}_{z,2}'^T\, \ldots\, \mathbf{n}_{z,N}'^T\right]^T = P\tilde{\mathbf{n}} \ , \tag{2.15}$$

$$H_F = P\Lambda P = \operatorname{diag}\left(\overbrace{H_{F,1}, H_{F,2}, \ldots, H_{F,N}}^{N}\right) \ , \tag{2.16}$$

with $R_{\mathbf{n}_z'\mathbf{n}_z'} = I$. Finally, (2.12) becomes

$$\mathbf{z}_i' = H_{F,i}\mathbf{v}_i' + \mathbf{n}_{z,i}' \ , \ i = 1, \ldots, N \ . \tag{2.17}$$

Therefore, $\mathbf{z}_i'$, $\mathbf{v}_i'$, and $\mathbf{n}_{z,i}'$ contain the received samples, transmitted symbols and noise samples of all users corresponding to tone i, and $H_{F,i}$ fully characterizes MIMO transmission within tone i. Thus, the original MIMO frequency selective channel of (2.6) is decomposed into the N MIMO channels of (2.17), as shown in Figure 2.4.

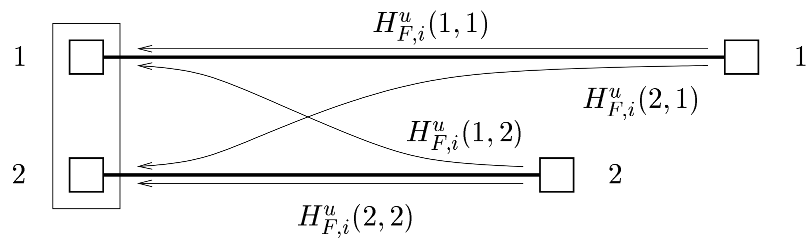

Figure 2.5: Upstream channel with two users.

2.1.4 Channel Properties

The upstream channel is considered first. A distinction between upstream and downstream is made by adopting the notation $H^u_{F,i}$ and $H^d_{F,i}$ respectively. An upstream example channel with only two users is shown in Figure 2.5. Note that each user's transmission reaches the central node through two paths, the "direct" path and the "crosstalk" path. However, the crosstalk coupled received signal originating from a user is always significantly weaker than the directly received signal originating from the same user. If $H(m,n)$ denotes the (m,n) element of matrix H, then for any tone i

$$|H^u_{F,i}(1,1)| \gg |H^u_{F,i}(2,1)| \, , \tag{2.18}$$

$$|H^u_{F,i}(2,2)| \gg |H^u_{F,i}(1,2)| \, . \tag{2.19}$$

Generalizing this observation for more users, the diagonal element of a column of $H^u_{F,i}$ is much larger in magnitude than the off-diagonal elements of the same column, so that for any $n \neq m$

$$|H^u_{F,i}(m,m)| \gg |H^u_{F,i}(n,m)| \, . \tag{2.20}$$

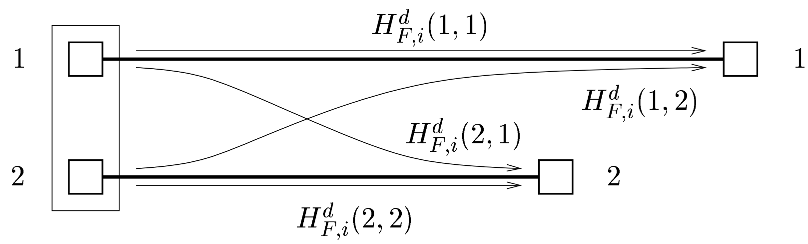

Figure 2.6: Downstream channel with two users.

Alternatively, this property is graphically depicted by

$$H^u_{F,i} = \left[\begin{array}{c|c|c|c} \mathrm{x} & & & \\ & \mathrm{x} & & \\ & & \cdots & \\ & & & \mathrm{x} \end{array}\right] , \tag{2.21}$$

where x indicates the "dominant" element within each column. Such matrices, where the diagonal elements of $H^u_{F,i}$ dominate "column-wise", are described in the following as *column-wise diagonally dominant*.

A similar property holds for the downstream channel as well. From Figure 2.6, depicting an example downstream channel with only two users, one observes that the central node's transmission reaches each user through both a "direct" and a "crosstalk" path. However, the crosstalk coupled signal received by a user is always significantly weaker than the direct signal received by the same user.

Extending this observation to more users, it is found that for any $n \neq m$

$$|H^d_{F,i}(m,m)| \gg |H^d_{F,i}(m,n)| \ . \tag{2.22}$$

A graphic depiction of the property is

$$H^d_{F,i} = \left[\begin{array}{cccc} \mathrm{x} & & & \\ \hline & \mathrm{x} & & \\ \hline & & \vdots & \\ \hline & & & \mathrm{x} \end{array}\right] . \tag{2.23}$$

Matrices of this kind, where the diagonal elements of $H^d_{F,i}$ dominate "row-wise", are described in the following as *row-wise diagonally dominant*.

In DSL, the magnitude difference is typically more than 20 dB. This can be verified from Figure 2.7, which refers to an extreme case, where a 500 foot pair neighbors a 6000 foot pair (and both consist of AWG-26 copper wire). The insertion losses are shown together with the FEXT coupling losses, assuming upstream transmission. The data were obtained using the standard models adopted by ANSI (see [18]). Clearly, the insertion loss of a signal is always smaller than the coupling loss that it experiences when it propagates into a neighboring pair. It must be stressed that this condition is not equivalent to saying that the crosstalk signals are weak. Notice that, in the example of Figure 2.7, the crosstalk coupling into loop 2 is much stronger than the transfer function of loop 2 at all frequencies above 1.7MHz.

From Figures 2.5 and 2.6, notice that the reciprocity property for twisted pair transmission implies that

$$H^u_{F,i}(2,1) = H^d_{F,i}(1,2) \ , \tag{2.24}$$

$$H^u_{F,i}(1,2) = H^d_{F,i}(2,1) \ . \tag{2.25}$$

Extending this observation to more users, one concludes that for all i

$$H^u_{F,i} = \left(H^d_{F,i}\right)^T \ . \tag{2.26}$$

However, it should be mentioned that this property does not hold in practice due to the existence of analog receiver and transmitter filters that differ between the central node and the remote terminals. This is why the upstream and downstream cases are

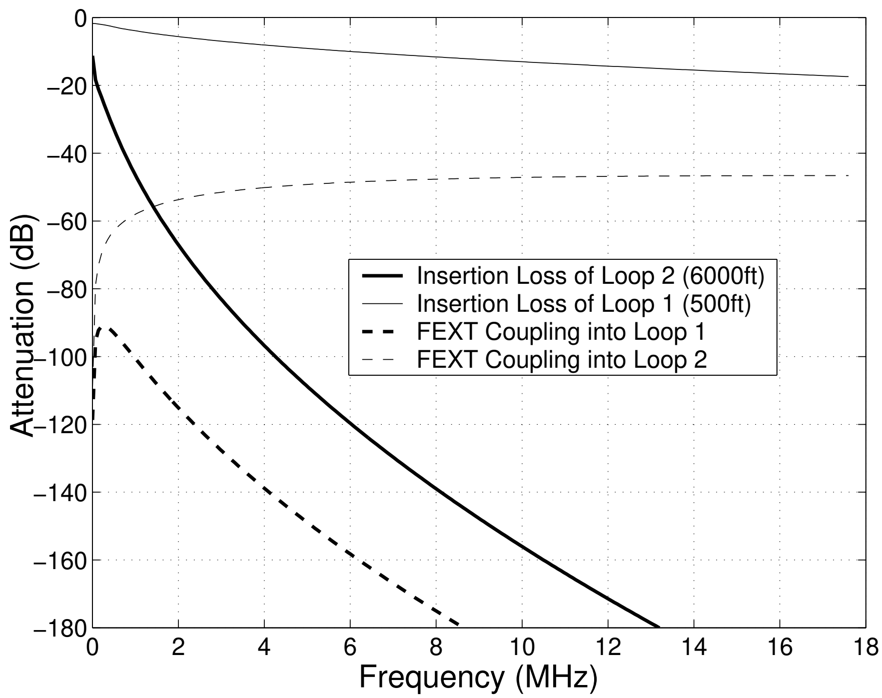

Figure 2.7: Example of insertion loss and FEXT coupling loss for upstream DSL.

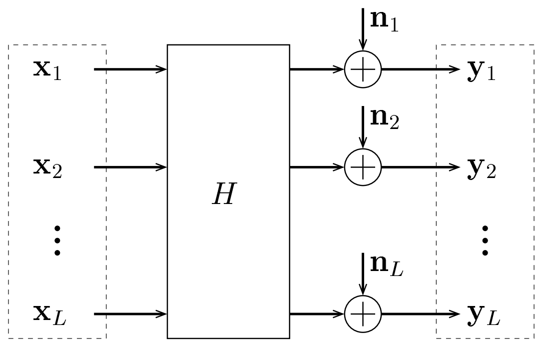

Figure 2.8: Gaussian vector channel.

treated separately in the following.

2.2 Multiuser Information Theory

In the previous section, multi-line channels were characterized and their distinct properties were revealed. This section investigates the multiuser aspects of such channels and reviews some relevant information theory results. These results are utilized in Chapter 5, where the performance of the proposed techniques is compared against bounds defined in the next subsections.

2.2.1 Vector Channel

Consider the Gaussian vector channel defined by (2.6) and depicted in Figure 2.8. For now, assume that the channel is used by a single user, who is capable of accessing all inputs and all outputs. In such a case, it is possible to have *coordination* among the inputs and among the outputs of the channel. Coordination refers to the ability to process the signals of all users jointly, such as performing linear pre-processing of the input signals, or joint minimum-distance decoding of the outputs.

A single-user channel is characterized by its *capacity*, which is defined as the highest rate in bits per channel use at which information can be sent with arbitrarily low probability of error [19]. Typical communication systems face constraints on their average and on their peak transmitted power. The focus of this section is on the average power constraint. In the context of multiple inputs, this constraint translates into a limit on the trace of the channel input covariance matrix. The single-user capacity of the Gaussian vector channel under an average power constraint is known to be

$$\begin{aligned} C^{SU} &= \max_{R_{\mathbf{xx}}} \frac{1}{2} \log |H R_{\mathbf{xx}} H^* + I| \\ &\quad \text{subject to} \quad \text{trace}\,(R_{\mathbf{xx}}) \leq \mathcal{E} \ , \end{aligned} \tag{2.27}$$

where $R_{\mathbf{xx}}$ is the input covariance matrix, and $\mathcal{E}$ is the maximum total transmitted energy. The optimum $R_{\mathbf{xx}}$ is such that its eigenvectors are equal to the right singular vectors of H, and its eigenvalues are obtained by water-filling across the singular values of H [20, 21, 22, 23].

This capacity result assumes coordination both for the channel inputs and for the channel outputs. Clearly, this is not the case for the multi-line channels, where multiple users exist. Since the multiple users are geographically dispersed, the ability to process the signals of all users jointly is limited either to the output signals (for upstream transmission), or to the input signals (for downstream transmission). Still, the next subsections show that (2.27) serves as a useful upper bound on the achievable performance of multi-line channels.

It must be noted that the inclusion of a cyclic prefix (as in Subsection 2.1.3) induces a loss with respect to the capacity of (2.27), as well as with respect to the capacity expressions of the next sections.

2.2.2 Upstream Multiuser Channel

Examining Figure 2.1 for the case of *upstream* communication, one observes that multiple users transmit information simultaneously to a common receiver. Such a channel is described as a Multiple Access Channel (MAC). Since the inputs and outputs are vector quantities, and in order to differentiate from the case of scalar inputs and outputs, this channel is specifically characterized as a vector MAC.

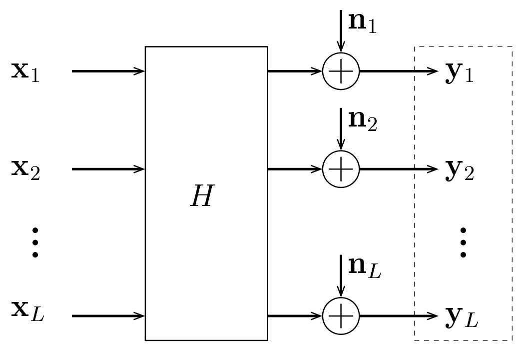

Figure 2.9: Gaussian vector multiple access channel.

In multiuser communication, the concept of capacity is replaced by the concept of the *capacity region*, which is defined as the set of achievable user rate tuples. A rate tuple is said to be achievable if users can transmit information at the prescribed rates with arbitrarily low probability of error [19].

The capacity region of the scalar MAC was found in [24]. The capacity region of the Gaussian vector MAC (shown in Figure 2.9) under power constraints for each user was derived in [25]. This region is expressed in [26] as

$$\begin{aligned}\mathcal{C}^{MAC} \quad = \quad & \bigcup_{R_{\mathbf{x}_1\mathbf{x}_1},\ldots,R_{\mathbf{x}_L\mathbf{x}_L}} \left[\begin{array}{l} (r_1,\ldots,r_L): \\ \sum_{i\in S} r_i \leq \frac{1}{2}\log\left|\sum_{i\in S} H_i R_{\mathbf{x}_i\mathbf{x}_i} H_i^* + I\right|, \forall S \subseteq \{1,\ldots,L\} \end{array} \right] \\ & \text{subject to} \quad \operatorname{trace}\left(R_{\mathbf{x}_i\mathbf{x}_i}\right) \leq \mathcal{E}_i \ , \ i = 1,\ldots,L \ , \end{aligned} \tag{2.28}$$

where $R_{\mathbf{x}_i\mathbf{x}_i}$ is the covariance matrix of $\mathbf{x}_i$, $H_i = \left[H_{1,i}^T \ H_{2,i}^T \ \ldots \ H_{L,i}^T\right]^T$, and $\mathcal{E}_i$ is the total energy allowed for user i. An illustration of the capacity region for 2 users is given in Figure 2.10, where it can be seen that the region is defined as the union of pentagons (shown with dashed lines) corresponding to all possible input covariance matrices.

The computation of the capacity region has been shown to belong to the class of convex programming problems [26], which allows for efficient algorithms to be employed. However, upper bounds on the region can be obtained with a much lower

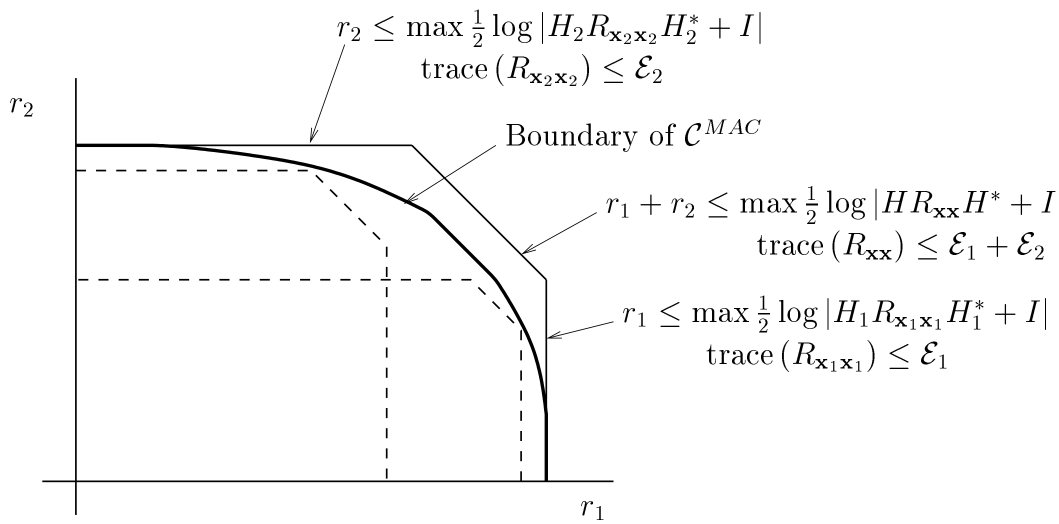

Figure 2.10: Capacity region of the Gaussian vector multiple access channel with 2 users.

computational cost. It is seen in later chapters that, for the specific case of multi-line channels, these upper bounds are tight.

Assuming that the entire channel is utilized by a single user, the capacity of this user can be derived. This capacity provides an upper bound on the rate of the specified user for the multiuser case. Also, by supposing that coordination is possible among the transmitters, the single user capacity of the channel can be obtained through (2.27). This value provides an upper bound on the sum capacity of all users in the multiuser case. Thus, the following region is a superset of $\mathcal{C}^{MAC}$,

$$\mathcal{C}^{MAC-UB} = \begin{bmatrix} (r_1, \ldots, r_L) : \\ r_i \leq \max_{\text{trace}\left(R_{\mathbf{x}_i\mathbf{x}_i}\right) \leq \mathcal{E}_i} \frac{1}{2} \log |H_i R_{\mathbf{x}_i\mathbf{x}_i} H_i^* + I| \,, i = 1, \ldots, L \\ \sum_{i=1}^{L} r_i \leq \max_{\text{trace}(R_{\mathbf{xx}}) \leq \sum_{i=1}^{L} \mathcal{E}_i} \frac{1}{2} \log |H R_{\mathbf{xx}} H^* + I| \end{bmatrix} . \tag{2.29}$$

The bounds are illustrated in Figure 2.10 for the 2 user case.

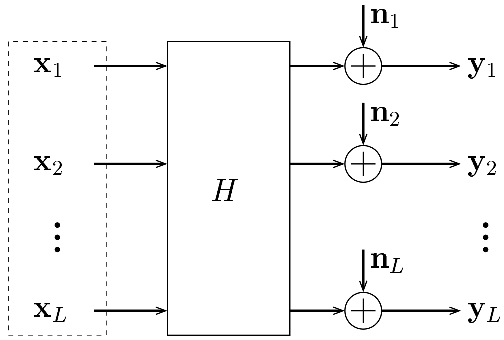

Figure 2.11: Gaussian vector broadcast channel.

2.2.3 Downstream Multiuser Channel

In the case of *downstream* communication, a common transmitter needs to send information to the receivers of multiple users (see Figure 2.1). This channel is described as a Broadcast Channel (BC), and for those instances where the inputs and outputs are vectors, it is described as a vector BC.

For the vector BC, coordination is possible among the transmitters, in which case the transmitter output is characterized by a total energy constraint. The capacity region remains unsolved except in the special case of the *degraded* BC [27]. Still, a number of results exist with regard to the sum capacity of the Gaussian vector BC (see Figure 2.11).

Reference [28] proposes (independently of [29, 30]) the use of the QR factorization in order to form a set of interference channels, where the interference is non-causally known to the transmitters. Then, *lattice precoding* [31] is adopted to achieve complete removal of the interference. This scheme is named *Ranked Known Interference* and is proven to be asymptotically sum capacity optimal both for low and high SNR. In [32], the solution for the sum capacity is expressed as the result of a game, with a "signal player" choosing a transmission covariance matrix to maximize the sum rate, and a "noise player" choosing a noise covariance matrix to minimize the sum rate.

Although the capacity region $\mathcal{C}^{BC}$ is unknown, upper bounds can be obtained

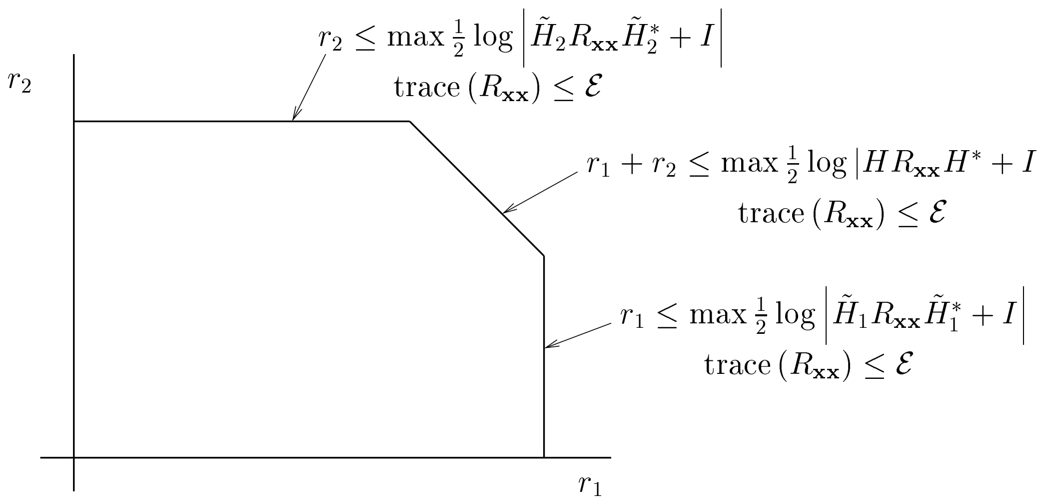

Figure 2.12: Capacity region bounds of the Gaussian vector broadcast channel with 2 users.

with a low computational cost. As shown in later chapters, for multi-line channels these upper bounds are tight. Similarly to upstream transmission, these bounds are obtained by assuming either that there is a single user receiving, or that coordination is possible at both sides. Thus, the following region is a superset of $\mathcal{C}^{BC}$,

$$\mathcal{C}^{BC-UB} = \left[\begin{array}{l} (r_1, \ldots, r_L) : \\ r_i \leq \max_{\text{trace}(R_{\mathbf{xx}}) \leq \mathcal{E}} \frac{1}{2} \log \left| \tilde{H}_i R_{\mathbf{xx}} \tilde{H}_i^* + I \right| \ , \ i = 1, \ldots, L \\ \sum_{i=1}^{L} r_i \leq \max_{\text{trace}(R_{\mathbf{xx}}) \leq \mathcal{E}} \frac{1}{2} \log |H R_{\mathbf{xx}} H^* + I| \end{array} \right] , \quad (2.30)$$

where $\mathcal{E}$ is the total transmitter output energy, and $\tilde{H}_i = [H_{i,1}\ H_{i,2}\ \ldots\ H_{i,L}]$. An illustration for 2 users is given in Figure 2.12.

Chapter 3

Generalized Decision Feedback Equalization

Decision feedback equalization (DFE) is a widely studied and very popular technique for Inter-Symbol Interference (ISI) suppression in channels with memory. In recent years, the concept of decision feedback has been employed in different contexts. Decision feedback equalizers for Multiple-Input-Multiple-Output (MIMO) channels have been proposed in [33, 34, 35]. Multiuser detectors employing decision feedback are described in [36, 37, 38], primarily aimed at improving Code-Division-Multiple-Access (CDMA) systems. The Bell-Labs Space-Time (BLAST) architecture utilizes decision feedback in wireless communication with multiple antennas [39, 40]. Also, equalization with decision feedback for block transmission in ISI channels has received considerable attention [41, 42].

The Generalized Decision Feedback Equalizer (GDFE) was originally proposed in [42] in the context of block transmission, but it was quickly realized that it is applicable in MIMO and multiuser channels. The GDFE is in several respects more general than other proposed structures. Additionally, its special cases include a number of known transmission techniques. In the next chapter, the GDFE serves as a starting point for the crosstalk cancellation techniques of multi-line channels.

Section 3.1 rederives the GDFE structure both with a Minimum-Mean-Square-Error (MMSE) and with a Zero-Forcing (ZF) criterion, using an approach that is different from that of previous proofs. To further highlight the generality of the GDFE, Section 3.2 first reviews some known special cases of the GDFE. Additionally, the relationships of the GDFE with the MMSE-DFE for ISI channels, and with the Vertical BLAST (V-BLAST) architecture are revealed. These last results are new developments and their complete proofs are given in Appendices A and B.

3.1 GDFE Derivation

3.1.1 Channel

The channel is described by

$$\mathbf{y} = H\mathbf{x} + \mathbf{n} \ , \tag{3.1}$$

where H defines an $N \times M$ channel matrix, $\mathbf{y}$ is the $N \times 1$ vector of received samples, $\mathbf{x}$ is the $M \times 1$ vector of transmitted symbols and $\mathbf{n}$ is the $N \times 1$ vector of noise samples. It is assumed that $R_{\mathbf{nn}} = I_N$. (Otherwise, a noise whitening matrix filter $R_{\mathbf{nn}}^{-1/2}$ must be included at the receiver.)

Several communication systems include transmit filtering, thus the channel input is expressed as

$$\mathbf{x} = F\mathbf{v} \ , \tag{3.2}$$

where $\mathbf{v}$ is the $N_v \times 1$ vector of input symbols, $R_{\mathbf{vv}}$ is positive definite and diagonal, and F represents the transmit filter. The assumption here is that $N_v \leq M$, which implies that F may expand the dimensionality of the transmitted symbols. The purpose of the transmit filter typically relates to optimizing the channel input, or to simplifying the receiver structure. Actually, the choice of F greatly influences the receiver structure. As explained later, an appropriate choice of F can completely eliminate the need for the equalizer to have feedback at the receiver.

3.1.2 MMSE-GDFE

The following derivation of the GDFE adopts a different approach than that of [42]. The GDFE equalization structure is shown in Figure 3.1.

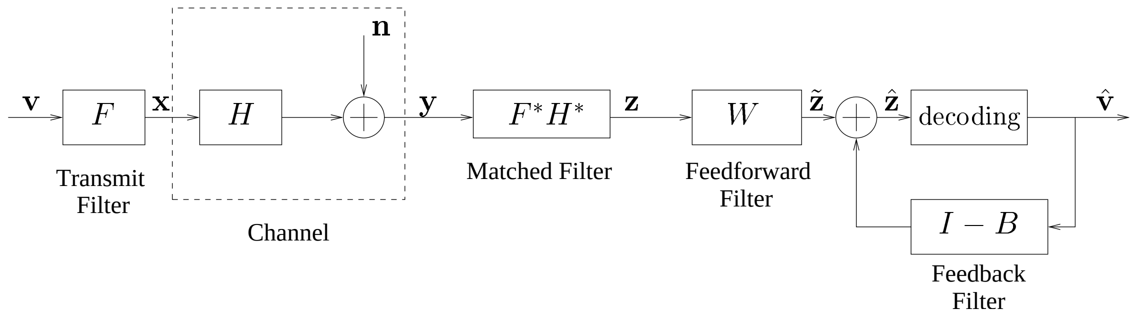

Figure 3.1: GDFE equalization structure.

Taking into account filtering at the transmitter and matched filtering at the receiver, the model

$$\mathbf{z} = R_f \mathbf{v} + \mathbf{n}_z \tag{3.3}$$

is obtained, where $\mathbf{z}$ is the output of the matrix matched filter, $R_f = F^*H^*HF$, $\mathbf{n}_z = F^*H^*\mathbf{n}$ and $R_{\mathbf{n}_z\mathbf{n}_z} = R_f$. As seen in Figure 3.1, $\mathbf{z}$ is processed by the feedforward filter W, and the feedback filter is $I - B$, where B is upper triangular and monic (diagonal elements equal 1). Assuming correct decisions, the error at the input of the decision device is

$$\mathbf{e} = \mathbf{v} - W\mathbf{z} - (I - B)\mathbf{v} = B\mathbf{v} - W\mathbf{z} \ . \tag{3.4}$$

The orthogonality principle implies that

$$BR_{\mathbf{vz}} = WR_{\mathbf{zz}} \Rightarrow W = BR_b \ , \tag{3.5}$$

where

$$R_b = R_{\mathbf{vz}}R_{\mathbf{zz}}^{-1} \Rightarrow R_b^{-1} = R_f + R_{\mathbf{vv}}^{-1} \ . \tag{3.6}$$

Note that by assumption $R_{\mathbf{vv}}$ is non-singular, therefore, so is R_b. The error covariance matrix is

$$\begin{aligned} R_{\mathbf{ee}} &= BR_{\mathbf{vv}}B^* - BR_bR_{\mathbf{zz}}R_bB^* \\ &= B(R_{\mathbf{vv}} - R_bR_{\mathbf{zz}}R_b)B^* \\ &= BR_bB^* \ . \end{aligned} \tag{3.7}$$

The criterion used for determining B is the minimization of the trace of $R_{\mathbf{ee}}$. This corresponds to the minimization of the sum of the mean square errors at the input of the symbol-by-symbol detection device.

Proposition 3.1.1 *Suppose that $R_{\mathbf{ee}} = BR_bB^*$, as in (3.7). Let the Cholesky decomposition of R_b^{-1} be*

$$R_b^{-1} = G^*S_oG \ , \tag{3.8}$$

where S_o is diagonal with positive elements, and G is upper triangular and monic. Then,

$$\mathit{trace}\,(R_{\mathbf{ee}}) \geq \sum_{i=1}^{N_v} s_i^{-1} \ , \tag{3.9}$$

where s_i is the i-th diagonal element of S_o. Equality is achieved by choosing $B = G$.

Proof: Observe that

$$\begin{aligned} \text{trace}\,(R_{\mathbf{ee}}) &= \text{trace}\left(BG^{-1}S_o^{-1}G^{-*}B^*\right) & (3.10)\\ &= \text{trace}\left[S_o^{-1} + \left(BG^{-1} - I\right)S_o^{-1}\left(BG^{-1} - I\right)^*\right] & (3.11)\\ &\geq \text{trace}\left(S_o^{-1}\right) \ , & (3.12) \end{aligned}$$

where (3.10) results from substitution using (3.7) and (3.8), (3.11) is due to

$$\text{trace}\left(S_o^{-1}\right) = \text{trace}\left(BG^{-1}S_o^{-1}\right) = \text{trace}\left(S_o^{-1}G^{-*}B^*\right) \ , \tag{3.13}$$

and (3.12) is a consequence of the fact that $(BG^{-1} - I)\,S_o^{-1}\,(BG^{-1} - I)^*$ is non-negative definite. Obviously, $B = G$ achieves equality. [1] ■

[1] This proof borrows ideas from the proof of Lemma 1 in [34].

An alternative interpretation for the choice of $B = G$ is that it achieves the diagonalization of $R_{\mathbf{ee}}$, in other words $\mathbf{e}$ is "whitened".

The optimal choice of B results in the following feedforward processing,

$$\tilde{\mathbf{z}} = S_o^{-1} G^{-*} \mathbf{z} = G\mathbf{v} - \mathbf{e} \ , \tag{3.14}$$

where $R_{\mathbf{ee}} = S_o^{-1}$. The upper triangular structure of G and the fact that $R_{\mathbf{ee}}$ is diagonal can be exploited to recover $\mathbf{v}$ with back-substitution and symbol-by-symbol detection as in Figure 3.1. Equation 3.14 is written as

$$\tilde{z}_i = v_i + g_{i,i+1} v_{i+1} + \ldots + g_{i,N_u} v_{N_v} - e_i \ , \ i = 1, \ldots, N_v \ , \tag{3.15}$$

where $\tilde{z}_i$, v_i, and e_i are the elements of $\tilde{\mathbf{z}}$, $\mathbf{v}$, and $\mathbf{e}$ correspondingly. Also, $g_{i,j}$ is the (i, j) element of G. If decision errors are assumed to be very rare, then feedback processing yields

$$\hat{z}_i = v_i - e_i \ , \ i = 1, \ldots, N_v \ , \tag{3.16}$$

where $\hat{z}_i$ are the elements of the decoder input, $\hat{\mathbf{z}}$, shown in Figure 3.1.

Therefore, a set of parallel channels is formed. The corresponding detection SNR's are

$$SNR_i = \mathcal{E}_i s_i \ , \ i = 1, \ldots, N_v \ , \tag{3.17}$$

where $\mathcal{E}_i$ is the i-th diagonal element of $R_{\mathbf{vv}}$.

3.1.3 ZF-GDFE

Assuming that R_f of (3.3) is non-singular, a ZF criterion can be employed. The following Cholesky factorization is obtained,

$$R_f = G^* S_o G \ . \tag{3.18}$$

Note that the G and S_o obtained here differ from those of (3.8). Processing the received vector $\mathbf{z}$ by $S_o^{-1}G^{-*}$ yields

$$S_o^{-1}G^{-*}\mathbf{z} = G\mathbf{v} + S_o^{-1}G^{-*}\mathbf{n}_z = G\mathbf{v} + \tilde{\mathbf{n}}_z \ , \tag{3.19}$$

where $R_{\tilde{\mathbf{n}}_z\tilde{\mathbf{n}}_z} = S_o^{-1}$ is diagonal. Since G is upper triangular and $\tilde{\mathbf{n}}_z$ has uncorrelated components, the input $\mathbf{v}$ can be recovered by back-substitution combined with symbol-by-symbol detection. Thus, the feedforward matrix is $W = S_o^{-1}G^{-*}$ and the feedback matrix is $I - B = I - G$.

3.2 Special Cases

The GDFE structure described in the previous section serves as the foundation of the crosstalk cancellation techniques that are proposed in Chapter 4. At this point, it is useful to outline the connections of the GDFE to other known transmission schemes. Additionally, new results are presented, such as the asymptotic equivalence between the MMSE-DFE of an ISI channel and the GDFE, and the relation between V-BLAST and the GDFE.

3.2.1 Channels with ISI

The GDFE structure effectively decomposes the matrix channel into a set of independent one-dimensional sub-channels. This decomposition is not unique, and different decompositions lead to distinct transmission systems.

The GDFE is applicable to channels with ISI, when block transmission and reception is employed. In such a case, the matrix H in (3.1) is an $N \times (N+\nu)$ convolution matrix corresponding to a channel with $\nu + 1$ taps as in (2.2). Evidently, the dimensionality of the output is smaller than the dimensionality of the input, which implies that some of the output samples (constituting the so-called "guard band") are ignored by the receiver.

In the vector coding method [20, 21, 22], the singular value decomposition of the channel matrix H is obtained,

$$H = U\Sigma V^* \ , \tag{3.20}$$

where U, V are unitary and Σ is diagonal with non-negative elements. The channel input vectors are chosen as the input singular vectors of the matrix and the channel output is appropriately demodulated by obtaining its inner products with the output singular vectors. Reference [42] proves that vector coding is a special case of the GDFE, where

$$\begin{aligned} F &= V \ , & (3.21)\\ S_o^{-1}G^{-*}F^*H^* &= \left(\Sigma^2 + R_{\mathbf{vv}}^{-1}\right)^{-1}\Sigma U^* \ , & (3.22)\\ I - G &= O \ . & (3.23) \end{aligned}$$

Thus, the feedback component is eliminated, and the feedforward section amounts to *rotation* by U^* and scaling as shown in Figure 3.2. The channel decomposition produces sub-channels with gains equal to the singular values of H.

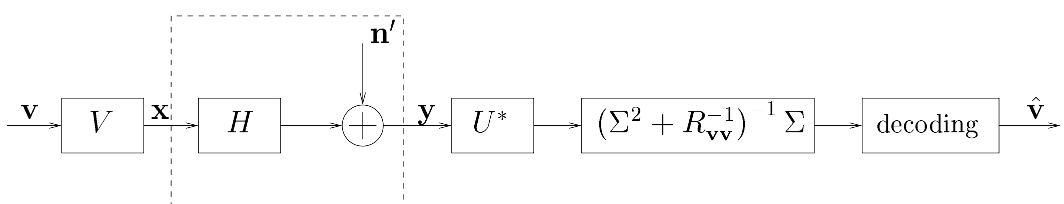

Figure 3.2: Vector coding equalization structure.

When block transmission assumes a cyclic prefix (CP) of length ν, then H becomes a *circulant* matrix as in (2.5). For any circulant matrix, the input singular vectors are equal to the columns of a matrix defining an Inverse Discrete Fourier Transform (IDFT) operation, and the output singular vectors are equal to the columns of a matrix defining a Discrete Fourier Transform (DFT) operation, as in

$$H = Q_{IDFT} D Q_{DFT} \ , \tag{3.24}$$

where D is diagonal with elements equal to the DFT values of the first row of H. The choice of

$$F = Q_{IDFT} , \tag{3.25}$$
$$S_o^{-1}G^{-*}F^*H^* = \left(D^*D + R_{\mathbf{vv}}^{-1}\right)^{-1} D^*Q_{DFT} , \tag{3.26}$$
$$I - G = O \tag{3.27}$$

results in the well-known Discrete Multi-Tone (DMT), or Orthogonal Frequency Division Multiplexing (OFDM) methods [15, 17, 43]. As shown in Figure 3.3, the transmitter performs an IDFT operation, and the receiver performs a DFT operation followed by scaling. Thus, DMT/OFDM can also be seen as a special case of the GDFE (see also [42]).

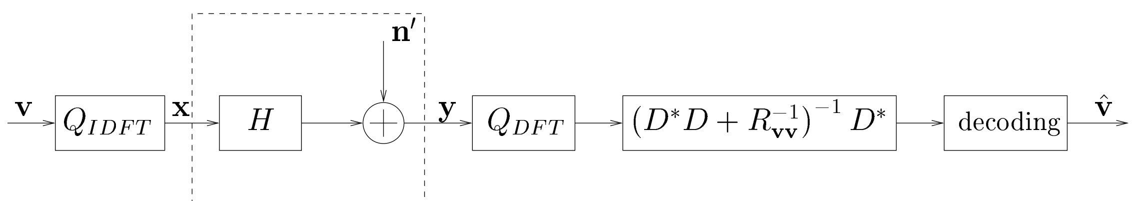

Figure 3.3: DMT/OFDM equalization structure.

Under certain conditions the GDFE converges to the MMSE-DFE equalizer (or to a set of MMSE-DFE equalizers) of an ISI channel [44] in the narrow sense that both structures perform the same signal operations at both the transmitter and the receiver. The convergence results are obtained in the limiting case where the number of dimensions of the GDFE becomes very large, and they apply to both the receiver equalization, and the transmitter optimization filtering. The derivation of these results is quite extensive and outside the main focus of this thesis, thus, it is included in Appendix A (see also [45]). The proofs are based on a fundamental equivalence theorem holding between the Cholesky factorization of a doubly-infinite Toeplitz covariance matrix and the spectral factorization of a doubly infinite stationary sequence.

3.2.2 MIMO Channels

MIMO channels represent a model for communication systems with multiple antennas, or with multiple lines. If coordination among all the inputs and among all the outputs of the channels can be performed, the vector coding principle of [22] is directly applicable. In cases where the MIMO channel has memory, vector coding may be combined with DMT, yielding Discrete Matrix Multi-Tone (DMMT) of [46] as a technique for space-time equalization. DMMT can also be shown to be a special case of the GDFE.

The V-BLAST architecture [40, 47] has been proposed for realizing very high spectral efficiencies over wireless channels with multiple antennas, and with independent Rayleigh fading between antenna pairs. The principle of successive interference cancellation is utilized, whereas, prior to detection of a symbol, interference from previously detected symbols is subtracted, and the received vector elements are linearly weighted to null (or reduce) interference from the yet undetected symbols. Appendix B proves that the receiver processing steps of this algorithm are actually equivalent to the receiver operations of the GDFE (see also [48]). It should be noted that this equivalence also applies for Diagonal BLAST (D-BLAST) [39], in which case an appropriate permutation matrix must be included in the GDFE transmitter.

Last, it has been observed that some instances of space-time codes may also be considered as special cases of the GDFE. Reference [49] shows the relationship between Orthogonal Space-Time Block Codes [50] and block codes capable of diagonalizing any arbitrary channel matrix with no channel knowledge at the transmitter. The latter codes can be realized by appropriate linear processing at both the transmitter and the receiver, and they may be viewed as a GDFE where both the feedforward and the feedback section have been made redundant.

Chapter 4

Coordinated Communication via QR Factorization

Crosstalk is a dominant constraint in the design of communication systems for multi-line channels. The need to mitigate the adverse effects of crosstalk on Digital Subscriber Line (DSL) systems has motivated significant research work. References [51, 52] exploited the cyclostationary character of crosstalk in the design of improved Viterbi decoders and decision feedback equalizers for DSL. In the absence of DSL receiver coordination, [53, 54] proposed wider-than-Nyquist system designs offering performance improvements. References [55, 56] recognized that the local subscriber loop is a Multiple-Input-Multiple-Output (MIMO) channel, and suggested joint linear processing of the signals at the transmitters or the receivers for the suppression of crosstalk.

This chapter studies multi-carrier communication in a multi-line channel, where coordination is possible at the central node only. The derivations have the Zero-Forcing (ZF) Generalized Decision Feedback Equalizer (GDFE) of Chapter 3 as a starting point. Section 4.1 presents the system designs for upstream and downstream communication, where the resulting crosstalk-cancelling structures involve non-linear processing, such as decision feedback and modulo precoding. The upstream scheme may be viewed as an application of successive interference cancellation for multi-line channels. The downstream scheme is a novel design, which has recently been independently proposed in [28] in the context of cellular wireless systems.

Section 4.2 takes the properties of multi-line channels into account and further examines the proposed systems. The following novel results are obtained: It is shown that the system performance approximates that of an equivalent channel with no crosstalk. Additionally, the system performance that is achieved with the ZF criterion is approximately equal to that with the Minimum-Mean-Square-Error (MMSE) criterion. The practical effects of channel estimation errors, computational complexity and error propagation are also addressed in this section. (See also [29, 30, 57].)

4.1 Crosstalk Cancellation

4.1.1 Channel

Recall the channel model of (3.1),

$$\mathbf{y} = H\mathbf{x} + \mathbf{n} \ . \tag{4.1}$$

Assuming that the matrix H corresponds to a multi-line channel as in Section 2.1, $\mathbf{y} = \left[\mathbf{y}_1^T \ \mathbf{y}_2^T \ \ldots \ \mathbf{y}_L^T\right]^T$, $\mathbf{x} = \left[\mathbf{x}_1^T \ \mathbf{x}_2^T \ \ldots \ \mathbf{x}_L^T\right]^T$, $\mathbf{n} = \left[\mathbf{n}_1^T \ \mathbf{n}_2^T \ \ldots \ \mathbf{n}_L^T\right]^T$, and $\mathbf{y}_i$, $\mathbf{x}_i$, $\mathbf{n}_i$ are blocks of samples corresponding to user i. Following the assumptions of Subsection 2.1.3, H consists of circulant block matrices $H_{i,j}$, and $R_{\mathbf{nn}} = I$. Also, H is considered non-singular, a fact that is further discussed later.

In the next two subsections, signal processing techniques are described for upstream and downstream transmission achieving both equalization against ISI and crosstalk cancellation. These techniques combine DMT modulation with either successive interference cancellation for upstream, or modulo precoding for downstream.

4.1.2 Upstream

For upstream transmission, it is possible to perform joint signal processing of the received signals, but the transmitters of the remote users must act independently.

The operation of the transmitters is defined by (3.2),

$$\mathbf{x} = F\mathbf{v} \ . \tag{4.2}$$

Since a multicarrier solution is sought, each transmitter performs an IDFT operation. The elements of $\mathbf{v}$ are interpreted as user symbols transmitted on frequency tones, and $\mathbf{v} = \left[\mathbf{v}_1^T \ \mathbf{v}_2^T \ \ldots \ \mathbf{v}_N^T\right]^T$, where $\mathbf{v}_i = \left[v_{i,1} \ v_{i,2} \ \ldots \ v_{i,L}\right]^T$ is a block of symbols of the L users corresponding to tone i.

Therefore, the transmit filter of the GDFE structure is now defined as

$$F = Q_{mIDFT}P \ , \tag{4.3}$$

where Q_{mIDFT}, P were defined in (2.11) and Subsection 2.1.3 correspondingly. The function of F can now be interpreted: Originally, $\mathbf{v}$ consists of N groups of L components, where each group holds data symbols of the L users. The matrix P performs a reordering of those symbols, so that each of the L resulting groups contains the N data symbols corresponding to one of the L users. Then, Q_{mIDFT} computes the IDFT on the symbols of each user. Clearly, the operation of the transmit filter does not result in any energy increase of the transmitted signal. Also, the data symbols of each user are processed independently of those of other users, which is consistent with the assumption that the transmitters are geographically separated.

The remainder of the ZF-GDFE modules are derived next. Recalling that the $H_{i,j}$ blocks of matrix H are circulant, $H_{i,j} = Q_{IDFT}\Lambda_{i,j}Q_{DFT}$, where $\Lambda_{i,j}$ is diagonal. Thus, R_f of (3.3) becomes

$$\begin{aligned} R_f &= F^*H^*HF && (4.4)\\ &= P\Lambda^*\Lambda P \ , && (4.5) \end{aligned}$$

where Λ consists of the $\Lambda_{i,j}$ blocks. The reordering imposed by the P matrix makes R_f a block diagonal matrix

$$R_f = \operatorname{diag}\left(R_{f,1}, R_{f,2}, \ldots, R_{f,N}\right) \ , \tag{4.6}$$

where $R_{f,i}$, $i = 1, \ldots, N$ are $L \times L$ matrices. The interpretation of this structure is that each block describes transmission within a "tone". The Cholesky factorization of R_f is

$$R_f = \mathrm{diag}\left(G_1^* S_{o,1} G_1, G_2^* S_{o,2} G_2, \ldots, G_N^* S_{o,N} G_N\right) \ , \tag{4.7}$$

where G_i are upper triangular and monic, and $S_{o,i}$ are diagonal with positive elements. It offers great aid to define the following QR factorization

$$H_F = P\Lambda P = \mathrm{diag}\left(H_{F,1}, H_{F,2}, \ldots, H_{F,N}\right) = \mathrm{diag}\left(Q_1 R_1, Q_2 R_2, \ldots, Q_N R_N\right) \ , \tag{4.8}$$

where $H_{F,i}$ characterizes MIMO transmission within tone i. Also, Q_i, $i = 1, \ldots, N$ are $L \times L$ unitary matrices, and R_i, $i = 1, \ldots, N$ are $L \times L$ upper triangular matrices. Simple substitution proves that $R_i = S_{o,i}^{\frac{1}{2}} G_i$, so the elements of $S_{o,i}^{\frac{1}{2}}$ are equal to the magnitude of the diagonal elements of R_i.

The GDFE receiver components are now examined. The special form of the feedback filter $G = \mathrm{diag}\left(G_1, G_2, \ldots, G_N\right)$ allows its breakup into N independent feedback filters, where each filter operates on the signals of a single tone. The combination of the matched filter and the feedforward filter is

$$\begin{aligned} S_o^{-1} G^{-*} F^* H^* &= S_o^{-1} G^{-*} P \Lambda^* Q_{mDFT} \\ &= S_o^{-1} G^{-*} P \Lambda^* P P Q_{mDFT} \ . \end{aligned} \tag{4.9}$$

Using (4.8)

$$\begin{aligned} &S_o^{-1} G^{-*} F^* H^* = \\ &S_o^{-1} G^{-*} \mathrm{diag}\left(R_1^* Q_1^*, R_2^* Q_2^*, \ldots, R_N^* Q_N^*\right) P Q_{mDFT} = \\ &S_o^{-1} G^{-*} \mathrm{diag}\left(G_1^* S_{o,1}^{\frac{1}{2}}, G_2^* S_{o,2}^{\frac{1}{2}}, \ldots, G_N^* S_{o,N}^{\frac{1}{2}}\right) \mathrm{diag}\left(Q_1^*, Q_2^*, \ldots Q_N^*\right) P Q_{mDFT} = \\ &\mathrm{diag}\left(S_{o,1}^{-\frac{1}{2}}, S_{o,2}^{-\frac{1}{2}}, \ldots, S_{o,N}^{-\frac{1}{2}}\right) \mathrm{diag}\left(Q_1^*, Q_2^*, \ldots Q_N^*\right) P Q_{mDFT} \ . \end{aligned} \tag{4.10}$$

At this point, the function of the "matched-feedforward" filter can be interpreted: The matrix Q_{mDFT} performs the DFT for each user, and matrix P reorders the samples into groups of tones. Each Q_i^* matrix "rotates" the samples corresponding to the i-th tone, in order to optimize the effect of decision feedback occuring later, and each $S_{o,i}^{-1/2}$ matrix performs a simple scaling. The operations taking place after reordering can be broken into N independent sections corresponding to tones, just like for the feedback filter.

Having determined the system modules, the signal processing operations are next outlined. The output of the DFT after re-ordering is

$$\mathbf{z}' = PQ_{mDFT}\mathbf{y} = P\Lambda P\mathbf{v} + \mathbf{n}'_z \Leftrightarrow \tag{4.11}$$

$$\mathbf{z}'_i = H_{F,i}\mathbf{v}_i + \mathbf{n}'_{z,i} \ , \ i = 1, \ldots, N \ , \tag{4.12}$$

where $\mathbf{z}' = \left[\mathbf{z}_1'^T \ \mathbf{z}_2'^T \ \ldots \ \mathbf{z}_N'^T\right]^T$, $\mathbf{n}'_z = PQ_{mDFT}\mathbf{n} = \left[\mathbf{n}_{z,1}'^T \ \mathbf{n}_{z,2}'^T \ \ldots \ \mathbf{n}_{z,N}'^T\right]^T$, and $R_{\mathbf{n}'_{z,i}\mathbf{n}'_{z,i}} = I_L$. The feedforward section of the receiver finally produces:

$$\begin{aligned} \tilde{\mathbf{z}}_i &= S_{o,i}^{-\frac{1}{2}} Q_i^* \mathbf{z}'_i & (4.13)\\ &= S_{o,i}^{-\frac{1}{2}} R_i \mathbf{v}_i + S_{o,i}^{-\frac{1}{2}} Q_i^* \mathbf{n}'_{z,i} & (4.14)\\ &= G_i \mathbf{v}_i + \tilde{\mathbf{n}}_{z,i} \ , \ i = 1, \ldots, N \ , & (4.15) \end{aligned}$$

where $R_{\tilde{\mathbf{n}}_{z,i}\tilde{\mathbf{n}}_{z,i}} = S_{o,i}^{-1}$. Taking into account the feedback section, and ignoring the effect of error propagation, the decoder output is finally expressed:

$$\begin{aligned} \hat{\mathbf{v}}_i &= \text{decode}\left[\tilde{\mathbf{z}}_i + (I - G_i)\mathbf{v}_i\right] & (4.16)\\ &= \text{decode}\left(\mathbf{v}_i + \tilde{\mathbf{n}}_{z,i}\right) \ , \ i = 1, \ldots, N \ . & (4.17) \end{aligned}$$

Since the noise vector contains uncorrelated components, each of the symbols can be decoded independently. Thus, L parallel channels are created within each tone. The detection SNR of user j on tone i then equals

$$SNR_{i,j} = s_{i,j}\mathcal{E}_{i,j} \ , \tag{4.18}$$

where $s_{i,j}$ is the j-th diagonal element of $S_{o,i}$, and $\mathcal{E}_{i,j} = E\left(|v_{i,j}|^2\right)$ is the energy of the symbol of user j on tone i.

The system has the structure shown in Figure 4.1, where each of the *canceler* blocks eliminates crosstalk within a tone. Observe that the computation of the parameters of the structure only requires the QR factorization of $H_F = P\Lambda P$.

Note that the above technique can be combined with coding, where each user employs its own code. In that case, (4.16) indicates that decoding the j-th user requires that users $j+1$ through L have already been decoded. Since decoding typically involves a delay, it is seen that the symbols of the j-th user must be buffered, until users $j+1$ through L have been decoded. Evidently, users with smaller indices are decoded at a later time, implying higher latency for those users. Appendix C describes an algorithm for the computation of *distance spectra* of codes with trellis description, and provides the distance spectra of popular convolutional codes.

The impact of error propagation was completely ignored in (4.17). Also, the possibility of changing the order of decoding the users (e.g. so that error propagation is minimized) has not been explored yet. These issues are examined in the context of multi-line channels in Section 4.2.

4.1.3 Downstream

In this case, the receivers are assumed to be geographically separated, which implies that the receiver section must be separable into N independent modules corresponding to the N users. However, there is now freedom in choosing the transmit filter as

$$F = Q_{mIDFT} P B \ , \tag{4.19}$$

where the matrix B is derived shortly. Then,

$$R_f = B^* P \Lambda^* \Lambda P B \ . \tag{4.20}$$

The aim is to find an appropriate B, which can make R_f diagonal so that the receiver feedback section disappears. Utilizing the Cholesky factorization, one finds

$$P\Lambda^*\Lambda P = \operatorname{diag}\left(G_1^* S_{o,1} G_1, G_2^* S_{o,2} G_2, \ldots, G_N^* S_{o,N} G_N\right) \ . \tag{4.21}$$

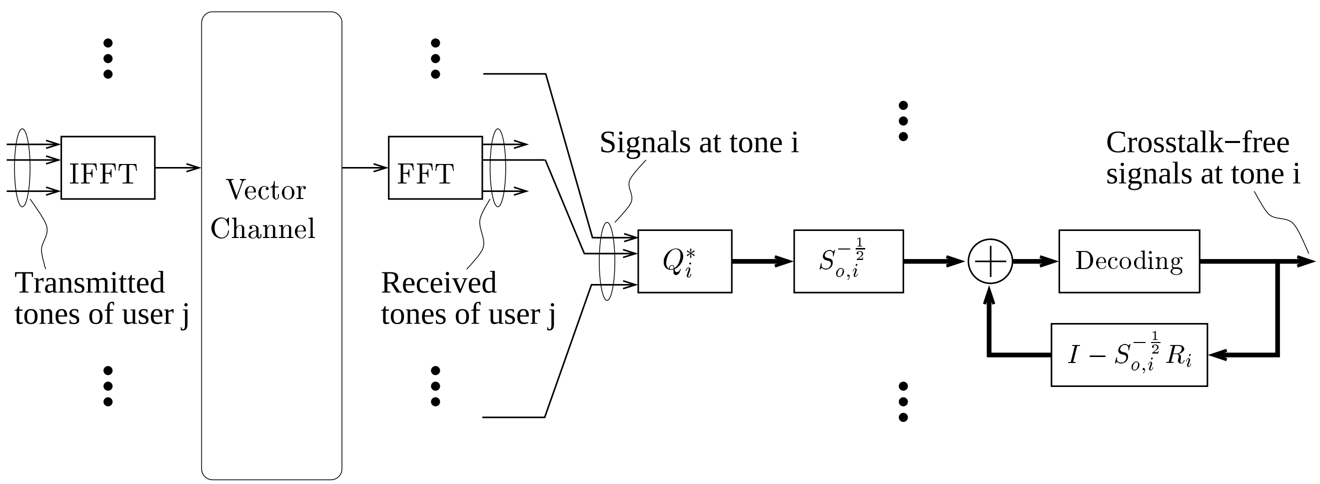

Figure 4.1: Crosstalk cancellation for upstream transmission.

Also, the following decomposition is defined

$$H_F^T = P\Lambda^T P = \mathrm{diag}\left(Q_1 R_1, Q_2 R_2, \ldots, Q_N R_N\right) \ , \tag{4.22}$$

where Q_i, $i = 1, \ldots, N$ are $L \times L$ unitary matrices, R_i, $i = 1, \ldots, N$ are $L \times L$ upper triangular matrices, and $S_{o,i}^{\frac{1}{2}}$ has diagonal elements equal to the magnitude of the diagonal elements of R_i. Observe that the difference with (4.8) is that H_F is here transposed. The choice of

$$B = \mathrm{diag}\left(\left(Q_1^*\right)^T R_1^{-T} S_{o,1}^{\frac{1}{2}}, \left(Q_2^*\right)^T R_2^{-T} S_{o,2}^{\frac{1}{2}}, \ldots, \left(Q_N^*\right)^T R_N^{-T} S_{o,N}^{\frac{1}{2}}\right) \tag{4.23}$$

produces $R_f = \mathrm{diag}\left(S_{o,1}, S_{o,2}, \ldots, S_{o,N}\right)$, which results in the trivial Cholesky factorization $R_f = G'^* S_o G'$ with $G' = I$. Therefore, there is no feedback section and the matched-feedforward filter becomes

$$\begin{aligned}
&S_o^{-1} G'^{-*} F^* H^* = \\
&S_o^{-1} B^* P \Lambda^* Q_{mDFT} = \\
&S_o^{-1} \mathrm{diag}\left(S_{o,1}^{\frac{1}{2}} \left(R_1^{-*}\right)^T Q_1^T, S_{o,2}^{\frac{1}{2}} \left(R_2^{-*}\right)^T Q_2^T, \ldots, S_{o,N}^{\frac{1}{2}} \left(R_N^{-*}\right)^T Q_N^T\right) \cdot \\
&\qquad \mathrm{diag}\left(\left(Q_1^*\right)^T \left(R_1^*\right)^T, \left(Q_2^*\right)^T \left(R_2^*\right)^T, \ldots, \left(Q_N^*\right)^T \left(R_N^*\right)^T\right) P Q_{mDFT} = \\
&S_o^{-\frac{1}{2}} P Q_{mDFT} \ .
\end{aligned} \tag{4.24}$$

Interpreting the receiver function, it is seen that the DFT is computed on the samples of each user and the output is scaled.

The transmitter function deserves more attention. Prior to IDFT modulation, the matrices $\left(Q_i^*\right)^T R_i^{-T} S_{o,i}^{1/2}$ operate independently on the symbols of each tone. The matrices $\left(Q_i^*\right)^T$ simply rotate the symbols, but $R_i^{-T} S_{o,i}^{1/2}$ may actually cause undesired energy increase.

Previous solutions to this problem are based on performing appropriate transmitter scaling to control the energy increase [58]. An alternative with much better performance is to utilize the concept of Tomlinson-Harashima precoding [59, 60, 61]. The module $R_i^{-T} S_{o,i}^{1/2}$ can be replaced by a feedback loop with $I - S_{o,i}^{-1/2} R_i^T$ in the feedback path. Then, in order to prevent energy increase, a modulo operator is placed in the forward path of the feedback loop, and a modulo operation is also applied at the receivers after scaling. This is explained in more detail in the following.

The signal processing steps are next described. The output of the N precoders can be expressed as

$$\tilde{\mathbf{v}}_i = \text{modulo}\left[\mathbf{v}_i + (I - S_{o,i}^{-1/2} R_i^T)\tilde{\mathbf{v}}_i\right] \ , \ i = 1, \ldots, N \ , \tag{4.25}$$

where $\mathbf{v}_i$ is the block of symbols of the L users corresponding to tone i. Also, the following definitions hold:

$$\text{modulo}\,(\mathbf{x}) = \begin{bmatrix} \Gamma_{M_1}(x_1) \\ \Gamma_{M_2}(x_2) \\ \vdots \\ \Gamma_{M_L}(x_L) \end{bmatrix} , \tag{4.26}$$

$$\Gamma_M\,(x) = x - Md\left\lfloor \frac{x + \frac{Md}{2}}{Md} \right\rfloor , \tag{4.27}$$

and M is the constellation size of x, while d is the constellation point spacing. (If x is complex, then $\Gamma_M\,[x] = \Gamma_{\sqrt{M}}\,[\Re(x)] + \jmath\Gamma_{\sqrt{M}}\,[\Im(x)]$.) An illustration of Γ_M is given in Figure 4.2, where the operation is visualized as "wrapping" the real line around a circle. The effect of (4.25) on the transmitted constellation is shown in Figure 4.3 for a simple example with only 2 users, and assuming real symbols.

The outputs of the precoders are then rotated by $(Q_i^*)^T$,

$$\mathbf{v}'_i = (Q_i^*)^T \tilde{\mathbf{v}}_i \ , \ i = 1, \ldots, N \Leftrightarrow \tag{4.28}$$

$$\mathbf{v}' = \text{diag}\left((Q_1^*)^T, (Q_2^*)^T, \ldots, (Q_N^*)^T\right) \tilde{\mathbf{v}} \ , \tag{4.29}$$

where $\tilde{\mathbf{v}} = \left[\tilde{\mathbf{v}}_1^T \ \tilde{\mathbf{v}}_2^T \ \ldots \ \tilde{\mathbf{v}}_N^T\right]^T$, and the channel input can then be expressed as

$$\mathbf{x} = Q_{mIDFT} P \mathbf{v}' \ . \tag{4.30}$$

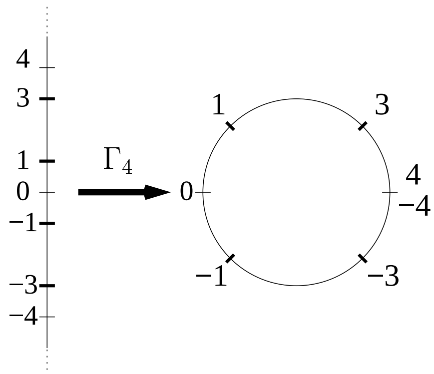

Figure 4.2: Illustration of the $\Gamma_M(x)$ operator with $M = 4$ and $d = 2$.

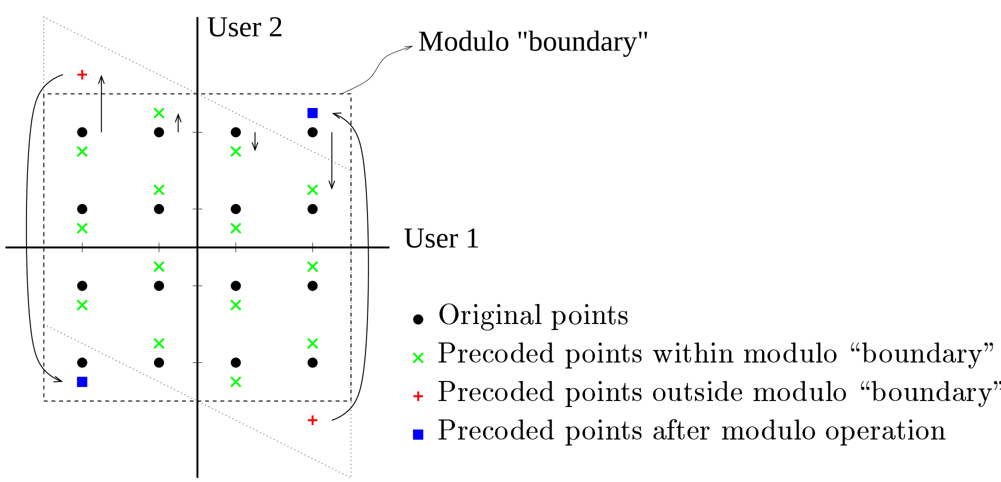

Figure 4.3: Effect of precoding on transmitted constellation for example of 2 users.

At the receiver, the following filtering is applied

$$\begin{aligned}
\mathbf{z}' &= S_o^{-\frac{1}{2}} P Q_{mDFT} \mathbf{y} && (4.31)\\
&= S_o^{-\frac{1}{2}} P \Lambda P \mathbf{v}' + S_o^{-\frac{1}{2}} P Q_{mDFT} \mathbf{n} && (4.32)\\
&= S_o^{-\frac{1}{2}} \mathrm{diag}\left(R_1^T, R_2^T, \ldots, R_N^T\right) \tilde{\mathbf{v}} + \tilde{\mathbf{n}}_z \ , && (4.33)
\end{aligned}$$

where $R_{\tilde{\mathbf{n}}_z \tilde{\mathbf{n}}_z} = S_o^{-1}$. Finally, the modulo operation is applied prior to decoding,

$$\begin{aligned}
\tilde{\mathbf{z}} &= \mathrm{modulo}\left(\mathbf{z}'\right) && (4.34)\\
&= \mathrm{modulo}\left(\mathbf{v} + \tilde{\mathbf{n}}_z\right) \ , && (4.35)
\end{aligned}$$

where $\mathbf{v} = \left[\mathbf{v}_1^T \ \mathbf{v}_2^T \ \ldots \ \mathbf{v}_N^T\right]^T$. In a well-designed system, $\tilde{\mathbf{n}}_z$ is small, so the approximation

$$\tilde{\mathbf{z}} = \mathbf{v} + \tilde{\mathbf{n}}_z \qquad (4.36)$$

holds. Since the noise vector components are uncorrelated, it is evident that parallel channels have been formed. Thus, the detection SNR of user j on tone i is

$$SNR_{i,j} = s_{i,j} \mathcal{E}_{i,j} \ , \qquad (4.37)$$

where $s_{i,j}$ is the j-th diagonal element of $S_{o,i}$, and $\mathcal{E}_{i,j} = E\left(|v_{i,j}|^2\right)$ is the energy of the symbol of user j on tone i.

The resulting system is depicted in Figure 4.4, where each of the *precoder* blocks eliminates crosstalk within a tone. Again, the determination of the system parameters only requires the computation of the QR factorization of $H_F^T = P \Lambda^T P$.

Note that this technique can also be combined with coding, where each user employs its own code, and coding only needs to be applied independently on the symbols of each user. Also, the effect of changing the order in which users are *precoded* (see (4.25)) is examined in Section 4.2.

With modulo precoding, as described in this subsection, the transmitted symbols lie in a hyper-cube (a square in the simple two-dimensional example of Figure 4.3). If the symbols were arranged inside a hyper-sphere, then, for a given total energy

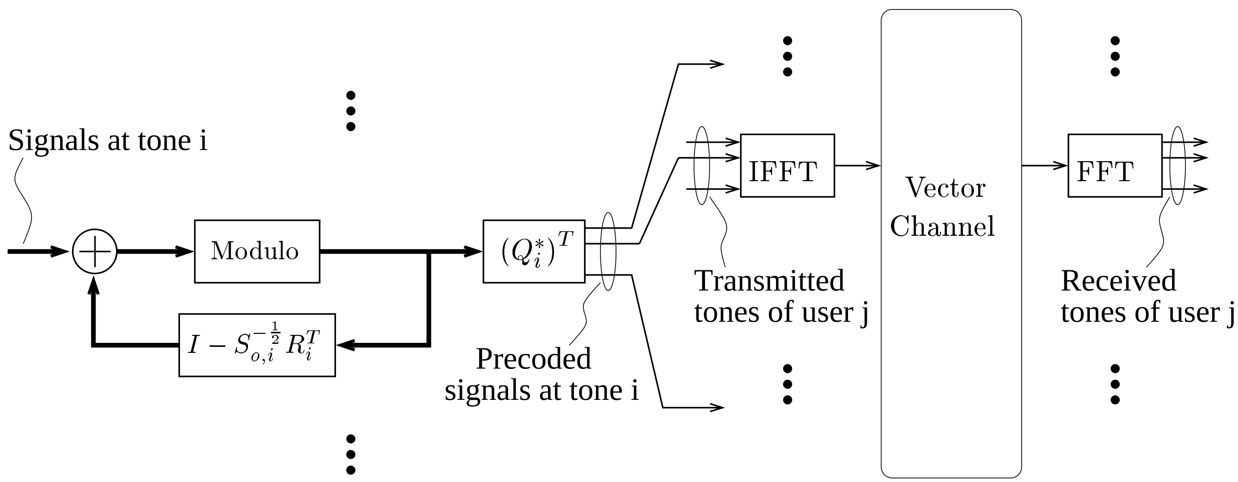

Figure 4.4: Crosstalk cancellation for downstream transmission.

constraint, more symbols would "fit" inside the hyper-sphere, and therefore more information would be transmitted per channel use. This effect also exists in single-user channels, and well-known techniques such as trellis shaping [62] have been developed, in order to make the transmitted constellations approximate hyper-spheres rather than hyper-cubes. The SNR gain associated with these techniques is termed *shaping gain*, and its theoretical maximum value is 1.53dB. For broadcast multi-user channels (e.g the downstream channel examined in this section), similar methods can be employed to realize this gain, and these include *lattice precoding* [31], and *trellis precoding* [63]. Reference [28] has independently proposed a transmission system, also adopting the QR factorization approach of this section, where lattice precoding is assumed.

4.2 Crosstalk Cancellation for Multi-line Channels

4.2.1 QR Factorization and Multi-line Channels

In the previous section, the derivation of the crosstalk cancellation parameters is based on the QR factorization. The procedure differs depending on whether upstream or downstream transmission is considered. In order to differentiate between upstream and downstream, the superscripts ^{u} and ^{d} are correspondingly used. For upstream transmission, (4.8) states

$$
\begin{aligned}
H_F^u = P\Lambda^u P &= \operatorname{diag}\left(H_{F,1}^u, H_{F,2}^u, \ldots, H_{F,N}^u\right) && (4.38)\\
&= \operatorname{diag}\left(Q_1^u R_1^u, Q_2^u R_2^u, \ldots, Q_N^u R_N^u\right)\ . && (4.39)
\end{aligned}
$$

For downstream transmission, (4.22) gives

$$
\begin{aligned}
H_F^d = P\Lambda^d P &= \operatorname{diag}\left(H_{F,1}^d, H_{F,2}^d, \ldots, H_{F,N}^d\right) && (4.40)\\
&= \operatorname{diag}\left(\left(R_1^d\right)^T \left(Q_1^d\right)^T, \left(R_2^d\right)^T \left(Q_2^d\right)^T, \ldots, \left(R_N^d\right)^T \left(Q_N^d\right)^T\right) . && (4.41)
\end{aligned}
$$

Assuming that (2.26) holds, it is deduced that $Q_i^u = Q_i^d$, $R_i^u = R_i^d$ for all i. However, (2.26) may not hold in practice, so upstream and downstream is treated independently in the following.

Recall the observations of Subsection 2.1.4, where it was shown that $H_{F,i}^u$, $i = 1, \ldots, N$ are column-wise diagonally dominant, while $H_{F,i}^d$, $i = 1, \ldots, N$ are row-wise diagonally dominant. These properties have interesting consequences for the crosstalk cancellation methods of Section 4.1.

The upstream case is considered first. Visualizing the columns of $H_{F,i}^u$ in vector space, it is seen that they are almost orthogonal to each other implying that Q_i^u is close to being an identity matrix. Thus, the magnitudes of the diagonal elements of R_i^u do not differ significantly from those of the diagonal elements of $H_{F,i}^u$. This is illustrated in Figure 4.5 for a two user case, where $\mathbf{h}_1$, $\mathbf{h}_2$ are the columns of $H_{F,i}^u$, $\mathbf{q}_1$, $\mathbf{q}_2$ are the columns of Q_i^u, r_{mn} is the (m, n) element of R_i^u, and $|r_{11}| = ||\mathbf{h}_1||$. The fact that $\mathbf{h}_1$, $\mathbf{h}_2$ are almost orthogonal to the axes implies that $|r_{22}| \simeq ||\mathbf{h}_2||$. The same conclusion holds even if the two columns are swapped, and this is also shown in Figure 4.5. The following proposition generalizes this observation for more users.

Proposition 4.2.1 *Let $\mathbf{h}_j$ be the j-th column, and h_{ij} be the (i, j) element of matrix H. Assume that*

$$|h_{ij}| \leq |h_{jj}| \tan\alpha \ , \ i \neq j \ , \tag{4.42}$$

where $0 \leq \alpha \leq 0.5$. Also, assume the QR factorization,

$$H = QR \ , \tag{4.43}$$

where Q is a unitary matrix, and R is an upper-triangular matrix. Letting r_{ij} be the (i, j) element of R, the following bounds then hold for all j:

$$|h_{jj}|\sqrt{1 - 4\alpha^2} \leq |r_{jj}| \leq |h_{jj}| \left[1 + (L-1)\frac{\tan^2\alpha}{2}\right] \ . \tag{4.44}$$

Proof: For the purpose of geometric illustration, the proof assumes real matrices, but it can be generalized for complex matrices. Notice that α has the geometric

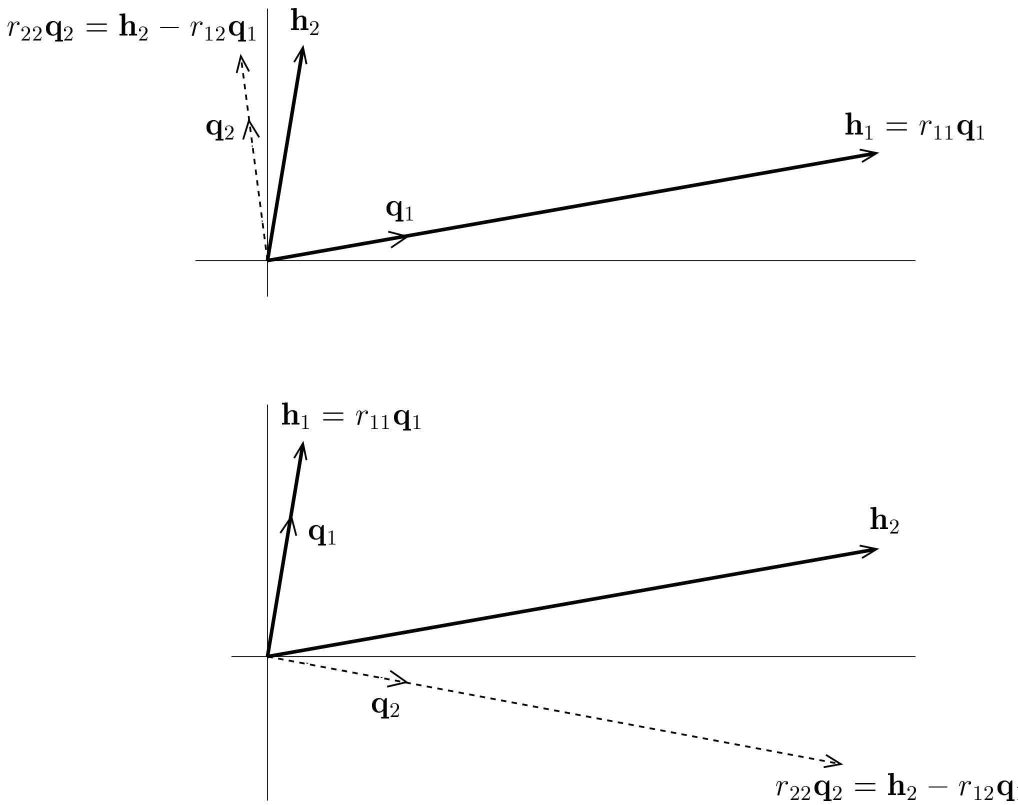

Figure 4.5: Illustration of QR factorization for the two possible orderings.

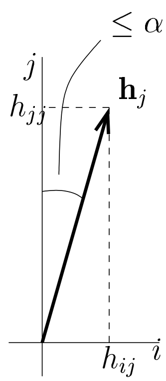

Figure 4.6: Geometric interpretation of α.

interpretation of being the maximum angle between axis j and vector $\mathbf{h}_j$ (see Figure 4.6).

Starting with $|r_{11}|$, one observes that

$$\begin{aligned} |r_{11}| &= |\mathbf{h}_1| = \sqrt{\sum_{i=1}^{L} |h_{i1}|^2} \\ &\leq |h_{11}|\sqrt{1+(L-1)\tan^2\alpha} \qquad (4.45) \\ &\leq |h_{11}|\left[1+(L-1)\frac{\tan^2\alpha}{2}\right] , \qquad (4.46) \end{aligned}$$

where (4.45) makes use of (4.42), and (4.46) results from $\sqrt{1+x} \leq 1 + x/2$. The following obvious lower bound can also be obtained

$$|r_{11}| \geq |h_{11}| \ . \qquad (4.47)$$

Next, bounds are obtained for $|r_{22}|$. In Figure 4.7, $\mathbf{q}_j$ is the j-th column of Q. Examining the figure, one sees that the upper bound on $|r_{22}|$ is achieved when $\mathbf{h}_1 \perp \mathbf{h}_2$, in which case $r_{12} = 0$ and

$$\begin{aligned} |r_{22}| &= |\mathbf{h}_2| \\ &\leq |h_{22}| \left[1 + (L-1)\frac{\tan^2 \alpha}{2}\right] . \end{aligned} \tag{4.48}$$

For the lower bound, letting $h_{2i} = 0$ for $i \neq 1, 2$ suffers no loss of generality. Then, $|r_{22}|$ is minimized, when the projection of $\mathbf{h}_2$ on $\mathbf{q}_1$ is minimized. As shown in Figure 4.7, this occurs when the angle equals $\pi/2 - 2\alpha$, in which case

$$|r_{12}| = |\mathbf{h}_2| \cos(\pi/2 - 2\alpha) = |\mathbf{h}_2| \sin 2\alpha . \tag{4.49}$$

Therefore,

$$\begin{aligned} |r_{22}| &= \sqrt{|\mathbf{h}_2|^2 + |r_{12}|^2 - 2|\mathbf{h}_2||r_{12}| \sin 2\alpha} && (4.50) \\ &\geq \sqrt{|\mathbf{h}_2|^2 + |\mathbf{h}_2|^2 \sin^2 2\alpha - 2|\mathbf{h}_2|^2 \sin 2\alpha \sin 2\alpha} && (4.51) \\ &= |\mathbf{h}_2|\sqrt{1 - \sin^2 2\alpha} \\ &\geq |h_{22}|\sqrt{1 - \sin^2 2\alpha} && (4.52) \\ &\geq |h_{22}|\sqrt{1 - 4\alpha^2} , && (4.53) \end{aligned}$$

where (4.50) results from vector addition, (4.51) makes use of (4.49), (4.52) uses the fact $|\mathbf{h}_2| \geq |h_{22}|$, and (4.53) results from $\sin(x) \leq x$, and holds for $\alpha < 0.5$.

Similar arguments that were used for deriving the bounds of $|r_{22}|$ can be used for $|r_{jj}|$. The upper bound on $|r_{jj}|$ is achieved when $\mathbf{h}_j \perp \mathbf{h}_k$, $k < j$, in which case

$$|r_{jj}| = |\mathbf{h}_j| \leq |h_{jj}| \left[1 + (L-1)\frac{\tan^2 \alpha}{2}\right] . \tag{4.54}$$

For the lower bound, it is seen that $|r_{jj}|$ is minimized, when the angle between $\mathbf{q}_{j-1}$ and $\mathbf{h}_j$ is minimized. Again, this angle must be greater than $\pi/2 - 2\alpha$, which leads to

$$|r_{jj}| \geq |h_{jj}|\sqrt{1 - 4\alpha^2} . \tag{4.55}$$

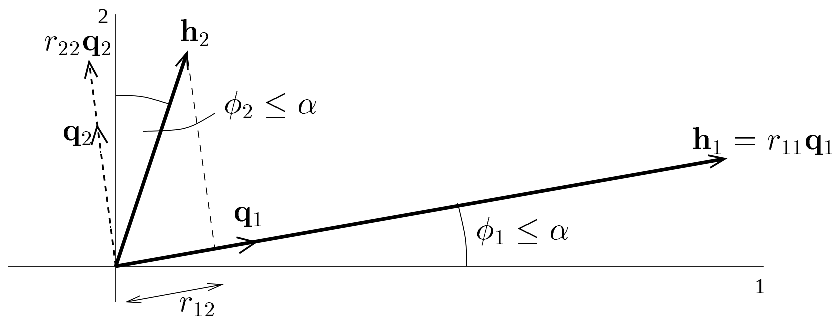

Figure 4.7: Illustration for QR bounds derivation.

In conclusion, the following bounds hold for all j:

$$|h_{jj}|\sqrt{1-4\alpha^2} \leq |r_{jj}| \leq |h_{jj}| \left[1+(L-1)\frac{\tan^2\alpha}{2}\right] \ . \tag{4.56}$$

■

Since $H^u_{F,i}$ are column-wise diagonally dominant, then they satisfy the conditions of the proposition with a very small α (see example of Figure 2.7). Therefore,

$$s^u_{i,j} \simeq |H^u_{F,i}(j,j)|^2 \ , \tag{4.57}$$

where $s^u_{i,j}$, $j=1,\ldots,L$ are the diagonal elements of $S^u_{o,i}$ (see (4.7)), which equal the squared values of the diagonal elements of R^u_i. The column-wise diagonally dominant property also implies that

$$|H^u_{F,i}(j,j)|^2 \simeq |\mathbf{h}^u_{F,i}(j)|^2 \ , \tag{4.58}$$

where $\mathbf{h}^u_{F,i}(j)$ is the j-th column of $H^u_{F,i}$.

Similar observations can be made for downstream transmission. Since $H^d_{F,i}$ is row-wise diagonally dominant, $\left(H^d_{F,i}\right)^T$ must be column-wise diagonally dominant.

Applying (4.57) and (4.58) to $\left(H^d_{F,i}\right)^T$ yields

$$\begin{aligned} s^d_{i,j} &\simeq |H^d_{F,i}(j,j)|^2 \ , && (4.59)\\ |H^d_{F,i}(j,j)|^2 &\simeq |\tilde{\mathbf{h}}^d_{F,i}(j)|^2 \ , && (4.60)\end{aligned}$$

where $s^d_{i,j}$, $j = 1,\ldots,L$ are the diagonal elements of $S^d_{o,i}$ (see (4.21)), which equal the squared values of the diagonal elements of R^d_i, and $\tilde{\mathbf{h}}^d_{F,i}(j)$ is the j-th row of $H^d_{F,i}$.

If one changes the order in which users are decoded, then this corresponds to multiplying $H^u_{F,i}$ on the right and on the left with a permutation matrix. For example, interchanging users m and n swaps rows m with n, and columns m with n,

$$\begin{bmatrix} & h_{1m} & h_{1n} & \\ h_{m1} & h_{mm} & h_{mn} & h_{mL} \\ h_{n1} & h_{nm} & h_{nn} & h_{nL} \\ & h_{Lm} & h_{Ln} & \end{bmatrix} \xrightarrow{\text{re-order}} \begin{bmatrix} & h_{1n} & h_{1m} & \\ h_{n1} & h_{nn} & h_{nm} & h_{nL} \\ h_{m1} & h_{mn} & h_{mm} & h_{mL} \\ & h_{Ln} & h_{Lm} & \end{bmatrix} . \qquad (4.61)$$

Such re-ordering affects Q^u_i and R^u_i of (4.39). However, observe that the matrix remains column-wise diagonally dominant, therefore, (4.57) and (4.58) still hold.

Changing the order in which users are precoded corresponds to multiplying $H^d_{F,i}$ on the right and on the left with a permutation matrix. Although this affects Q^d_i and R^d_i of (4.41), the matrix remains row-wise diagonally dominant, and (4.59), (4.60) continue to hold.

4.2.2 Performance Approximations

Equation (4.18) gives the detection SNR for upstream transmission using the crosstalk cancellation of Subsection 4.1.2,

$$SNR^u_{i,j} = s^u_{i,j}\mathcal{E}^u_{i,j} \ . \qquad (4.62)$$

On the other hand, assume upstream transmission in a multi-line channel with *no crosstalk*. Clearly, $H^u_{F,i}$ is then diagonal, and the detection SNR of user j on tone i

of this hypothetical system is

$$\widehat{SNR}^u_{i,j} = |H^u_{F,i}(j,j)|^2 \mathcal{E}^u_{i,j} \ . \tag{4.63}$$

Finally, assume that there is *only one user* transmitting upstream in the multi-line channel, and that the receiver only needs to optimally decode this single user. In this case, $H^u_{F,i}$ is all zeroes, except for the column corresponding to the transmitting user. The detection SNR for user j on tone i is then equal to

$$\widehat{\widehat{SNR}}^u_{i,j} = |\mathbf{h}^u_{F,i}(j)|^2 \mathcal{E}^u_{i,j} \ . \tag{4.64}$$

From (4.57) and (4.58), it is deduced that

$$SNR^u_{i,j} \simeq \widehat{SNR}^u_{i,j} \simeq \widehat{\widehat{SNR}}^u_{i,j} \ . \tag{4.65}$$

The situation for upstream is similar. Equation (4.37) gives the downstream SNR for user j on tone i utilizing the crosstalk mitigation method of Subsection 4.1.3,

$$SNR^d_{i,j} = s^d_{i,j} \mathcal{E}^d_{i,j} \ . \tag{4.66}$$

The detection SNR in a channel with no crosstalk is

$$\widehat{SNR}^d_{i,j} = |H^d_{F,i}(j,j)|^2 \mathcal{E}^d_{i,j} \ . \tag{4.67}$$

Finally, the detection SNR with only a single user receiving is

$$\widehat{\widehat{SNR}}^d_{i,j} = |\tilde{\mathbf{h}}^d_{F,i}(j)|^2 \mathcal{E}^d_{i,j} \ . \tag{4.68}$$

Finally, the use of (4.59), (4.60) results in

$$SNR^d_{i,j} \simeq \widehat{SNR}^d_{i,j} \simeq \widehat{\widehat{SNR}}^d_{i,j} \ . \tag{4.69}$$

As explained in Chapter 5, the detection SNR can be used to calculate an approximation for the achievable data rate, and thus, it defines a very useful performance metric. Therefore, (4.69) reveals that the performance of the crosstalk cancellation methods of Section 4.1 for multi-line channels is very close to the performance in a hypothetical system where no crosstalk exists. Additionally, this performance is very close to that achieved by assuming that only one user is utilizing the multi-line channel.

This result is illustrated for an example DSL channel. Figure 4.8 shows the upstream *channel* SNRs for the same loop topology as in Figure 2.7, either assuming that no crosstalk exists, or that crosstalk is cancelled by the method of Subsection 4.1.2. Channel SNR is defined as equal to the SNR with unit transmission energy. The background noise is white with PSD equal to -140 dBm/Hz. Since the QR factorizations differ when the columns of $H^u_{F,i}$ are swapped, the two possible orderings are displayed. Clearly, the plots corresponding to the same loop overlap almost perfectly, which means that QR cancellation is as good as having no crosstalk, and that the impact of ordering can be ignored. Also, this serves as evidence that the crosstalk ZF criterion is adequate, and employing an MMSE criterion will not yield noticeable improvement.

4.2.3 Practical Considerations

Multi-line channels can be characterized as stationary, since they do not change significantly over time. Thus, the channel is typically identified by training during an initialization period, and the resulting channel estimation is very accurate. Still, it is illuminating to examine the effect of channel estimation errors.

Let the estimated upstream and downstream channel matrices for tone i be

$$\hat{H}^u_{F,i} = H^u_{F,i} + E^u_i \ , \tag{4.70}$$

$$\hat{H}^d_{F,i} = H^d_{F,i} + E^d_i \ , \tag{4.71}$$

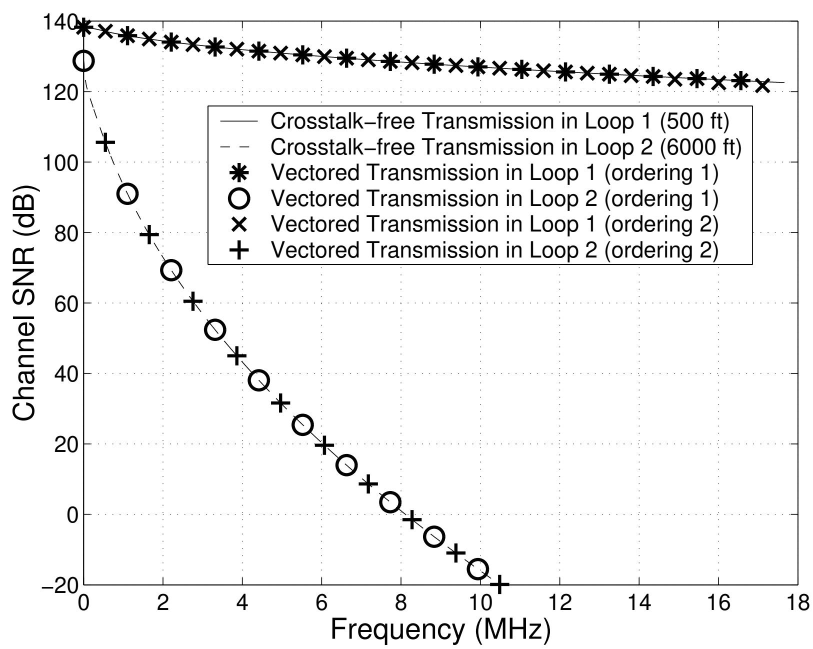

Figure 4.8: Channel SNRs with no crosstalk and with crosstalk cancellation (shown for both possible user orderings).

where E_i^u, E_i^d are the corresponding channel estimation errors. Then, the QR decompositions of $\hat{H}_{F,i}^u$, $\left(\hat{H}_{F,i}^d\right)^T$ produce

$$H_{F,i}^u = \hat{Q}_i^u \hat{R}_i^u - E_i^u \ , \tag{4.72}$$

$$H_{F,i}^d = \left(\hat{R}_i^d\right)^T \left(\hat{Q}_i^d\right)^T - E_i^d \ , \tag{4.73}$$

where $\hat{Q}_i^u$, $\hat{R}_i^u$, $\hat{Q}_i^d$, $\hat{R}_i^d$ are the QR factor estimates. Matrices $\hat{S}_{o,i}^u$ and $\hat{S}_{o,i}^d$ are defined as diagonal matrices with elements equal to the squared diagonal elements of $\hat{R}_i^u$ and $\hat{R}_i^d$. Starting from (4.13) and (4.16), the effect on upstream communication can be computed,

$$\begin{aligned}
\tilde{\mathbf{z}}_i &= \left(\hat{S}_{o,i}^u\right)^{-\frac{1}{2}} \left(\hat{R}_i^u - \left(\hat{Q}_i^u\right)^* E_i^u\right) \mathbf{v}_i + \tilde{\mathbf{n}}_{z,i} \Rightarrow \\
\tilde{\mathbf{v}}_i &= \text{decode}\left(\mathbf{v}_i - \left(\hat{S}_{o,i}^u\right)^{-\frac{1}{2}} \left(\hat{Q}_i^u\right)^* E_i^u \mathbf{v}_i + \tilde{\mathbf{n}}_{z,i}\right) \\
&= \text{decode}\left(\left(1 - \left(\hat{S}_{o,i}^u\right)^{-\frac{1}{2}} \text{diagonal}\left(\left(\hat{Q}_i^u\right)^* E_i^u\right)\right) \mathbf{v}_i - \right. \\
&\quad \left. \left(\hat{S}_{o,i}^u\right)^{-\frac{1}{2}} \left(\left(\hat{Q}_i^u\right)^* E_i^u - \text{diagonal}\left(\left(\hat{Q}_i^u\right)^* E_i^u\right)\right) \mathbf{v}_i + \tilde{\mathbf{n}}_{z,i}\right) \ .
\end{aligned} \tag{4.74}$$

From the expression above, one sees that the estimation errors impact transmission by introducing a “bias” in the detection (first term in (4.74)), and also by resulting in some residual crosstalk (second term in (4.74)).

A similar analysis can be applied for downstream communication, but modulo arithmetic complicates the expressions. Ignoring the modulo operations and employing (4.25), (4.29), and (4.31) gives

$$\begin{aligned}
\tilde{\mathbf{z}}_i &= \left(\hat{S}_{o,i}^d\right)^{-\frac{1}{2}} \left(\left(\hat{R}_i^d\right)^T \left(\hat{Q}_i^d\right)^T - E_i^d\right) \mathbf{v}_i' + \tilde{\mathbf{n}}_{z,i} \\
&= \mathbf{v}_i - \left(\hat{S}_{o,i}^d\right)^{-\frac{1}{2}} E_i^d \left(\hat{Q}_i^d\right)^{T*} \tilde{\mathbf{v}}_i + \tilde{\mathbf{n}}_{z,i} \\
&= \mathbf{v}_i - \left(\hat{S}_{o,i}^d\right)^{-\frac{1}{2}} E_i^d \left(\hat{Q}_i^d\right)^{T*} \left(\hat{R}_i^d\right)^{-T} \left(\hat{S}_{o,i}^d\right)^{\frac{1}{2}} \mathbf{v}_i + +\tilde{\mathbf{n}}_{z,i} \ .
\end{aligned} \tag{4.75}$$

Just as before, the effects of the estimation errors can be separated into a detection bias term and a residual crosstalk term.

Equations (4.74) and (4.75) reveal that the impact of channel estimation errors is aggravated, when any of the diagonal elements of $S^u_{o,i}$, or $S^d_{o,i}$ is small. The diagonally dominant structure of $H^u_{F,i}$ and $H^d_{F,i}$ for multi-line channels makes it practically impossible for $H^u_{F,i}$ and $H^d_{F,i}$ to be singular. However, an ill-conditioned channel (implying small elements for $S^u_{o,i}$ and $S^d_{o,i}$) cannot be ruled out, thus increasing the impact of channel estimation errors and posing several computational problems. Such cases may arise at high frequencies, for example in DSL loop topologies similar to that of Figure 2.7. Nevertheless, the energy allocation algorithms proposed in Chapter 5 prevent the occurence of such phenomena by not allowing transmission within those tones and by those users that correspond to small elements of $S^u_{o,i}$ or $S^d_{o,i}$.

Also, it is interesting to comment on the complexity of crosstalk cancellation with QR factorization. The computational cost is decomposed into the cost of the QR factorizations and the cost associated with signal processing. Multi-line channels are stationary, so the QR factorizations need to be computed infrequently (during initialization). Generally, this requires $\mathcal{O}(L^3)$ flops per tone (e.g. using the Householder transform as in [64]). Another $\mathcal{O}(L^2)$ flops per tone are required for the operations of (4.15) and (4.29). However, these costs can be greatly reduced in certain instances by taking advantage of the crosstalk environment characteristics.

It is known that the crosstalk noise in a DSL pair originates mostly from just three or four neighboring pairs (as in Figure 4.9), which implies that $H^u_{F,i}$, or $H^d_{F,i}$ are practically sparse with only three or four relatively large off-diagonal elements per row. Therefore, approximating $H^u_{F,i}$, $H^d_{F,i}$ as sparse matrices, Givens rotations can be employed to triangularize these matrices with a reduced cost of $\mathcal{O}(L^2)$ flops. The real-time computational burden due to the canceler and precoder blocks can also be reduced by employing Givens rotations. It can be shown that the operations of (4.15) and (4.29) require only $\mathcal{O}(L)$ flops per tone.

The effects of error propagation in Subsection 4.1.2 were ignored, but in practice they must be taken into account. First, it is argued that such effects are in most cases not severe. Then, a method to control error propagation is outlined.

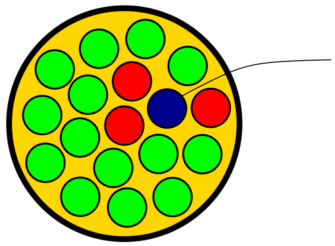

Figure 4.9: Cross-section of cable.

It should be pointed out that for block decision feedback operations, error propagation is not a "catastrophic" event (as in DFEs for ISI mitigation), since it always terminates when all users in (4.16) have been decoded. Also, in typical systems, each user's probability of symbol error is designed to be at a specified level P_e (smaller than 10^{-7} for DSL). In the worst case, the probability of symbol error of the last decoded user equals LP_e, which for a small L gives a small (but perhaps significant) increase. The sparse structure of the channel matrix in DSL implies that this worst case probability is actually closer to $L_1 P_e$, where L_1 equals 3 to 4.

In order to control this phenomenon, the following method can be used: Each user's transmission is allowed to have a different probability of symbol error requirement, so that users that are decoded first are designed such that they are more reliable. If errors are assumed to lead to detection errors for all later decisions in the scheme, then the probability of error of user j equals

$$P_{e,j} = P_{n,j} + P_{e,j-1} \ , \tag{4.76}$$

where $P_{n,j}$ is the probability that the noise of user j exceeds the minimum distance (ignoring error propagation). By setting $P_{n,j} = P_{n,j-1}/\mathcal{M}$, it is found that

$$P_{e,j} = P_{n,1} \sum_{m=0}^{j} \left(\frac{1}{\mathcal{M}}\right)^m < \frac{\mathcal{M}}{\mathcal{M}-1} P_{n,1} \ , \tag{4.77}$$

where $\mathcal{M}$ is a constant, chosen appropriately to control error propagation.

Chapter 6 investigates blind receiver training, in which case error propagation effects are much more critical.

Chapter 5

Transmission Optimization with Coordinated Communication

THE DESIGN OF ANY ADVANCED COMMUNICATION SYSTEM must address the problem of allocating the available resources (e.g. bandwidth and transmitter power) in such a way so that the system performance is maximized. Additionally, multiuser designs must also consider the distribution of these resources among the users.

The problem of optimizing the transmission parameters of multi-line channels has provided motivation for previous research work. Reference [21] proposed an algorithm, analogous to water-filling, for maximizing the capacity of a DSL system suffering from crosstalk from the same kind of systems. In [65], the issue of optimizing ADSL transmission in the presence of crosstalk from other DSL sources was studied. The question of the best spectrum use for simultaneous upstream and downstream transmission was explored in [66]. A distributed multi-user algorithm for optimizing the transmitted Power Spectral Densities (PSDs) of VDSL systems was described in [67, 68].

This chapter examines transmission optimization in light of the crosstalk cancellation techniques of Chapter 4, and gives solutions for specific transmission scenarios. Section 5.1 shows that energy allocation in systems employing crosstalk cancellation amounts to independent water-filling for each user. When power back-off or joint energy allocation and frequency planning must be considered, the optimization problems are solved through efficient numerical algorithms. Section 5.2 gives simulation results assuming the application of the previously described techniques in practical cases with realistic parameters corresponding to VDSL systems.

5.1 Transmission Optimization Scenarios

5.1.1 Energy Allocation

The optimization objective is the maximization of the weighted sum of the data rates of all users,

$$\max \sum_{j=1}^{L} a_j r_j \ , \tag{5.1}$$

where $a_j \geq 0$ is the weight assigned to the j-th user , and r_j is the achievable data rate of the j-th user, which may refer to either the upstream or the downstream direction.

In order to compute the data rate, the *gap* approximation is employed. The concept of the SNR-gap for flat channels is here briefly reviewed. It is well-known that the capacity of a Gaussian channel is [19]

$$c = \frac{1}{2}\log_2\left(1+|g|^2\mathcal{E}\right) \ \text{bits/use} \ , \tag{5.2}$$

where g is the channel gain, $\mathcal{E}$ is the average transmitted power, and unit noise variance is assumed. According to the *gap* approximation, the data rate of a practical communication system for the same channel is (see [69], and the earlier work in [70])

$$b = \frac{1}{2}\log_2\left(1+\frac{|g|^2\mathcal{E}}{\Gamma}\right) \ \text{bits/use} \ , \tag{5.3}$$

where Γ is defined as the SNR-gap. (Note that the approximation holds only for $b \geq 0.5$.) A simple interpretation is the following: Assume that a specific data rate and probability of symbol error are defined. Then, Γ signifies the difference (in dB) between the SNR required for the capacity to equal the intended data rate and the

SNR required by the practical system to achieve the intended data rate with the prescribed probability of symbol error. The relationship

$$\Gamma = \Gamma_{uncoded} - \gamma_{code} + \gamma_{margin} \text{ (dB)} \tag{5.4}$$

holds, where $\Gamma_{uncoded}$ is the SNR-gap of a system with no channel coding, γ_{code} is the *coding gain*, and γ_{margin} is the required *noise margin*. The parameter $\Gamma_{uncoded}$ is inversely related to the required probability of symbol error. The quantity γ_{margin} defines the additional available margin against sudden noise increase, or corresponding channel degradation.

In the following, Frequency Division Duplex (FDD) is assumed for the separation of upstream and downstream. The SNRs within each tone and for each user are given by (4.18) and (4.37). Taking into account the fact that vectoring essentially "diagonalizes" the channel and assuming no error propagation in upstream, the upstream and downstream achievable rates are obtained as

$$r_j^u = \sum_{i \in \mathcal{N}_u} \frac{1}{2} \log_2 \left(1 + \frac{s_{i,j}^u \mathcal{E}_{i,j}^u}{\Gamma} \right) , \tag{5.5}$$

$$r_j^d = \sum_{i \in \mathcal{N}_d} \frac{1}{2} \log_2 \left(1 + \frac{s_{i,j}^d \mathcal{E}_{i,j}^d}{\Gamma} \right) , \tag{5.6}$$

where $\mathcal{N}_u$ and $\mathcal{N}_d$ are the sets of upstream and downstream tone indices correspondingly, which depend on the FDD plan. Also, $s_{i,j}^u$, $\mathcal{E}_{i,j}^u$ and $s_{i,j}^d$, $\mathcal{E}_{i,j}^d$ are defined as in (4.18) and (4.37).

The parameters with respect to which optimization takes place are $\mathcal{E}_{i,j}^u$ for upstream, and $\mathcal{E}_{i,j}^d$ for downstream transmission, and these are constrained by limits on the transmitted energy. In upstream transmission, the total transmitted energy of user j is constrained by

$$\text{trace}\left(R_{\mathbf{x}_j \mathbf{x}_j}\right) \leq \mathcal{E}_j^u , \tag{5.7}$$

where $\mathcal{E}_j^u$ is the maximum allowed upstream transmitted energy of user j. From (4.3) it is seen that for user j

$$\mathbf{x}_j = Q_{IDFT}\left[v_{1,j}\ v_{2,j}\ \ldots\ v_{N,j}\right]^T\ . \tag{5.8}$$

Thus, from (5.7) it is deduced that

$$\sum_{i\in\mathcal{N}_u} \mathcal{E}^u_{i,j} \le \mathcal{E}^u_j\ . \tag{5.9}$$

For downstream transmission, a total transmit energy constraint is defined for each line emanating from the central node. This is a realistic constraint for multi-line channels, which usually corresponds to power limitations of the analog line-driver amplifier circuits of the modems. For line j, this is expressed as

$$\operatorname{trace}\left(R_{\mathbf{x}_j\mathbf{x}_j}\right) \le \mathcal{E}^d_j\ , \tag{5.10}$$

where $\mathcal{E}^d_j$ is the maximum allowed energy transmitted downstream on line j. From (4.30) it is seen that the input to line j is

$$\mathbf{x}_j = Q_{IDFT}\left[v'_{1,j}\ v'_{2,j}\ \ldots\ v'_{N,j}\right]^T\ . \tag{5.11}$$

Next, $R_{\mathbf{x}_j\mathbf{x}_j}$ must be expressed in terms of $R_{\mathbf{v}_i\mathbf{v}_i}$, $i = 1,\ldots,N$. Although $R_{\mathbf{v}'_i\mathbf{v}'_i} = (Q^*_i)^T R_{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i} (Q_i)^T$ from (4.28), the modulo operation of (4.25) presents an obstacle in deriving $R_{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i}$ from $R_{\mathbf{v}_i\mathbf{v}_i}$. However, it has been observed through simulation that

$$R_{\mathbf{v}'_i\mathbf{v}'_i} \simeq R_{\mathbf{v}_i\mathbf{v}_i} \tag{5.12}$$

is a valid approximation.

Simulation results confirming this claim are next presented for the pessimistic DSL loop topology of Figure 2.7. For simulation purposes it is assumed that the transmission PSD is flat at −60dBm/Hz, the noise PSD is flat at −140dBm/Hz, and the SNR-gap is 12dB. Thus, an integer bit distribution is derived, where the maximum number of bits per tone is set to be 12. The precoder is simulated in the

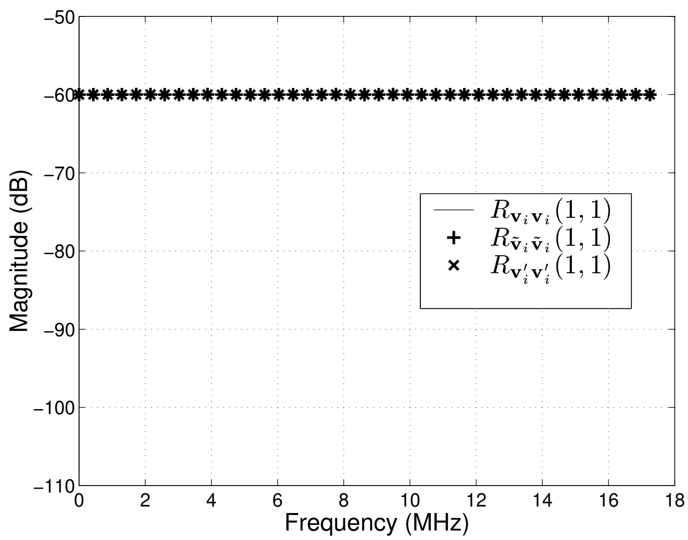

Figure 5.1: Magnitudes of elements $(1,1)$ of $R_{\mathbf{v}_i\mathbf{v}_i}$, $R_{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i}$, and $R_{\mathbf{v}'_i\mathbf{v}'_i}$.

time domain, and the sample covariance matrices of $\tilde{\mathbf{v}}_i$, $\mathbf{v}'_i$ and $\mathbf{v}_i$ are computed after 10000 iterations for a number of tones. Figures 5.1, 5.2, and 5.3 show elements $(1,1)$, $(2,2)$ and $(1,2)$ of the sample covariance matrices correspondingly.

It is seen in Figure 5.1 that user 1 transmits information on all tones, and that the energy increase at the precoder output is indiscernible. From Figure 5.2, it is evident that user 2 does not transmit information on tones with frequency higher than 2MHz. There is no energy increase below 2MHz, but at higher frequencies, the rotation operation of (4.28) causes the precoder output to have a very small amount of energy (30dB below the nominal value). For practical purposes, it can be claimed that there is no energy increase in any frequency. As seen in Figure 5.3, the correlation at the precoder output is very small at frequencies below 2MHz. At higher frequencies, the correlation becomes stronger, but always remains 15dB below the nominal transmission PSD level. Thus, the approximation can be made that the precoder outputs are uncorrelated and (5.12) is a reasonable approximation.

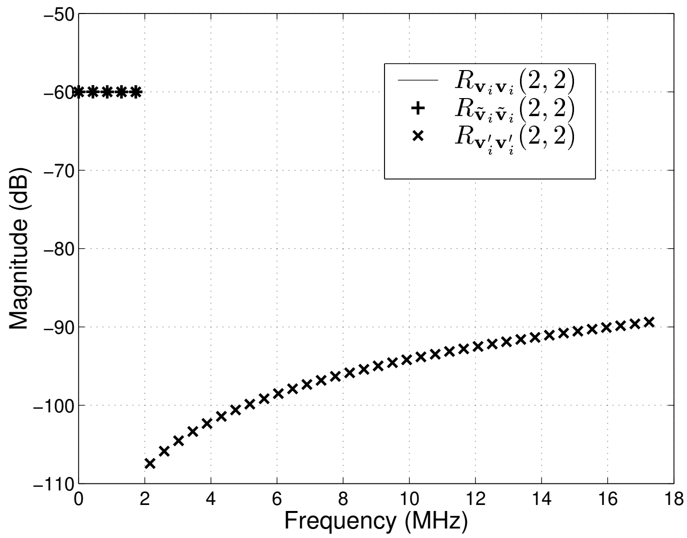

Figure 5.2: Magnitudes of elements $(2,2)$ of $R_{\mathbf{v}_i\mathbf{v}_i}$, $R_{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i}$, and $R_{\mathbf{v}'_i\mathbf{v}'_i}$.

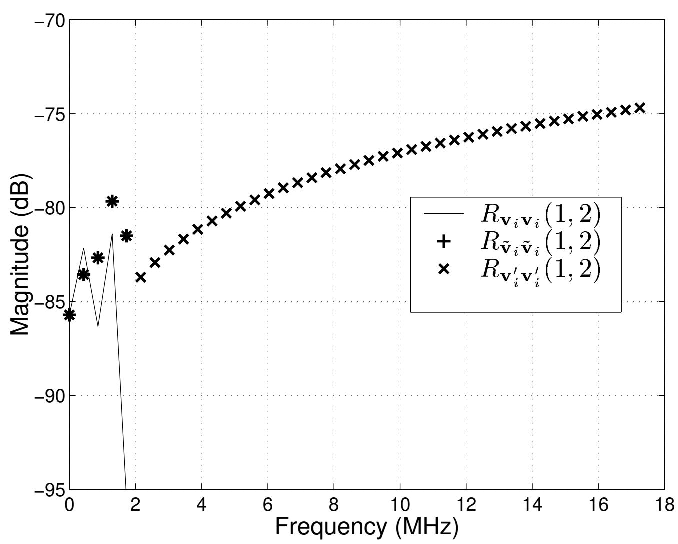

Figure 5.3: Magnitudes of elements $(1,2)$ of $R_{\mathbf{v}_i\mathbf{v}_i}$, $R_{\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i}$, and $R_{\mathbf{v}'_i\mathbf{v}'_i}$.

With the approximation of (5.12), Equation (5.11) gives

$$\sum_{i \in \mathcal{N}_d} \mathcal{E}_{i,j}^d \leq \mathcal{E}_j^d \ . \tag{5.13}$$

With this in mind, it is deduced that independent constraints apply for each user. Therefore, the energy allocation problem of (5.1) becomes independent for each user, and thus the a_k weights are irrelevant in this scenario. The optimization problem for each transmission direction is broken down into L waterfilling problems expressed by

$$\max \quad \sum_{i \in \mathcal{N}_u} \frac{1}{2} \log_2 \left(1 + \frac{s_{i,j}^u \mathcal{E}_{i,j}^u}{\Gamma} \right) \tag{5.14}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{N}_u} \mathcal{E}_{i,j}^u \leq \mathcal{E}_j^u \ , \tag{5.15}$$

and by

$$\max \quad \sum_{i \in \mathcal{N}_d} \frac{1}{2} \log_2 \left(1 + \frac{s_{i,j}^d \mathcal{E}_{i,j}^d}{\Gamma} \right) \tag{5.16}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{N}_d} \mathcal{E}_{i,j}^d \leq \mathcal{E}_j^d \ , \tag{5.17}$$

where $j = 1, \ldots, L$. Solutions to these problems are derived by the well-known waterfilling algorithm [19]. Simulation results indicating the performance benefits for VDSL are shown in Subsection 5.2.1.

5.1.2 Energy Allocation with Power Backoff

The premise of all previous derivations is that coordination is possible at the central node for the signals of a group of neighboring lines. This does not preclude the operation of other *alien* systems in neighboring lines, which on one hand cause crosstalk into the coordinated systems, and on the other hand suffer from crosstalk originating from the coordinated systems. This phenomenon is very visible in DSL, where the current approach in controlling it is to impose limits on the transmitted PSDs, so

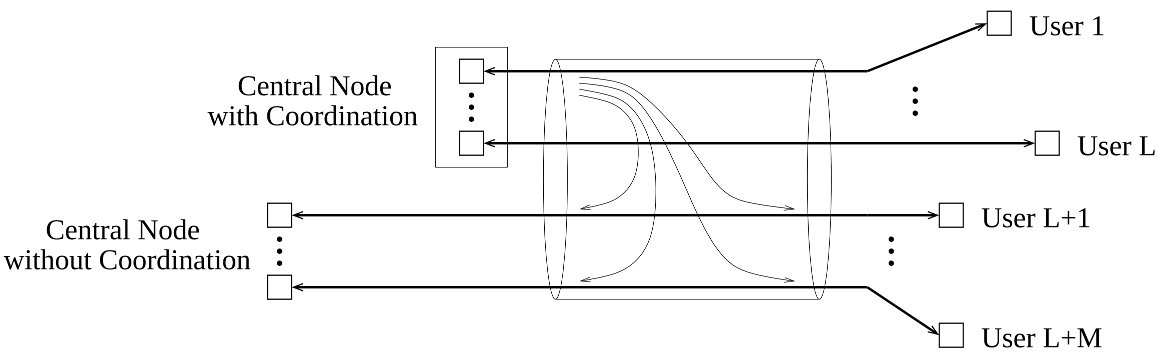

Figure 5.4: Topology with co-existence of coordinated and non-coordinated systems.

that the performance of systems is not excessively affected by crosstalk.

Additionally, a number of multi-line systems (e.g. VDSL) suffer from the fact that upstream signals on short lines detrimentally affect upstream performance on long lines (similarly to the near-far situation in wireless communications). In order to avoid imposing an overly restrictive universal PSD mask, power back-off methods have been proposed (see [71]), which effectively make the PSD mask dependent on the loop length of the specific line. An alternative approach to control this effect is the concept of iterative waterfilling [67, 68].

A DSL-inspired scenario, where the downstream communication of neighboring systems may suffer considerably is shown in Figure 5.4. This situation will occur increasingly often as the installation of ONUs progresses, while twisted pair connections to the COs remain (see Section 1.1).

This section examines the problem of transmission optimization in coordinated systems with the additional constraint of protecting the non-coordinated neighboring systems. The resulting method differs significantly from previous approaches, which either adopted a universal PSD mask, or used power back-off methods relying only on loop length information.

Equation (2.12) is augmented to include the received samples of alien systems,

$$\begin{bmatrix} \mathbf{z}' \\ \mathbf{z}^a \end{bmatrix} = \begin{bmatrix} H_F & C_F^a \\ C_F & H_F^a \end{bmatrix} \begin{bmatrix} \mathbf{v}' \\ \mathbf{v}^a \end{bmatrix} + \begin{bmatrix} \mathbf{n}'_z \\ \mathbf{n}_z^a \end{bmatrix} , \tag{5.18}$$

where $\mathbf{z}^a$, $\mathbf{v}^a$, and $\mathbf{n}_z^a$ are vectors of the received samples, of the transmitted symbols and of the noise samples respectively of the alien systems. The definitions of the block matrices C_F^a, C_F and H_F^a depend on both the channel and the characteristics of the alien systems, and although H_F is block diagonal (see (2.16)), this property does not generally hold for the other matrices.

In the following, it is assumed that H_F, H_F^a, C_F, and C_F^a are known both at the central node and at the remote nodes of the coordinated systems. In practice, the identification of C_F and C_F^a may present difficulties, such as in the DSL case where the coordinated central node is controlled by an entity that is different from the one that controls the non-coordinated central node. Reference [72] approaches this problem by introducing a "third party" that is allowed to access received data from both the coordinated and the non-coordinated central node. An alternative solution is to perform channel identification by applying the expectation-maximization (EM) algorithm in the context of a multiple-input single-output system [73].

With this in mind, a criterion is devised to mitigate the effect of crosstalk into the alien systems. Limiting the FEXT induced on the alien systems in the mean square sense results in the following constraints for the upstream and downstream energy allocation,

$$\sum_{j=1}^{L} \sum_{i \in \mathcal{N}_u} |c_{k,(j-1)N+i}|^2 \mathcal{E}_{i,j}^u \leq \mathcal{E}_k^{u,c} , \; k = 1, \ldots, MN^a \tag{5.19}$$

$$\sum_{j=1}^{L} \sum_{i \in \mathcal{N}_d} |c_{k,(j-1)N+i}|^2 \mathcal{E}_{i,j}^d \leq \mathcal{E}_k^{d,c} , \; k = 1, \ldots, MN^a , \tag{5.20}$$

where M is the number of neighboring systems, N^a is the number of "dimensions" (e.g. number of tones) per neighboring system, $\mathcal{E}_k^{u,c}$, $\mathcal{E}_k^{d,c}$ are the maximum allowable crosstalk energies in sample k of the neighboring systems for upstream and downstream, and $c_{m,n}$ is the (m,n) element of the $MN^a \times LN$ matrix C_F. Note that this approach can be generalized, so that both the induced FEXT and NEXT are restricted.

The set of inequalities in (5.19) and (5.20), combined correspondingly with those of (5.9) and (5.13), form a set of linear inequality constraints. Including the rate maximization objective of (5.1) yields the optimization problems

$$\max \quad \sum_{j=1}^{L} a_j \sum_{i \in \mathcal{N}_u} \frac{1}{2} \log_2 \left(1 + \frac{s_{i,j}^u \mathcal{E}_{i,j}^u}{\Gamma}\right) \tag{5.21}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{N}_u} \mathcal{E}_{i,j}^u \leq \mathcal{E}_j^u \ , \ j = 1, \ldots, L \ , \tag{5.22}$$

$$\sum_{j=1}^{L} \sum_{i \in \mathcal{N}_u} |c_{k,(j-1)N+i}|^2 \mathcal{E}_{i,j}^u \leq \mathcal{E}_k^{u,c} \ , \ k = 1, \ldots, MN^a \ , \tag{5.23}$$

and

$$\max \quad \sum_{j=1}^{L} a_j \sum_{i \in \mathcal{N}_d} \frac{1}{2} \log_2 \left(1 + \frac{s_{i,j}^d \mathcal{E}_{i,j}^d}{\Gamma}\right) \tag{5.24}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{N}_d} \mathcal{E}_{i,j}^d \leq \mathcal{E}_j^d \ , \ j = 1, \ldots, L \ , \tag{5.25}$$

$$\sum_{j=1}^{L} \sum_{i \in \mathcal{N}_d} |c_{k,(j-1)N+i}|^2 \mathcal{E}_{i,j}^d \leq \mathcal{E}_k^{d,c} \ , \ k = 1, \ldots, MN^a \ . \tag{5.26}$$

The objective functions (5.21) and (5.24) are concave, since they are sums of log functions, which are known to be concave. The constraints (5.22), (5.23), (5.25), and (5.26) define linear inequalities, therefore, they form convex sets. Thus, the problem is characterized as convex and solutions can be efficiently produced using convex programming techniques. Such an example is given in Subsection 5.2.3.

Extra restrictions (such as PSD masks or limits on the number of bits per tone) can easily be included in the above optimization problems, since they only require the introduction of additional linear inequality constraints.

5.1.3 Joint Energy Allocation and Frequency Planning

The previous subsections assumed that the FDD plan is fixed, in other words $\mathcal{N}_u$ and $\mathcal{N}_d$ were pre-determined. However, a *dynamically configured band plan* offers significant advantages in certain circumstances. Such a plan is common for all users, but is determined during modem initialization (or re-training), depending on the specific transmission environment, as well as on the user requirements.

The disadvantages of a fixed band plan are exemplified in DSL in the presence of bridged taps. It is then likely that deep "notches" in the channel frequency response due to bridged taps may lead to transmission in one direction facing a disproportionate degradation. On the other hand, adopting a dynamic plan in such a case enables a fairer distribution of the impact of the notches on both upstream and downstream. Additionally, a dynamic band plan gives extra flexibility by allowing the provisioning of either symmetric or asymmetric rates between upstream and downstream.

The optimization objective is now expressed by

$$\max \sum_{j=1}^{L} \left(a_j^u r_j^u + a_j^d r_j^d \right) \ , \tag{5.27}$$

where $a_j^u, a_j^d \geq 0$ are the weights assigned to upstream and downstream transmission for user j, and r_j^u, r_j^d are the achievable upstream and downstream rates of user j. The optimization parameters involve not just the energies assigned but also the selection of upstream and downstream tones. If (5.5) and (5.6) are employed for the computation of the user rates, then the problem requires the partition of the set of tones into $\mathcal{N}_u$ and $\mathcal{N}_d$. This implies a *binary constraint*, where each tone must be assigned to one of two choices, upstream or downstream. Then, the optimum solution requires searching over the exponentially large set of possible tone assignments. Unfortunately, even with a moderately large number of tones, such a solution has a prohibitively high complexity.

If instead the binary constraints are relaxed, the computations are greatly simplified. This idea has previously been exploited in similar problems: in [74] for subcarrier allocation in multiuser Orthogonal-Frequency-Division-Multiplexing, in [75] for the computation of the FDMA capacity of the Gaussian multiple access channel in the presence of ISI, and in [76] for the computation of the FDMA capacity of the broadcast channel.

Initially, it is assumed that each tone is time-shared between upstream and downstream, thus obtaining the following achievable rates

$$r_j^u = \sum_{i=1}^{N} t_i^u \frac{1}{2} \log_2 \left(1 + \frac{s_{i,j}^u \mathcal{E}_{i,j}^u}{t_i^u \Gamma}\right) , \tag{5.28}$$

$$r_j^d = \sum_{i=1}^{N} t_i^d \frac{1}{2} \log_2 \left(1 + \frac{S_{i,j}^d \mathcal{E}_{i,j}^d}{t_i^d \Gamma}\right) , \tag{5.29}$$

where t_i^u, t_i^d define the fraction of time in tone i used for upstream and downstream transmission correspondingly, and $t_i^u + t_i^d = 1$, $t_i^u, t_i^d \geq 0$. The existence of t_i^u and t_i^d in the denominators inside the log expressions implies that the assigned energy is "boosted", since transmission takes place over only some fraction of time. The energy constraints for user j are

$$\sum_{i=1}^{N} \mathcal{E}_{i,j}^u \leq \mathcal{E}_j^u , \tag{5.30}$$

$$\sum_{i=1}^{N} \mathcal{E}_{i,j}^d \leq \mathcal{E}_j^d . \tag{5.31}$$

Therefore, the optimization problem has the following form:

$$\begin{aligned}
\max \quad & \sum_{j=1}^{L}\left[a_j^u \sum_{i=1}^{N} t_i^u \frac{1}{2}\log_2\left(1+\frac{s_{i,j}^u \mathcal{E}_{i,j}^u}{t_i^u \Gamma}\right)+\right. \\
& \left. a_j^d \sum_{i=1}^{N} t_i^d \frac{1}{2}\log_2\left(1+\frac{s_{i,j}^d \mathcal{E}_{i,j}^d}{t_i^d \Gamma}\right)\right] && (5.32)\\
\text{subject to} \quad & \sum_{i=1}^{N} \mathcal{E}_{i,j}^u \le \mathcal{E}_j^u \ , \ j=1,\ldots,L \ , && (5.33)\\
& \sum_{i=1}^{N} \mathcal{E}_{i,j}^d \le \mathcal{E}_j^d \ , \ j=1,\ldots,L \ , && (5.34)\\
& t_i^u + t_i^d = 1 \ , \ i=1,\ldots,N \ . && (5.35)
\end{aligned}$$

The objective function in (5.32) is concave, because it is a sum of functions of the form $x\log(1+y/x)$, which are convex for $x, y \ge 0$ (see [75]). The constraint sets are clearly convex, since they are defined by the linear inequalities (5.33), (5.34) and by the hyperplane of (5.35). Therefore, the problem is characterized as convex and a variety of methods can be used efficiently to obtain a solution.

Such a solution actually yields a hybrid between an FDD and a Time Division Duplexing (TDD) implementation. Since an FDD implementation is required, an approximate solution is obtained by rounding t_i^u and t_i^d. Naturally, this is suboptimal, but when the number of tones is fairly large, it has been observed (see Subsection 5.2.4 for simulation results) that the optimal t_i^u and t_i^d are equal to 0 or 1 for almost all tones. Thus, the approximate solution is adequately close to the optimal one.

5.2 Simulation Results for VDSL

The simulation results presented next refer to the VDSL environment. The general parameters and the DMT specifications are shown in Table 5.1, and they are taken from [4, 77]. The specified coding gain, noise margin and probability of error target lead to an SNR-gap of 12dB. In practice, DMT systems have a maximum constellation

Number of DMT Tones	4096
Tone Width	4.3125 KHz
Symbol Rate	4 KHz
Coding Gain	3.5 dB
Noise Margin	6 dB
Symbol Error Probability	$< 10^{-7}$
Maximum Power	14.5 dBmW
Cable Type	26-Gauge (0.4 mm)
Source/Load Resistance	100 Ohm
Amateur Radio Bands	as specified in [4]

Table 5.1: Simulation parameters.

Upstream Bands (MHz)	Downstream Bands (MHz)
0.03 – 0.138	0.138 – 2.5
2.5 – 3.75	3.75 – 7.5
7.5 – 14.5	14.5 – 17.7

Table 5.2: Upstream-downstream frequency assignment.

size for transmission within a tone. In the following, no such limit is taken into account, since the aim is to demonstrate the achievable upper bounds.

5.2.1 Energy Allocation and Channel Estimation Errors

The energy allocation method of Subsection 5.1.1 amounts to performing water-filling independently for each user. The upstream/downstream frequency division band plan is chosen as in Table 5.2, and the noise models used are specified in [4]. A loop topology with 20 users is assumed, with all loops having the same length. Figures 5.5 and 5.6 compare the achievable data rates with and without coordination versus the loop length for upstream and downstream transmission respectively. Plots are shown assuming either noise A or noise F. Although no PSD constraint is enforced, the resulting PSDs were observed to be typically below -52dBmW/Hz, and the peak never exceeded the value of -45dBmW/Hz.

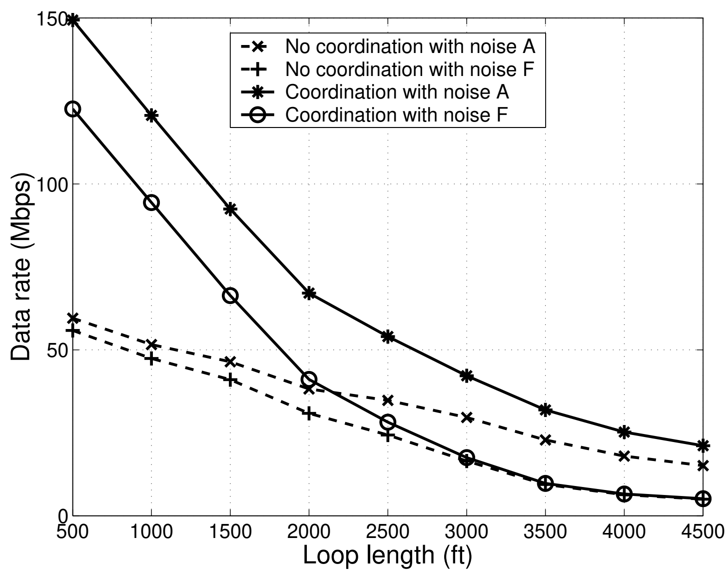

Figure 5.5: Upstream VDSL data rates.

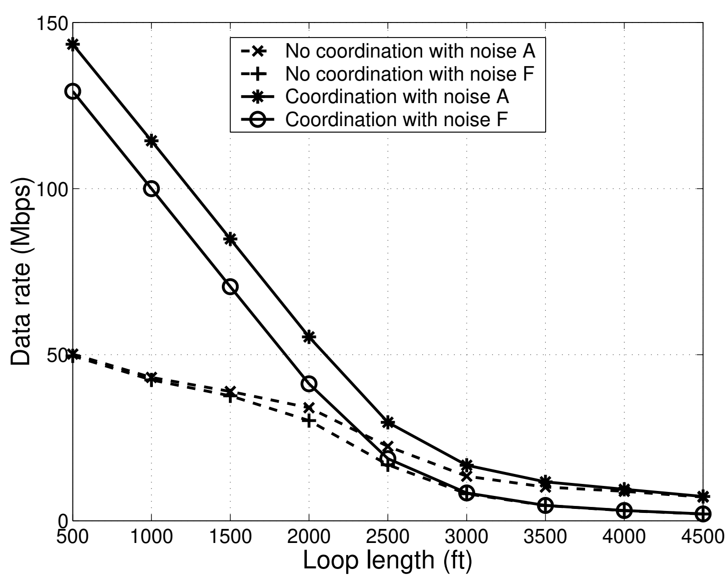

Figure 5.6: Downstream VDSL data rates.

Coordination clearly improves performance for short loops, and this can be viewed in two ways: For a given loop length, coordination allows the achievement of much higher data rates. The rate increases are considerable for lengths shorter than 4500ft in noise A, or 2500ft in noise F. The gains become spectacular in short loops, where transmission is FEXT-limited. Also, it is seen that coordination extends the maximum loop length given a data rate requirement. As an example, in Figure 5.6, a downstream rate requirement of 50Mbps limits non-coordinated systems with noise of type A to loops shorter than 1150ft, while coordination extends the reach to 2650ft.

Next, the impact of channel estimation errors in a worst-case scenario is assessed. Equations (4.74) and (4.75) indicate that the effects of such errors are residual crosstalk, which may be treated as noise, and detection biasing, which can be modeled as a decrease in the minimum distance of the constellation points (e.g. see [78]). Here, the effect of residual crosstalk is examined, and simulation results are shown, where the energy allocation algorithm accounts for estimation errors by considering them as extra noise. By making the approximations $Q_i^u \approx I$, $Q_i^d \approx I$, and that E_i^u, E_i^d (see (4.72, 4.73)) have zero diagonal elements, the covariance matrices of the extra noise term in tone i are

$$R^u_{\mathbf{n}_e\mathbf{n}_e} = \left(\hat{S}^u_{o,i}\right)^{-\frac{1}{2}} E_i^u R_{\mathbf{v}_i\mathbf{v}_i} \left(E_i^u\right)^* \left(\hat{S}^u_{o,i}\right)^{-\frac{1}{2}} , \tag{5.36}$$

$$R^d_{\mathbf{n}_e\mathbf{n}_e} = \left(\hat{S}^d_{o,i}\right)^{-\frac{1}{2}} E_i^d R_{\mathbf{v}'_i\mathbf{v}'_i} \left(E_i^d\right)^* \left(\hat{S}^d_{o,i}\right)^{-\frac{1}{2}} , \tag{5.37}$$

for upstream and downstream correspondingly. Assume a worst-case situation, where E_i^u, E_i^d have all their off-diagonal elements equal to ε and the variance of $\mathbf{v}_i$ and $\mathbf{v}'_i$ is σ_u^2. With the additional approximation of uncorrelated components of $\mathbf{v}'_i$, it is found that the energies of the noise contributions due to channel estimation errors in tone i for user j are

$$\left(\sigma^u_{i,j}\right)^2 = (L-1)\left(\hat{s}^u_{i,j}\right)^{-1} |\varepsilon|^2 \sigma_u^2 , \tag{5.38}$$

$$\left(\sigma^d_{i,j}\right)^2 = (L-1)\left(\hat{s}^d_{i,j}\right)^{-1} |\varepsilon|^2 \sigma_u^2 , \tag{5.39}$$

for upstream and downstream transmission correspondingly. Also, $\hat{s}^u_{i,j}$, $\hat{s}^d_{i,j}$, $j =$

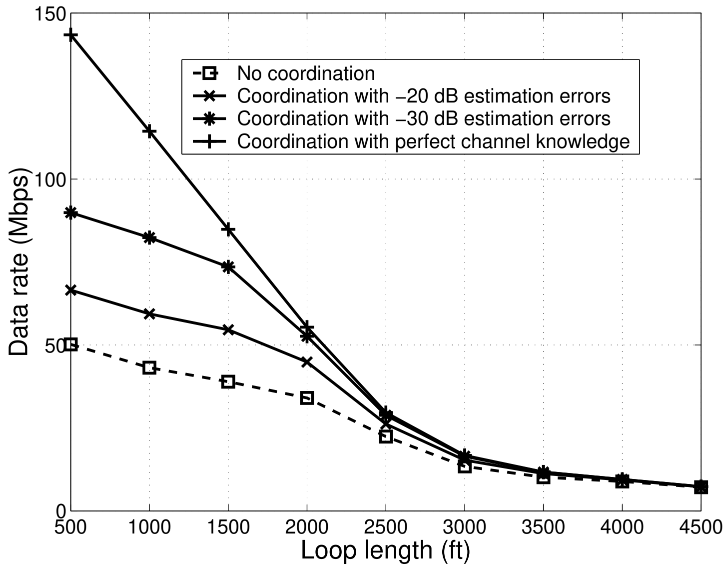

Figure 5.7: Upstream VDSL data rates (noise A) with channel estimation errors.

$1, \ldots, L$ are the diagonal elements of $\hat{S}^u_{o,i}$ and $\hat{S}^d_{o,i}$ correspondingly.

Making the pessimistic assumption that $\sigma_u^2 = -50$dBmW, Figures 5.7 and 5.8 show the impact of channel estimation errors on the achievable upstream and downstream data rates for the noise A environment. The curves are parameterized by the *relative* estimation error $|\varepsilon|/\sqrt{s_{i,j}}$, and it is seen that as channel identification becomes less accurate, the performance advantage of the coordinated system diminishes significantly for the shorter loops.

5.2.2 Comparison with Information Theory Bounds

The rate bounds defined by (2.29) and (2.30) are next compared against the achievable rates of the proposed coordination techniques in a realistic DSL scenario. It is assumed that there are only 2 users, where line 1 is 1500ft long and line 2 is 2000ft long. The SNR gap of 12dB applies for all computed rates, including the bounds, and noise model F is considered. Also, the band plan is as in Table 5.2.

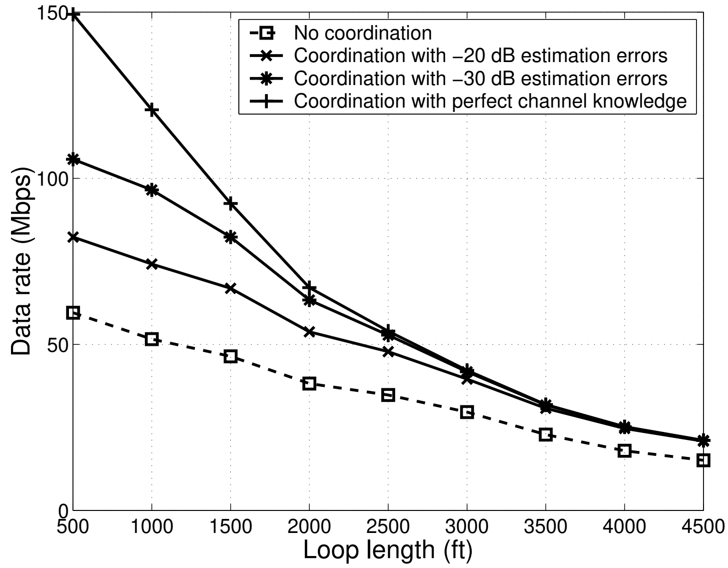

Figure 5.8: Downstream VDSL data rates (noise A) with channel estimation errors.

Figure 5.9 shows the upstream rate regions and theoretical bounds. As expected, the rate improvement from crosstalk cancellation is significant, especially for user 2. But additionally, the rates achieved with crosstalk cancellation almost coincide with the single user and vector channel bounds of the multiple access channel. Figure 5.10 shows the corresponding downstream rate regions and theoretical bounds. Again, the rate increase due to crosstalk mitigation is large, and additionally, the sum rate is very close to the bound obtained from the vector channel. The user rates are smaller than the single user bounds, but that can be justified by the fact that individual energy contraints apply for each coordinated user, whereas (2.30) employs a total energy constraint applying to all users.

Evidently, the rate region bounds are tight. This observation holds for all multi-line channels, and is due to the diagonally dominant structure of the channel matrices. The crosstalk cancellation methods employing the QR factorization are capable of approaching these bounds very closely (the difference is less than 1% in the previous VDSL example).

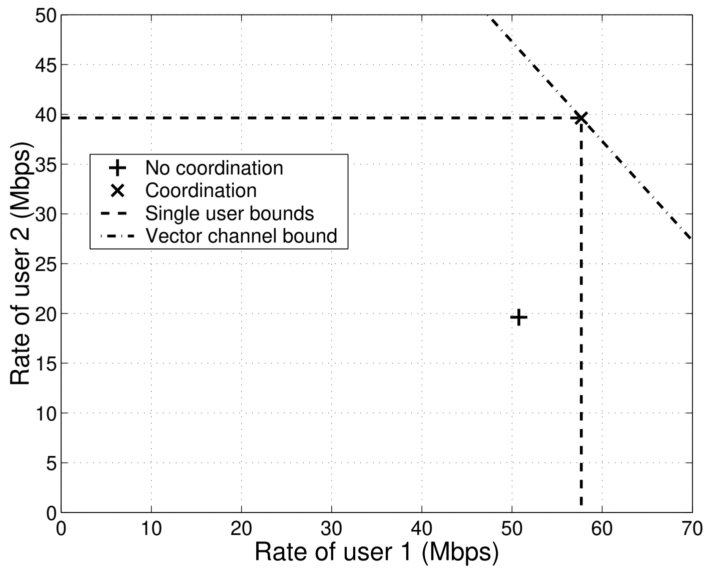

Figure 5.9: Comparison of upstream rate regions.

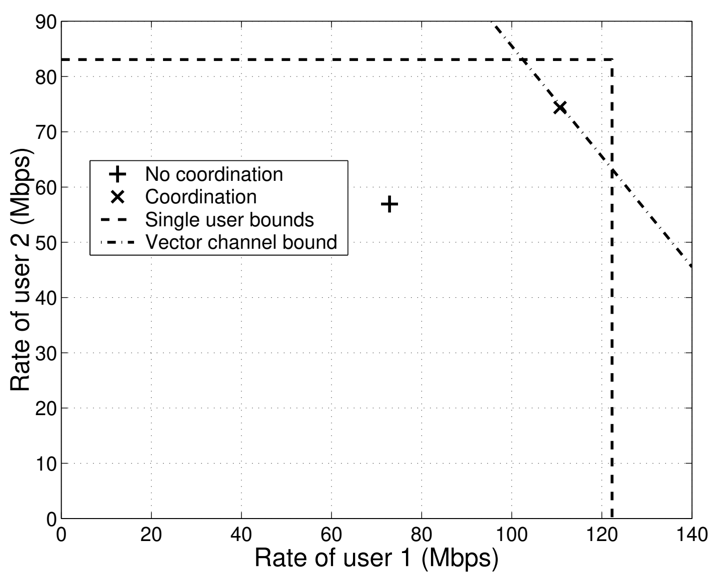

Figure 5.10: Comparison of downstream rate regions.

5.2.3 Energy Allocation with Power Back-Off

Another VDSL example is given here, where energy allocation is further constrained by the requirement that some "alien" DSL system must also be protected against crosstalk from the coordinated system. Assume that there are two coordinated users and a single affected CO-based system as in Figure 5.4, where downstream transmission from the CO is severely impaired. The loop lengths of the coordinated users are 2500ft and 2000ft, the loop length of the non-coordinated user is 3000ft, and the crosstalk coupling length between the coordinated and the alien systems is 1500ft.

Assuming the existence of a VDSL system on the affected line, a possible strategy to mitigate the impact of crosstalk is to postulate that the induced crosstalk must not exceed the crosstalk PSD caused by a certain number of VDSL disturbers with loop lengths equal to the length of the affected line. (This is analogous to the equalized FEXT power back-off method described in [71]). Employing the FEXT model of [18], (5.20) must then be satisfied with

$$\mathcal{E}_k^{d,c} = PSD(f_k) k_{FEXT} l f_k^2 |H(f_k)|^2 \left(\frac{N_d}{49}\right)^{0.6}, \; k = 1, \ldots, N \;, \tag{5.40}$$

where f_k is the center frequency of tone k, $H(f)$ and l are the transfer function and the length of the affected line (in feet), $PSD(f)$ is the VDSL PSD mask, N_d is the number of disturbers, and k_{FEXT} is a constant equal to $7.999 \cdot 10^{-20}$. The coefficients $c_{j,l}$ of (5.20) are computed by making the approximation that signals transmitted in a tone do not cause crosstalk in neighboring tones, and $N_d = 20$ is chosen.

Figure 5.11 shows the achievable downstream rate curve, where the different points are obtained by adjusting the weights in (5.1). User 1 corresponds to the 2500ft loop, and user 2 corresponds to the 2000ft loop. Noise A is assumed and the transmission parameters are set as in the previous subsection. The dashed curve indicates the achievable rate bounds in the absence of crosstalk constraints. The solid curve reveals the possible trade-off between the 2 users for meeting the requirement to protect the alien system.

The numerical results of this and the following subsection were obtained by using

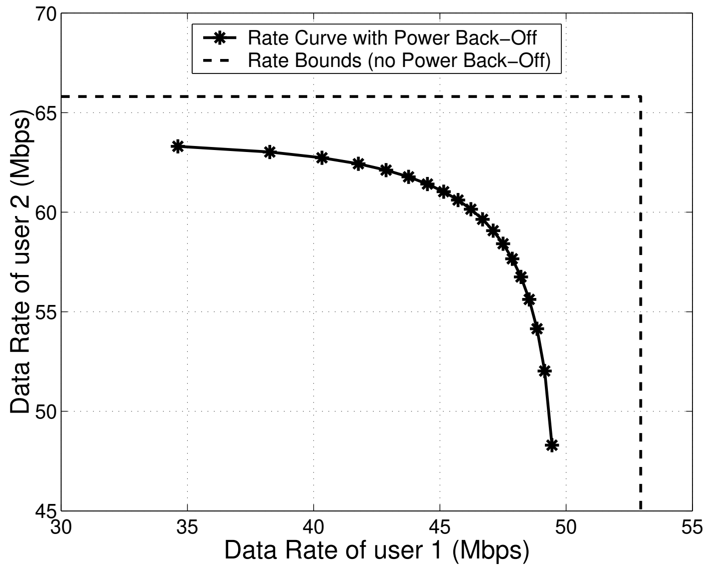

Figure 5.11: Rate region for 2 users with power back-off.

Maximum number of superbasics	1000
Feasibility tolerance	10^{-14}
Optimality tolerance	10^{-20}

Table 5.3: MINOS parameters for power back-off optimization.

the AMPL mathematical programming language [79], together with the nonlinear programming package MINOS [80]. The MINOS package is suitable for solving large-scale optimization problems, and is effective with both linear and non-linear programs. The optimization problems of Subsections 5.1.2 and 5.1.3 involve non-linear objectives and linear constraints. The method employed by MINOS in such cases is a reduced-gradient method with quasi-Newton approximations to the reduced Hessian. Table 5.3 summarizes the MINOS parameters used to obtain the plot of Figure 5.11.

Maximum number of superbasics	1000
Feasibility tolerance	10^{-14}
Optimality tolerance	10^{-8}
Minor damping parameter	10^{-4}

Table 5.4: MINOS parameters for joint energy allocation and frequency planning.

5.2.4 Joint Energy Allocation and Frequency Planning

Finally, it is assumed that the frequency plan for upstream/downstream duplex is programmable according to the loop characteristics and the service requirements. Results for two users are shown, where user 1 and user 2 have loop lengths of 1500ft and 3000ft correspondingly. Noise A is considered, and all transmission parameters are the same as in the previous subsection, with the difference being that the number of tones is set to 128 in order to reduce the computational burden. Additionally, radio frequency ingress in the amateur radio bands is ignored, thus allowing transmission over the whole available bandwidth. The MINOS package is again used with the parameters shown in Table 5.4.

By having $\alpha_1^u = \alpha_2^u$, $\alpha_1^d = \alpha_2^d$, and adjusting the upstream/downstream weights, the rate curves of Figure 5.12 are produced. The rates achieved with the currently proposed VDSL band plans 997 and 998 (see [4]) are also shown. The optimal frequency allocation for almost symmetric transmission is shown in Figure 5.13. The corresponding achievable rates are $r_1^u = 89.7$Mbps, $r_1^d = 90.8$Mbps, $r_2^u = 32.0$Mbps, and $r_2^d = 29.3$Mbps, which are obtained with $\alpha_1^u = \alpha_2^u = 0.4998$.

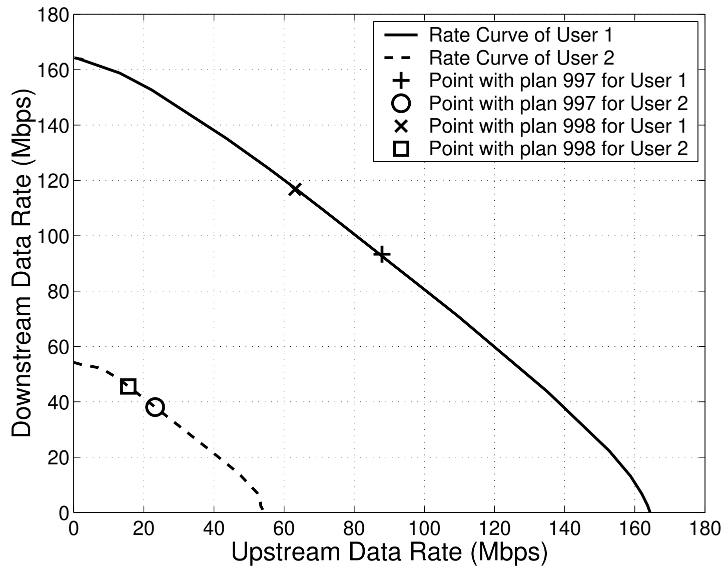

Figure 5.12: Upstream/downstream rate regions for 2 users with optimum frequency duplexing.

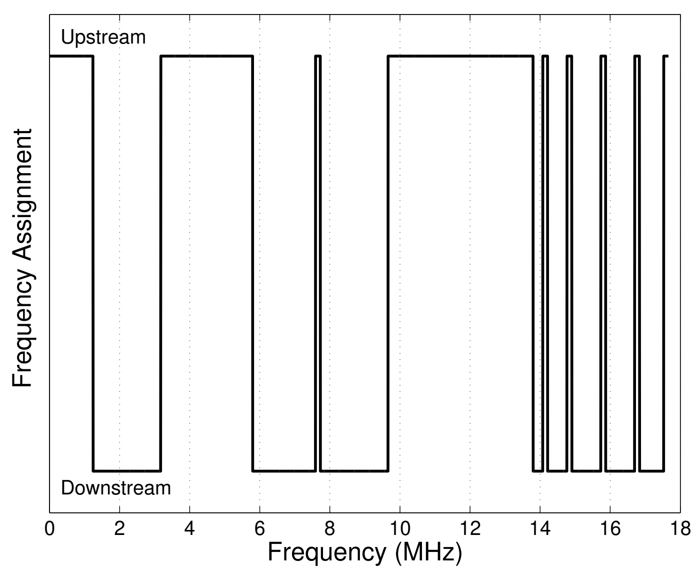

Figure 5.13: Optimum frequency duplexing plan with symmetric service for upstream and downstream.

Chapter 6

Adaptive Technique for Upstream Coordinated Communication

THE PREVIOUS CHAPTERS MADE THE ASSUMPTION that the communication systems are able to identify the multi-line channel, and that this information is available at both the transmitters and the receivers. A standard approach for channel identification is to transmit during modem initialization a *training sequence*, which is known a priori to both the transmitters and the receivers. However, if the channel changes over time, *adaptive* solutions need to be adopted, otherwise training causes a large loss in the effective data rate. In other scenarios, a training sequence may be unavailable, in which case *blind* techniques must be employed.

This chapter derives an adaptive implementation of the receiver for upstream communication with crosstalk cancellation. A significant number of research results have already been produced for adaptive Multiple-Input Multiple-Output (MIMO) Decision Feedback Equalization (DFE). Reference [81] proposed an adaptive decorrelating detector with decision feedback for Code-Division-Multiple-Access (CDMA) based on a least squares formulation. Then, [82] used a stochastic gradient algorithm to adapt the filter coefficients of a CDMA DFE receiver. An adaptive method for a CDMA DFE receiver was presented in [83], together with a blind realization when knowledge of signature sequences and timing is available. Also, the MIMO DFE adaptation problem is solved in [84] by utilizing an appropriate state-space model.

This chapter describes a novel adaptive method for MIMO DFE equalization (see also [85]). In contrast with previously proposed techniques, this method makes no use of training, and requires no knowledge about the channel or its structure. Additionally, it is applicable to multi-line upstream systems. Section 6.1 describes the technique for receiver adaptation based on the decomposition of a unitary matrix into Givens rotation terms, and motivated by the concept of a *bootstrap detector* [86]. Simulation results are included in Section 6.2 demonstrating the performance of the method.

6.1 Blind Crosstalk Cancellation for Upstream using Givens Rotations

6.1.1 Crosstalk Cancellation for Upstream

Crosstalk cancellation with upstream multi-carrier communication in a multi-line channel was examined in Subsection 4.1.2. As seen in Figure 4.1, the transmission filter is $F = Q_{mIDFT}P$. The receiver filtering consists of two parts. The section PQ_{mDFT} performs DFT and reordering, as in (4.12),

$$\mathbf{z}'_i = H_{F,i}\mathbf{v}_i + \mathbf{n}'_{z,i}\ ,\ i = 1, \ldots, N\ , \tag{6.1}$$

where $R_{\mathbf{n}'_{z,i}\mathbf{n}'_{z,i}} = I_L$. This is followed by rotation and scaling applied for each tone, as in (4.13),

$$\tilde{\mathbf{z}}_i = S_{o,i}^{-\frac{1}{2}} Q_i^* \mathbf{z}'_i\ , \tag{6.2}$$

$$\tilde{\mathbf{z}}_i = S_{o,i}^{-\frac{1}{2}} R_i \mathbf{v}_i + \tilde{\mathbf{n}}_{z,i}\ , \tag{6.3}$$

where $R_{\tilde{\mathbf{n}}_i\tilde{\mathbf{n}}_i} = S_{o,i}^{-1}$. Observe that the rotation causes $\tilde{\mathbf{z}}_i$, $\mathbf{v}_i$ to be related by the *triangular* matrix $S_{o,i}^{-\frac{1}{2}} R_i$. Decision feedback is employed and the decoder output is

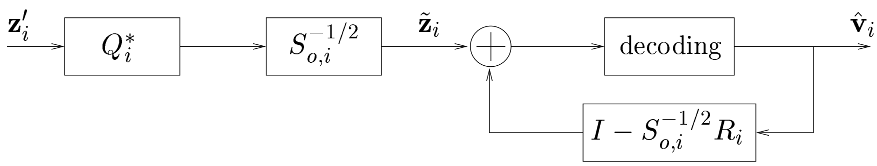

Figure 6.1: Crosstalk canceler for tone i.

expressed as in (4.16),

$$\hat{\mathbf{v}}_i = \text{decode}\left[\tilde{\mathbf{z}}_i + (I - S_{o,i}^{-\frac{1}{2}} R_i)\hat{\mathbf{v}}_i\right] \tag{6.4}$$

$$= \text{decode}\left(\mathbf{v}_i + \tilde{\mathbf{n}}_{z,i}\right) \ , \ i = 1, \ldots, N \ . \tag{6.5}$$

The equalization structure is shown in Figure 6.1. The transmitter is independent of the channel, and the only receiver parameters that depend on the channel are Q_i, and R_i, while $S_{o,i}$ is diagonal and composed of the squared diagonal elements of R_i. The next subsection shows that Q_i can be expressed as the multiplication of multiple terms, where each term is characterized by a single parameter.

In the following, the i subscript is dropped and transmission within each tone is considered. For notational convenience, $D = S_{o,i}^{-\frac{1}{2}}$ is defined. Additionally, all quantities are assumed real in order to allow for intuitive geometrical interpretations. If solutions with complex quantities are desired, a possible approach is to represent complex matrices and vectors as real matrices and vectors with twice as many dimensions.

6.1.2 Givens Rotations and QR Factorization

This section reviews the application of Givens rotations for the derivation of the QR factorization [64, Chapter 5]. A Givens rotation is an orthogonal matrix of the form

$$G(i,j,\theta) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & \cos(\theta) & \dots & \sin(\theta) & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & -\sin(\theta) & \dots & \cos(\theta) & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}, \tag{6.6}$$

where all off-diagonal entries are 0 except for (i,j) and (j,i) that equal $\sin(\theta)$ and $-\sin(\theta)$ respectively, and all diagonal entries are 1 except for (i,i) and (j,j) that are both equal to $\cos(\theta)$.

When a vector is multiplied by G^T, then only elements i and j are affected. The projection of the vector on the two-dimensional plane defined by the i and j coordinates is rotated in the counter-clockwise direction by θ radians. If the original vector entries are $(\dots, x_i, \dots, x_j, \dots)$, then the choice $\theta = \tan^{-1}(-x_j/x_i)$ zeros the j-th coordinate.

Thus, left multiplication of the matrix H_F by G^T can be used to zero an element of row j using the element of the same column at row i as a “pivot”. The application of this procedure for all lower triangular entries of H_F can be used to obtain its QR decomposition. An example of the method for a 3×3 case is shown:

$$\begin{bmatrix} X & X & X \\ X & X & X \\ X & X & X \end{bmatrix} \xrightarrow{G_1^T(1,2,\theta_1)} \begin{bmatrix} X & X & X \\ 0 & X & X \\ X & X & X \end{bmatrix} \xrightarrow{G_2^T(1,3,\theta_2)}$$
$$\begin{bmatrix} X & X & X \\ 0 & X & X \\ 0 & X & X \end{bmatrix} \xrightarrow{G_3^T(2,3,\theta_3)} \begin{bmatrix} X & X & X \\ 0 & X & X \\ 0 & 0 & X \end{bmatrix} . \tag{6.7}$$

With H_F being an $L \times L$ matrix, at most $L(L-1)/2$ Givens rotations are needed to zero all its lower triangular entries, so that

$$G^T_{N(N-1)/2} \ldots G_2^T G_1^T H_F = R \ , \tag{6.8}$$

where G_k are Givens rotations and R is upper triangular. Therefore, Q can be expressed as

$$Q = G_1 G_2 \ldots G_{N(N-1)/2} \ , \tag{6.9}$$

and thus, the QR decomposition is obtained. Note that, when the matrix is sparse (as described in Subsection 4.2.3), then Q is expressed with fewer than $L(L-1)/2$ Givens rotations.

6.1.3 Blind Algorithm Description

A *blind* method to compute the Givens rotations is described next. Notice that an alternative definition of the triangularization condition of (6.3) is the following set of $L(L-1)/2$ equations,

$$E\left(\tilde{z}_m v_n\right) = 0 \ , \ m = 2, \ldots, L \ , \ n = 1, \ldots, m-1 \ , \tag{6.10}$$

where $\tilde{z}_m$, v_n are the elements of $\mathbf{z}$ and $\mathbf{n}$. If one defines

$$\tilde{\mathbf{z}}^{(k)} = G_k^T \dots G_2^T G_1^T \mathbf{z}' \ , \tag{6.11}$$

then (6.10) can be replaced by

$$E\left(\tilde{z}_m^{(k)} v_n\right) = 0 \ , \ m = 2, \dots, L \ , \ n = 1, \dots, m-1 \ , \tag{6.12}$$

where $\tilde{z}_m^{(k)}$ are the elements of $\tilde{\mathbf{z}}^{(k)}$, and it is assumed that the k-th rotation zeros element (m, n) of $G_{k-1}^T \dots G_2^T G_1^T H_F$. Using (6.7) as an example, the requirement $E(\tilde{z}_2^{(1)} x_1) = 0$ determines the parameter θ_1 of G_1. Similarly, $E(\tilde{z}_3^{(2)} x_1) = 0$ determines θ_2 of G_2, and $E(\tilde{z}_3^{(3)} x_2) = 0$ determines θ_3 of G_3.

Assuming that the decoded outputs of the equalizer are reliable, (6.12) can be replaced by

$$E\left(\tilde{z}_m^{(k)} \hat{v}_n\right) = 0 \ , \ m = 2, \dots, L \ , \ n = 1, \dots, m-1 \ , \tag{6.13}$$

where $\hat{v}_n$ are the elements of $\hat{\mathbf{v}}$. Motivated by the idea of a *bootstrap detector* [86], the following adaptive algorithm is proposed for estimating θ_k,

$$\bar{\theta}_k \leftarrow \bar{\theta}_k - \mu_k \tilde{z}_m^{(k)} \mathrm{sign}\left(\hat{v}_n\right) \ , \tag{6.14}$$

where $\bar{\theta}_k$ is the estimate of θ_k, $\mu_k > 0$ is an appropriately chosen step-size, and $k = 1, \dots, L(L-1)/2$. When $\tilde{z}_m^{(k)}$ is uncorrelated with $\hat{v}_n$, then $\bar{\theta}_k$ remains unchanged on average. If $\tilde{z}_m^{(k)}$ is positively correlated with $\hat{v}_n$, then $\bar{\theta}_k$ on average decreases in order to eliminate the correlation, and if the correlation is negative, then $\bar{\theta}_k$ increases. For adequately small noise and interference, this method allows convergence to the exact solution of (6.12). Previous work [87] has shown that the bootstrap approach is fairly robust even in environments with relatively large noise. This has also been demonstrated here.

Evidently, (6.14) defines a blind method for the determination of Q in the QR decomposition of H_F. Note that the use of (6.12) to determine θ_k assumes that all angles $\theta_1, \dots, \theta_{k-1}$ have been accurately estimated. This implies for (6.14) that for any k, convergence for $\bar{\theta}_k$ requires a satisfactory level of convergence for $\bar{\theta}_{k-1}$. In practice,

$\bar{\theta}_{k-1}$ contains some error, however, this error is not detrimental to the convergence of the algorithm. Simulations indicate that the convergence time increases linearly with L.

The need to estimate R blindly is handled next. Assuming that Q has achieved convergence, one finds that

$$\tilde{\mathbf{z}}^{(L(L-1)/2)} = R\mathbf{v} + Q^T \mathbf{n}'_z \ . \tag{6.15}$$

In order to perform decision feedback detection, the strictly upper triangular entries of R must be estimated. Let $\check{\mathbf{z}}$ (with elements $\check{z}_i$) be the vector exactly before the decoding device, which ideally must contain no interference. The following set of equations determines $r_{mn}\ ,\ n > m$:

$$E\left(\check{z}_m v_n\right) = 0 \ ,\ m = 1, \ldots, L \ ,\ n = m+1, \ldots, L \ . \tag{6.16}$$

Thus, the adaptation algorithm proposed for r_{mn} is

$$\bar{r}_{mn} \leftarrow \bar{r}_{mn} + \lambda_{mn} \check{z}_m \mathrm{sign}\left(\hat{v}_n\right) \ , \tag{6.17}$$

where $\bar{r}_{mn}$ is the estimate of r_{mn}, $\lambda_{mn} > 0$ is an appropriately chosen step size, and $m = 1, \ldots, L \ ,\ n = m+1, \ldots, L$.

Even if $r_{mn}\ ,\ m = 1, \ldots, L\ ,\ n = m+1, \ldots, L$ have been perfectly estimated, $\check{\mathbf{z}}$ must be scaled prior to decoding. A method to estimate the magnitude of the diagonal elements of R (which equal the elements of D^{-1}) is by averaging over time the magnitude of the elements $\check{z}_i$,

$$|\bar{r}_{ll}| = \frac{< |\check{z}_l| >}{\sqrt{\mathcal{E}_l}} \ , \tag{6.18}$$

where $\mathcal{E}_l$ is the average energy of $\mathbf{v}_l$, which must be known to the receiver. A number of time averaging variants can be employed, and the simulations shown later make use of a simple sliding window technique.

Still, the signs of $r_{ll}\ ,\ l = 1, \ldots, L$ are unknown, and it is clear that it is impossible to resolve them blindly. Thus, after convergence has been achieved, a single training symbol needs to be transmitted to determine the signs of r_{ll}.

Observe that given H_F, the rotation angles θ_k , $k = 1, \ldots, L(L-1)/2$ are not unique. It can be shown that the derived decomposition into $G_1, G_2, \ldots, G_{L(L-1)/2}$ and R is unique up to multiplication of these matrices by diagonal matrices with diagonal elements equal to ± 1. This has no consequence for the equalization method.

Also, it should be noted that the performance of this scheme is significantly affected by the decoding order. However, any reordering operation would have to make use of some channel knowledge that would violate the assumption of a blind technique.

The blind equalization structure is graphically depicted in Figure 6.2 for the case of $L = 3$. Observe that one difference with Figure 6.1 is that the scaling matrix D has been moved to the location exactly before the decoding device.

6.2 Simulation Results

The examples used in this section are not constrained to multi-line channels. In these examples, the average transmitted energies, $\mathcal{E}_l$, $l = 1, \ldots, L$, are assumed known at the receiver. The quantities that are estimated are the rotation angles, $\theta_1, \ldots, \theta_{L(L-1)/2}$, and the elements of R.

The convergence of the algorithm is first demonstrated using a simple example with $L = 2$. The case of $L = 2$ was chosen for illustration, however, simulations indicate that convergence is achieved for higher L, although the required convergence time increases approximately linearly with L.

The channel matrix is assumed to be

$$H_F = \begin{bmatrix} 4 & 1.5 \\ 6 & 4 \end{bmatrix} , \tag{6.19}$$

with noise variace $\sigma^2 = 1$ on all dimensions. The modulation scheme is 4-PAM, and the minimum distance between constellation points at the transmitter is 2.

The adaptation parameters are chosen to be $\mu_k = 5 \cdot 10^{-4}$ for all k, and $\lambda_{mn} = 10^{-3}$ for all m and n. The time averaging window size is 1000 symbols. Figure 6.3 shows

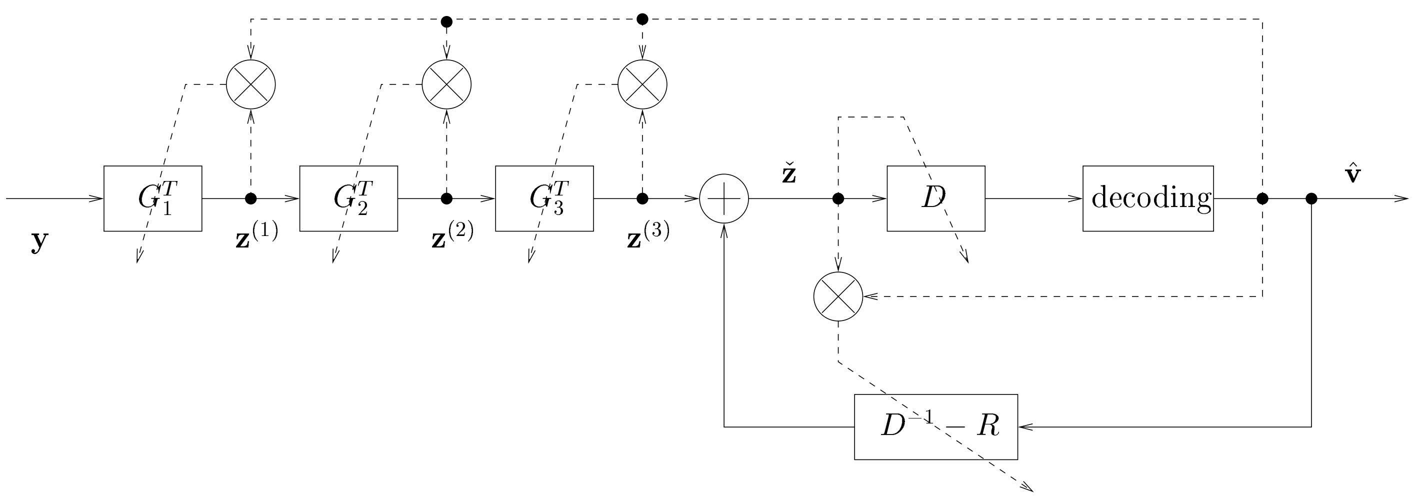

Figure 6.2: Blind upstream Decision Feedback Equalizer example with $L = 3$.

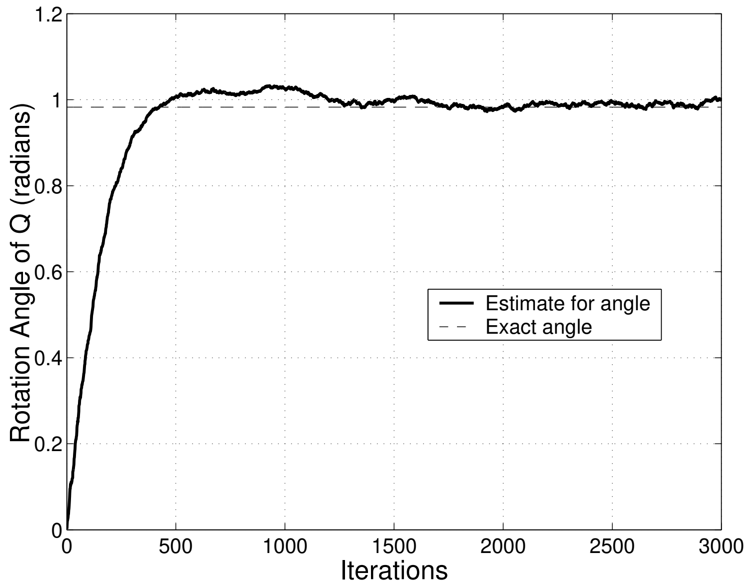

Figure 6.3: Adaptation of the rotation angle corresponding to Q.

the adaptation of the rotation angle as well as the exact rotation angle required to make H_F upper triangular. Figure 6.4 shows the magnitudes of the estimates of R as well as the magnitudes of the exact values. In these figures, the estimates are compared against the values obtained by QR decomposition assuming perfect channel knowledge. Finally, Figure 6.5 shows the squared errors of the two users (measured at the inputs of the decision devices) and compares them against the theoretical errors assuming perfect channel knowledge. It is seen that the measured error of user 1 converges to the theoretical value, but that of stream 2 remains significantly higher than the theoretical value. This discrepancy is almost entirely due to error propagation effects. Since stream 2 has a significantly large error, detection errors occur frequently, which then lead to a large error for stream 1.

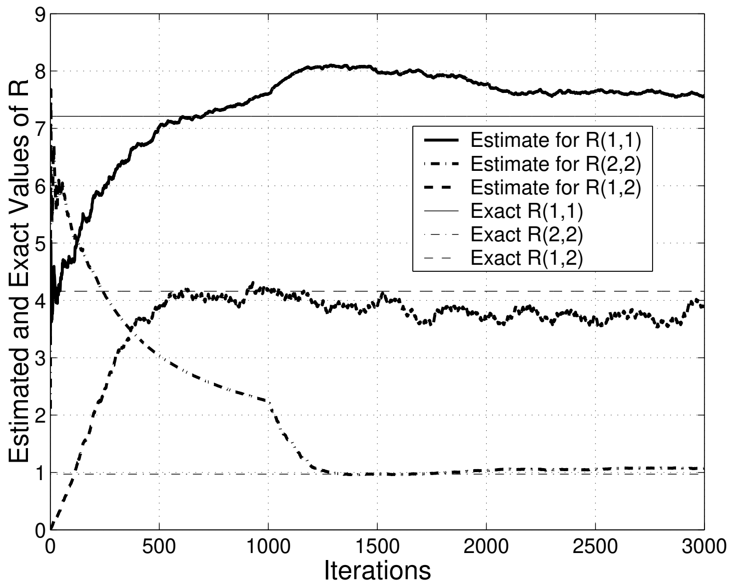

Figure 6.4: Adaptation of the upper triangular entries of R.

Next, the channel matrix is assumed to be

$$H = \begin{bmatrix} 1 & 0.7 & 0.5 \\ 0.7 & 1 & 0.2 \\ 0.5 & 0.2 & 1 \end{bmatrix}, \tag{6.20}$$

and its QR factorization gives the upper triangular matrix

$$R = \begin{bmatrix} -1.187 & -1.095 & -0.910 \\ 0 & -0.672 & 0.068 \\ 0 & 0 & 0.677 \end{bmatrix}. \tag{6.21}$$

Simulation results are now presented studying the probabilities of symbol error for the 3 users. Each simulation run involves 10^6 test symbols on which the probabilities of error are measured. For each SNR setting, 10 such runs are performed in order to average over the possible adaptation outcomes. The number of training symbols is

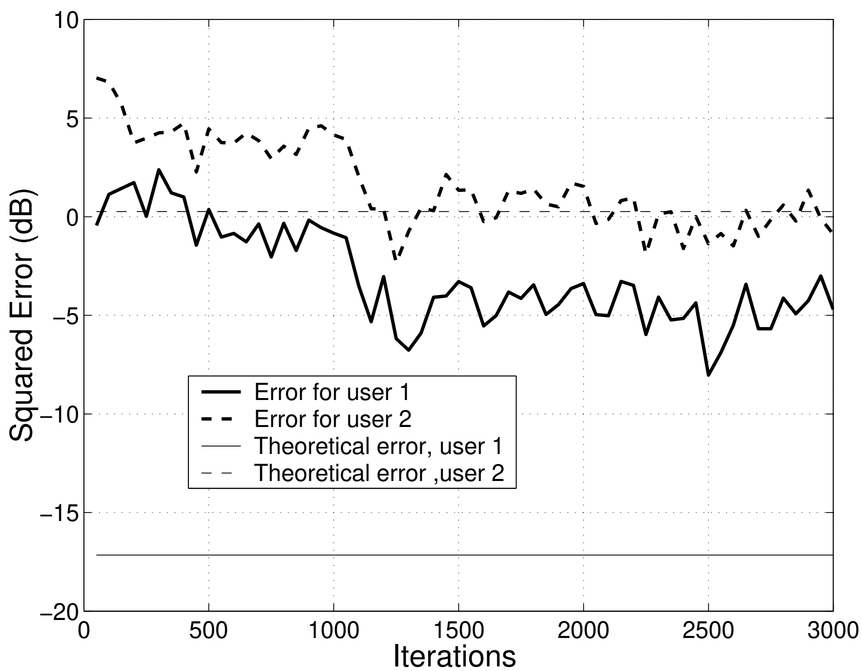

Figure 6.5: Error prior to decoding.

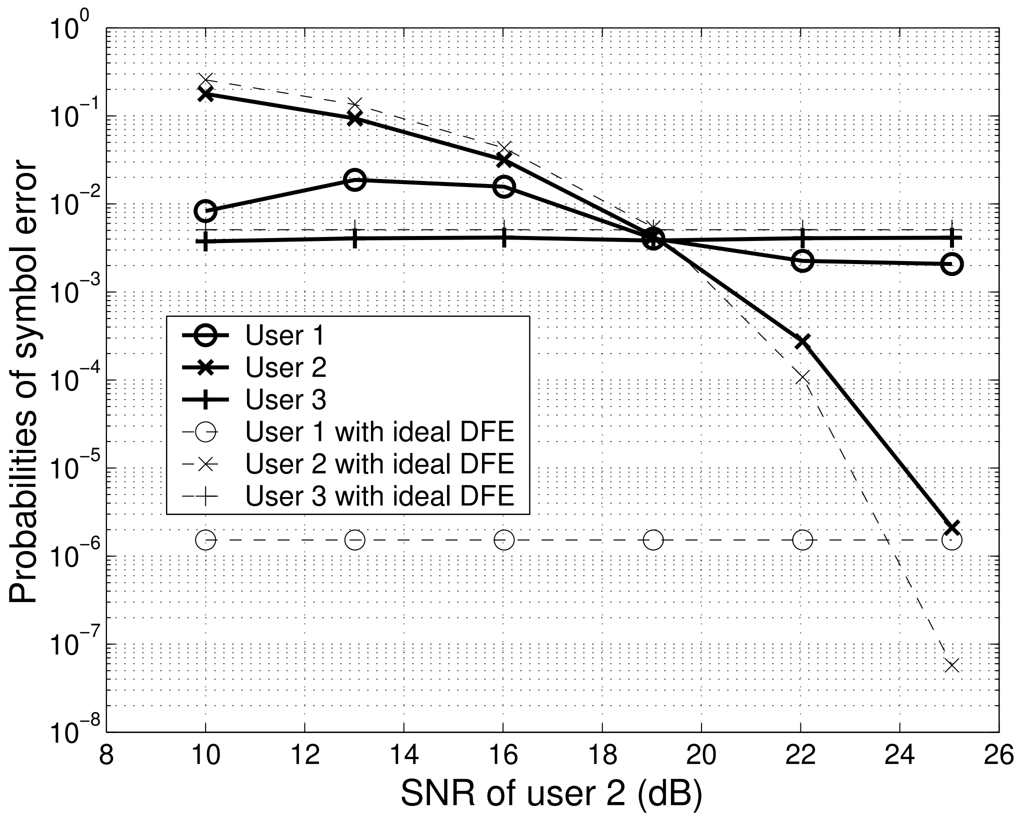

Figure 6.6: Probability of error of the 3 users, varying the SNR of user 2.

10^3, and the adaptation parameters are chosen as $\mu_k = 8 \cdot 10^{-5}$, and $\lambda_{mn} = 10^{-3}$. All other parameters remain the same as before.

Figure 6.6 shows the probabilities of error for the three users, which are compared with the theoretical values obtained from a DFE with perfect channel knowledge, where error propagation effects are ignored. The SNR of a user is defined as the receiver SNR in the absence of interference. The SNRs of users 1 and 3 are fixed at 19dB and the SNR of user 2 varies between 4 and 19dB in increments of 3.

The first observation is that the probability of error of user 3 is very close (and actually lies a bit lower) than the theoretical value. The assumed parameters result in a large value for the probability of error, which has consequences in the detection of the other two users. The curve of user 2 closely follows the theoretical curve, up to the point where the theoretical probability of error of user 2 becomes significantly smaller than the probability of error of user 3. From that point on, user 2 appears to be affected by propagation of errors originating from user 3. However, the effect is not dominant, and this can be explained by examining (6.21), where it is evident that the effect of decision feedback on user 2 is weak. On the other hand, the detection of user 1 strongly depends on subtracting the interference of users 2 and 3, therefore error propagation is dominant. At low SNR, user 1 is degraded due to errors emanating from user 2, but for large SNR, the degradation is due to errors from user 3. Thus, stream 1 is incapable of attaining the theoretical values.

Therefore, the divergence from the theoretical values is mainly due to the inherent characteristics of error propagation. In all cases, the algorithm achieves satisfactory convergence even at relatively low SNR values and with strong interference.

Chapter 7

Conclusions and Future Work

THIS DISSERTATION INVESTIGATED the benefits of coordinated communication for broadband access networks with multiple lines. It was shown that the current design methodology, which treats the system as a set of point-to-point communication links, performs substantially lower than what can be achieved by exploiting the co-location of transceivers at the central node.

This chapter summarizes the key results of this work and suggests topics for future research.

7.1 Conclusions

Multi-line communication systems constitute a fundamental segment of the information technology infrastructure. The most evident examples of such systems are Digital Subscriber Line (DSL) systems, Local Area Networks (LANs) over copper, and serial links for chip-to-chip communication. These systems have a characteristic topology, where multiple lines emanate from a central node and reach remote terminals. This topology offers the opportunity for coordinating the transmissions and the receptions of all lines at the central node side.

The employment of coordination techniques requires that a multiuser view of the multi-line channel is adopted. This dissertation showed that the multi-line channel has distinct properties which lead to interesting information theoretic conclusions. It was seen that the upstream capacity region approximates a rectangle for 2 users, or more generally a rectangular parallelepiped (a "box") for multiple users. Additionally, for both upstream and downstream, the capacity region boundary lies very close to the rate sum bound corresponding to the single user capacity of the vector channel.

Also, specific coordination strategies were proposed, where crosstalk cancellation is utilized. These techniques involve a combination of the principle of multicarrier transmission with the structure of the Generalized Decision Feedback Equalizer (GDFE). The upstream scheme is a novel application of successive interference cancellation for multi-line systems. The downstream scheme extends the concept of precoding to multiuser systems and is a new development (independently proposed in [28]).

The following interesting observations were made about the crosstalk cancellation schemes of the multi-line systems: The benefit of employing a Minimum-Mean-Square-Error (MMSE) criterion instead of a Zero-Forcing (ZF) criterion for reducing crosstalk is typically very minor. Even with ZF, the achievable performance approximates that in a hypothetical environment with no crosstalk. Moreover, the achievable rate sum is very close to the single user capacity of the vector channel. Additionally, it was shown that the implementation cost of the ZF solution can be reduced substantially by decomposing the feed-forward matrix into a set of Givens rotations. Based on this fact, an adaptive technique for blind receiver training in the case of upstream transmission was developed, where each rotation term is adapted independently using a decorrelation criterion.

The issue of transmission optimization in the context of the crosstalk cancellation schemes was also addressed. When a coordinated system operates independently, the optimum energy allocation across the users and across frequency was shown to require independent water-fillings for each user. If neighboring, non-coordinated systems must be protected from excessive crosstalk, then it was shown that the energy allocation problem can be solved numerically by exploiting the convex form of the problem. For the case where energy allocation is combined with frequency planning, a near-optimum solution was obtained by applying a numerical, low-complexity algorithm, that has previously been used in different scenarios, such as subcarrier assignment in multiuser Orthogonal-Frequency-Division-Multiplexing (OFDM).

The performance benefits of coordination for DSL can be summarized as follows: The data rate gains become significant in loops shorter than 4000 feet. In loops shorter than 1500 feet, the data rates increase by more than a factor of 2. In the absence of cooordination, a user requesting an internet connection at a broadband speed of 100Mbps cannot be serviced using DSL technology. Assuming the use of coordination, the user can be accomodated, provided that he is located within 1300 feet of an Optical Network Unit (ONU). If this user has access to 2 DSL lines, then he can be 2200 feet away from an ONU.

Coordination in DSL holds the potential to provide truly broadband speeds and at the same time eliminate the need to replace the most expensive segment of the copper wire infrastrucure.

7.2 Future Work

Enhancements of the crosstalk cancellation methods presented in Chapter 4 are possible areas of future research. Downstream transmission was seen to experience error propagation due to the decision feedback operations. Some early work indicates that eliminating feedback processing incurs a rather small loss in the achievable performance, which is potentially offset by complexity reduction. In this case, the receiver architecture is simplified to operations of *maximum ratio combining* across all lines and within each tone. Interestingly, such a scheme is still capable of approximating well the information theoretic bounds of Chapter 2.

The proposed scheme for upstream transmission employs modulo precoding. However, more general schemes such as *lattice precoding* or *trellis precoding* are known to have improved performance. Applying these existing techniques to multi-line channels and evaluating their benefit is a direction which is worth exploring.

Almost all the results presented in the previous chapters assumed that prior knowledge of the multiuser channel is available. Although the literature on MIMO channel identification is extensive, the multi-line case presents new challenges. When a new line is activated, it may significantly alter the characteristics of crosstalk within the coordinated system. On the one hand, the channel information needs to be updated in order to account for the new line, so that transmission on the new line commences as soon as possible. On the other hand, there must be minimal disruption in the operation of the previously active lines. These requirements are conflicting, and lead to a number of interesting research questions.

It was shown in Chapter 5 that transmission optimization with power back-off and with frequency planning can be formulated as a convex programming problem. Although algorithms for convex problems are becoming increasingly popular, their implementation in practical systems may face complexity difficulties. Future work could include the investigation of sub-optimum, low-complexity algorithms for these scenarios.

Another scenario of interest is when a coordinated system is adjacent to a non-coordinated system, or possibly to some other coordinated system. In such cases, each system must act independently in order to both optimize its transmission and to prevent excessive crosstalk from seriously affecting the neighboring system. Although one possible approach is the method of Subsection 5.1.2, a promising alternative is the combination of the concept of iterative waterfilling [67, 68] with the coordinated transmission techniques of Chapter 4.

Appendix A

Convergence of the Generalized DFE to the MMSE-DFE

THIS APPENDIX PROVES that under certain conditions the Generalized Decision Feedback Equalizer (GDFE) [42] asymptotically converges to the infinite length Minimum-Mean-Square-Error Decision Feedback Equalizer (MMSE-DFE) [44] in the limit of a large number of dimensions. The results extend to the case where the optimum spectrum consists of several disjoint bands, which implies the operation of multiple corresponding independent MMSE-DFE's.

The MMSE-DFE and the GDFE (applied to block transmission with ISI) are derived in Sections A.1 and A.2 respectively. The GDFE is derived with a different approach than that of Chapter 3, in order to show the parallels between the two systems. Section A.3 shows the convergence results under the assumption that all frequencies are utilized, or equivalently all GDFE dimensions are used. Then, Section A.4 addresses the same problem in the context of having unused frequency regions, or equivalently having some unused GDFE dimensions. Finally, Section A.5 gives an example to illustrate the convergence results.

A.1 Optimum Transmission with MMSE-DFE

A.1.1 Transmission Optimization

Suppose that $P(D)$ is the D-transform of the impulse response of a scalar channel, and without loss of generality assume that the noise is white with unit variance, and,

$$Y(D) = P(D)X(D) + N'(D) \ . \tag{A.1}$$

Given an energy constraint on $X(D)$, waterfilling provides the solution for the optimum channel input autocorrelation $R_{xx}(D)$.

The Paley-Wiener criterion states that if

$$\int_{-\pi}^{\pi} |\ln R_{xx}(e^{-\jmath\omega})| d\omega < \infty \ , \tag{A.2}$$

then the spectral factorization exists,

$$R_{xx}(D) = \mathcal{E} G_x(D) G_x^*(D^{-*}) \ , \tag{A.3}$$

where $G_x(D) = \sum_{n=-\infty}^{\infty} g_{x,n} D^n$ is monic ($g_{x,0} = 1$), causal ($g_{x,n} = 0$ for $n < 0$) and minimum phase (all its poles are outside the unit circle, and all its zeros are on or outside the unit circle), and $\mathcal{E}$ is a positive constant. Therefore, the channel input $X(D)$ is obtained by $X(D) = G_x(D)V(D)$, where $V(D)$ is a white sequence with energy $\mathcal{E}$, representing the data sequence that needs to be communicated.

However, the Paley-Wiener criterion is not satisfied when the optimized input spectrum contains frequency "gaps" where no input energy is assigned (e.g. Figure A.1). In those cases, the channel is partitioned into L sub-channels, where independent MMSE-DFE's operate. Assuming that the pulse response of the i-th sub-channel is $P_i(D)$, then the optimum sub-channel input autocorrelation $R_{xx,i}(D)$ satisfies the Paley-Wiener criterion. Thus, the transmission optimization filter obtained is $G_{x,i}(D)$ and the transmission filtering operation is $X_i(D) = G_{x,i}(D)V_i(D)$, where $X_i(D)$ is the i-th sub-channel input, and $V_i(D)$ is the data sequence with energy

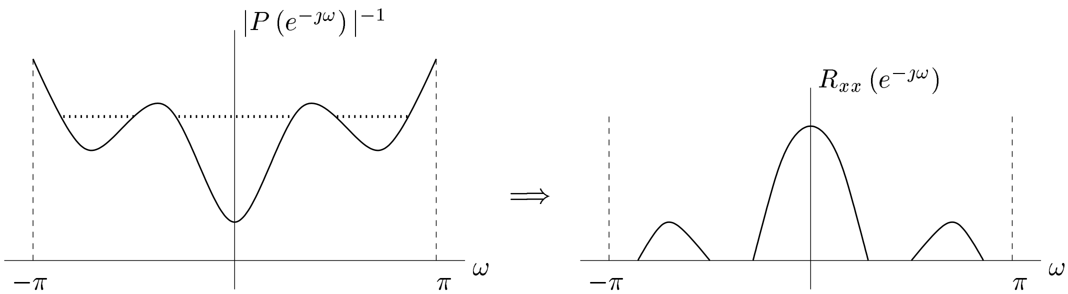

Figure A.1: Example of waterfilling and resulting input spectrum.

$\mathcal{E}_i$ being transmitted in the i-th band.

A.1.2 MMSE-DFE Receiver

Let $P(D)$ and $G_x(D)$ be the pulse response and the transmit optimization filter corresponding to a channel (or possibly sub-channel). Taking into account transmit optimization and receiver matched filtering, the *forward canonical model* is obtained,

$$Z(D) = R_f(D)V(D) + N(D) \ , \tag{A.4}$$

where $Z(D)$ is the matched filter output, $R_f(D) = G_x^*(D^{-*})P^*(D^{-*})P(D)G_x(D)$ and $R_{nn}(D) = R_f(D)$. The *backward canonical model* is obtained as:

$$V(D) = R_b(D)Z(D) + E(D) \Rightarrow \tag{A.5}$$

$$Z(D) = R_b^{-1}(D)V(D) - R_b^{-1}(D)E(D) \ , \tag{A.6}$$

where $R_b(D)$ is the MMSE Linear Equalizer of the channel and $E(D)$ is the MMSE error sequence, with $R_b^{-1}(D) = R_f(D) + \mathcal{E}^{-1}$ and $R_{ee}(D) = R_b(D)$. The Paley-Wiener criterion for $R_b^{-1}(D)$ is always satisfied, so

$$R_b^{-1}(D) = s_o G(D)G^*(D^{-*}) \ , \tag{A.7}$$

where s_o is a positive constant, and $G(D)$ is monic, causal and minimum-phase. If $Z(D)$ is processed by the anti-causal filter $s_o^{-1}G^{-*}(D^{-*})$, then (A.6) becomes

$$s_o^{-1}G^{-*}(D^{-*})Z(D) \quad = \quad G(D)V(D) - E'(D) \;, \tag{A.8}$$

where $E'(D) = G(D)E(D)$ and $R_{e'e'}(D) = s_o^{-1}$. Since $G(D)$ is monic, causal and minimum-phase, a DFE can be readily implemented with the feedback filter being $1 - G(D)$. The detection SNR is $SNR_{MMSE-DFE} = s_o\mathcal{E}$. The complete MMSE-DFE equalization structure is shown in Figure A.2.

A.2 Optimum Transmission with GDFE

A.2.1 Transmission Optimization

Block transmission and reception in a scalar channel with ISI is now assumed. Let P be an $N \times (N + \nu)$ convolution matrix (corresponding to a channel with $\nu + 1$ taps) and let the noise be white with unit variance, and

$$\mathbf{y} = P\mathbf{x} + \mathbf{n}' \;, \tag{A.9}$$

where P is a Toeplitz matrix and its first row is $[p_0\; p_1\; \cdots\; p_\nu\; 0\; \cdots\; 0]$. Given an energy constraint on the input vector $\mathbf{x}$, an optimum channel input covariance matrix $R_{\mathbf{xx}}$ can be found by waterfilling in a set of parallel subchannels, whose gains are equal to the singular values of the matrix P [46].

The covariance matrix $R_{\mathbf{xx}}$ may be singular. In the assumed block transmission model, ν input dimensions are always "lost", and singularities may also arise because transmitting in the null space of P is a waste of energy, and because waterfilling optimization may require that some dimensions are not used. Instead of the frequency gaps (encountered with MMSE-DFE), one now finds "gaps" in the dimensions utilized. The non-singularity condition is expressed as

$$\ln |R_{\mathbf{xx}}| < \infty. \tag{A.10}$$

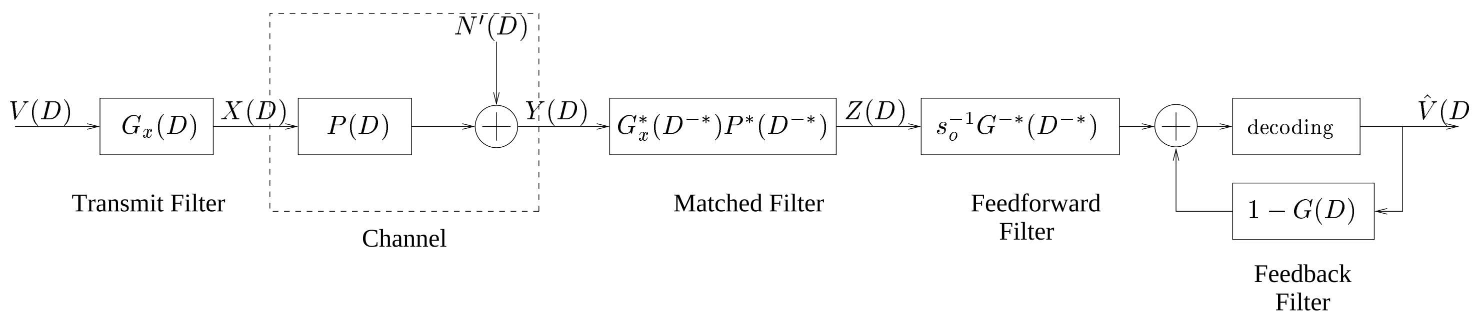

Figure A.2: MMSE-DFE equalization structure.

There are several methods to construct a channel input with the desired covariance matrix, and each of them results in a different structure. Here, a method is given which facilitates the proof of the convergence to the MMSE-DFE. Assuming that $R_{\mathbf{xx}}$ is non-singular, the Cholesky factorization $R_{\mathbf{xx}} = G_x R_{\mathbf{vv}} G_x^*$ is obtained, where G_x is upper triangular and monic (diagonal elements equal 1), and $R_{\mathbf{vv}}$ is diagonal. Then, $\mathbf{x}$ can be derived by $\mathbf{x} = G_x \mathbf{v}$.

Clearly, $R_{\mathbf{xx}}$ is singular for the channel described by (A.9), unless the unused dimensions are "removed". Let A be an $(N+\nu) \times N^\star$ matrix, where $N^\star$ is the rank of $R_{\mathbf{xx}}$. Defining $\mathbf{x} = A\mathbf{u}$, the channel model becomes

$$\mathbf{y} = PA\mathbf{u} + \mathbf{n}' \ . \tag{A.11}$$

By appropriately choosing A, the optimum channel input covariance matrix $R_{\mathbf{uu}}$ can be made non-singular of the same rank as $R_{\mathbf{xx}}$. This means that all dimensions of PA are used, and, additionally, matrix A eliminates those input dimensions of P which are not useful for communication. There is no unique method to find A, however the choices of Sections A.3 and A.4 serve to illustrate the equivalence between frequency gaps and dimensionality gaps.

Then, the Cholesky factorization of $R_{\mathbf{uu}}$ gives $R_{\mathbf{uu}} = G_u R_{\mathbf{vv}} G_u^*$. The matrix G_u serves as the transmission optimization section of the transmitter, and $\mathbf{u} = G_u \mathbf{v}$, where $\mathbf{v}$ is the communicated data. Vector $\mathbf{u}$ goes through the "dimension-expanding" matrix A, and finally gives the channel input $\mathbf{x}$.

A.2.2 GDFE Receiver

Taking into account singularity elimination and transmission optimization at the transmitter and receiver matched filtering, the *forward canonical model* is obtained,

$$\mathbf{z} = R_f \mathbf{v} + \mathbf{n} \ , \tag{A.12}$$

where $\mathbf{z}$ is the matched filter output, $R_f = G_u^* A^* P^* P A G_u$ and $R_{\mathbf{nn}} = R_f$. The *backward canonical model* is also obtained as

$$\mathbf{v} = R_b \mathbf{z} + \mathbf{e} \Rightarrow \mathbf{z} = R_b^{-1} \mathbf{v} - R_b^{-1} \mathbf{e} \ , \tag{A.13}$$

where R_b is the matrix MMSE Linear Equalizer of the channel and $\mathbf{e}$ is the MMSE error vector, with $R_b^{-1} = R_f + R_{\mathbf{vv}}^{-1}$ and $R_{\mathbf{ee}} = R_b$ [42]. Since R_b^{-1} is always non-singular, the following is derived,

$$R_b^{-1} = G^* S_o G \ , \tag{A.14}$$

where S_o is a diagonal matrix with positive elements, and G is upper triangular and monic. If $\mathbf{z}$ is processed by the lower triangular matrix $S_o^{-1} G^{-*}$, then the backward canonical model becomes

$$S_o^{-1} G^{-*} \mathbf{z} = G\mathbf{v} - G\mathbf{e} = G\mathbf{v} - \mathbf{e}' \ , \tag{A.15}$$

where $R_{\mathbf{e}'\mathbf{e}'} = S_o^{-1}$. Since G is an upper triangular matrix, the input vector $\mathbf{v}$ can be recovered by decision feedback. With the assumption that decision errors are very rare, a set of $N^\star$ decoupled sub-channels is produced, and the corresponding detection SNR's are $SNR_i = \mathcal{E}_{v,i} S_{o,i}$, $i = 1, \ldots, N^\star$, where $\mathcal{E}_{v,i}$ is the i-th diagonal element of $R_{\mathbf{v}v}$, and $S_{o,i}$ is the i-th diagonal element of S_o. This equalization structure is called a generalized DFE (GDFE) and is shown in Figure A.3.

A.3 Convergence of GDFE to MMSE-DFE without unused dimensions

Let $P(D)$ be the pulse response of a channel with length $\nu + 1$. By assumption, the optimized input spectrum $R_{xx}(e^{-\jmath\omega})$ is strictly positive at all frequencies. Therefore, the spectral factorization of $R_{xx}(D)$ can be obtained and the transmission filter $G_x(D)$, the feedforward filter $S_o^{-1} G^{-*}(D^{-*})$ and feedback filter $1 - G(D)$ are derived.

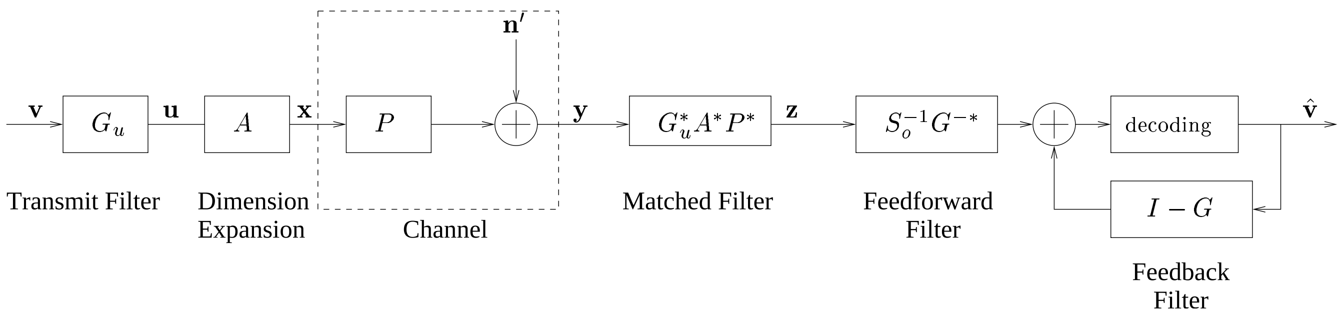

Figure A.3: GDFE equalization structure.

Next, the GDFE structure for the same channel is considered, where P is the corresponding $N \times (N+\nu)$ convolution matrix. The use of a cyclic prefix (CP) may be modeled with

$$A_{cp} = \begin{bmatrix} & I_N \\ I_\nu & O_{\nu\times(N-\nu)} \end{bmatrix} , \tag{A.16}$$

which makes $\tilde{P} = PA_{cp}$ a circulant matrix, and thus,

$$\mathbf{y} = \tilde{P}\mathbf{u} + \mathbf{n}' . \tag{A.17}$$

It is known that a circulant matrix R has an eigenvalue decomposition $R = QDQ^*$, where $Q = Q_{IDFT}$ is a matrix performing the N-point Inverse Discrete Fourier Transform (IDFT), $Q^* = Q_{DFT}$ is the corresponding Discrete Fourier Transform (DFT) matrix, and D's diagonal elements - the eigenvalues of R - equal the DFT values of the first row of R.

Lemma A.3.1 *Let $R_{\mathbf{uu}}$ be circulant with eigenvalues $\lambda_{u,j}$ equal to the energy assigned by waterfilling to a set of N parallel sub-channels with gains equal to the DFT values of the first row of $\tilde{P}$. Then, this is an optimum covariance matrix of* $\mathbf{u}$.

Proof: The matrix $\tilde{P}$ can be decomposed as $\tilde{P} = Q_{IDFT}\Lambda Q_{DFT}$, where Λ is diagonal with elements λ_j equal to the DFT values of the channel response. Maximizing the mutual information of the circulant channel is equivalent to the maximization of

$$|I + \tilde{P}R_{\mathbf{uu}}\tilde{P}^*| = |I + \Lambda Q_{DFT}R_{\mathbf{uu}}Q_{IDFT}\Lambda^*| . \tag{A.18}$$

Defining the diagonal elements of $Q_{DFT}R_{\mathbf{uu}}Q_{IDFT}$ as $\lambda_{u,j}$, and applying Hadamard's inequality, it is found that

$$|I + \Lambda Q_{DFT}R_{\mathbf{uu}}Q_{IDFT}\Lambda^*| \leq \prod_{j=1}^{N} \left(1 + |\lambda_j|^2\lambda_{u,j}\right) , \tag{A.19}$$

with equality if $Q_{DFT}R_{\mathbf{uu}}Q_{IDFT}$ is diagonal, which holds when $R_{\mathbf{uu}}$ is circulant. Note that the $\lambda_{u,j}$'s are then equal to the eigenvalues of $R_{\mathbf{uu}}$, and their optimum values are determined by waterfilling in a set of N parallel channels with gains λ_j. ■

The following can now be shown:

Proposition A.3.1 *For $N \to \infty$, the elements of a row of $R_{\mathbf{uu}}$ converge to the coefficients $r_{xx,n}$ of the autocorrelation function $R_{xx}(D) = \sum_{n=-\infty}^{\infty} r_{xx,n} D^n$.*

Proof: According to Lemma A.3.1, the optimum covariance matrix of $\mathbf{u}$ can be determined by dividing the channel into N sub-channels and waterfilling in them. When the number of dimensions N becomes infinite, the number of sub-channels also becomes infinite. Therefore, the infinite set of eigenvalues constitutes a continuous representation of the energy distribution across the band. Since the optimum energy distribution in the band of interest is unique, the set of eigenvalues will be identical to the optimized input spectrum $R_{xx}(e^{-\jmath\omega})$ of the MMSE-DFE. This can be written as $\lambda_{u,j} \to R_{xx}(e^{-\jmath\omega})$, where it is assumed that the eigenvalues $\lambda_{u,j}$ are ordered, so that higher values of the index j correspond to higher frequencies ω. With $N \to \infty$, $R_{\mathbf{uu}}$ is a doubly infinite Toeplitz matrix, and all its rows become identical,

$$R_{\mathbf{uu}} = Q_{IDFT} \mathrm{diag}\left(\ldots, \lambda_{u,j-1}, \lambda_{u,j}, \lambda_{u,j+1}, \ldots\right) Q_{DFT} \ . \tag{A.20}$$

The eigenvalues $\lambda_{u,j}$ are equal to the Discrete-Time-Fourier-Transform (DTFT) values of a row of $R_{\mathbf{uu}}$. But they are also equal to $R_{xx}(e^{-\jmath\omega})$, which is the DTFT of $r_{xx,n}$, therefore, a row of $R_{\mathbf{uu}}$ is identical to the $r_{xx,n}$ coefficients. (See [88] for alternative proof.) ■

The cyclic prefix results in some energy waste for the GDFE, but at the limit of $N \to \infty$ this effect disappears. The following theorem is fundamental for the rest of the paper:

Theorem A.3.1 *Let $R(D)$ be the autocorrelation function of a stationary stochastic process satisfying the Paley-Wiener criterion. Also, let R be the doubly infinite Toeplitz covariance matrix corresponding to this stochastic process.*

The spectral factorization of $R(D)$ exists and $R(D) = \mathcal{E} G(D) G^*(D^{-*})$, where $G(D) = \sum_{n=0}^{\infty} g_n D^n$ is monic, causal and minimum phase, and $\mathcal{E}$ is a positive constant. The Cholesky factorization of R also exists and $R = LDL^*$, where L is doubly infinite, Toeplitz, lower triangular and monic with each row being $[\ldots\ l_2\ l_1\ 1\ 0\ 0\ \ldots]$.

Matrix D is doubly infinite, Toeplitz, and diagonal with each row being $[\ldots\ 0\ 0\ d\ 0\ 0\ \ldots]$. Then,

$$l_n = g_n \ , \ n \geq 1 \ , \tag{A.21}$$

$$d = \mathcal{E} > 0 \ . \tag{A.22}$$

Also, the Cholesky factorization $R = UDU^*$ exists, where U is doubly infinite, Toeplitz, upper triangular and monic with each row being $[\ldots\ 0\ 0\ 1\ u_1\ u_2\ \ldots]$, and $u_n = g_n^*$, $n \geq 1$.

Proof: See [89, Chap. 6]. ■

Finally, the following theorem can be proved:

Theorem A.3.2 *Assume the optimum MMSE-DFE and GDFE structures for a specific channel, where the Paley-Wiener criterion is satisfied. The components of the GDFE are the following matrices: the transmission optimization matrix G_u, the feed-forward matrix $S_o^{-1}G^{-*}$ and the feedback matrix $I-G$. The components of the MMSE-DFE are the following filters: the transmission optimization filter $G_x(D)$, the feedforward filter $s_o^{-1}G^{-*}(D^{-*})$ and the feedback filter $1-G(D)$.*

When $N \to \infty$, the rows of the GDFE matrices become identical and the elements of a row are equal to the terms of the corresponding MMSE-DFE filter.

Proof: It is sufficient to show the equivalence of G_u, G, and S_o with $G_x(D)$, $G(D)$, and s_o respectively. By assumption, $R_{xx}(e^{-\jmath\omega})$ has no frequency gaps where zero energy is assigned, so by Proposition A.3.1, the eigenvalues of the doubly infinite $R_{\mathbf{uu}}$ are strictly positive, and $R_{\mathbf{uu}}$ is non-singular. The transmission optimization matrix for GDFE is then found by obtaining the Cholesky factorization $R_{\mathbf{uu}} = G_u R_{\mathbf{vv}} G_u^*$. By Theorem A.3.1, the rows of G_u are identical to the MMSE-DFE transmission optimization filter terms given in $G_x(D)$. Also, the diagonal elements of $R_{\mathbf{vv}}$ are equal to $\mathcal{E}$.

Next, the receiver side is examined. The matrix $R_f = G_u^* A^* P^* P A G_u$ exhibits Hermitian symmetry, but is not Toeplitz for a finite number of dimensions due to the fact that G_u is not Toeplitz. However, at the limit of an infinite number of dimensions,

G_u becomes Toeplitz and so does R_f. Therefore, the rows of the doubly-infinite matrix R_f are identical.

Observe that at the limit the matrix A becomes an identity matrix. Since multiplication of the doubly infinite convolution matrices is equivalent to computing the convolution of the corresponding sequences, the elements of the rows of R_f are identical to the terms of $R_f(D)$. The same relationship is true for R_b^{-1} and $R_b^{-1}(D)$ which are derived from R_f and $R_f(D)$ respectively. By applying Theorem A.3.1 on R_b^{-1} and $R_b^{-1}(D)$, it is deduced that the rows of G (as in (A.14)) will be identical to the terms of $G(D)$ (as in (A.7)). Additionally, the diagonal elements of S_o will equal s_o (the inverse of the MMSE error variance). ■

If one views the GDFE as a time-varying equalizer, one observes that with $N \to \infty$ it becomes time-invariant and performs the same functions as the MMSE-DFE.

A.4 Convergence of GDFE to MMSE-DFE with unused dimensions

Let P be the corresponding $N \times (N + \nu)$ convolution matrix, and $\tilde{P} = PA_{cp}$ be the resulting matrix after introduction of a cyclic prefix. Then,

$$\mathbf{y} = \tilde{P}\mathbf{w} + \mathbf{n}' \ . \tag{A.23}$$

Let $R_{\mathbf{ww}}$ be the optimum covariance matrix of $\mathbf{w}$, which according to Lemma A.3.1 has the property of being circulant, and its eigenvalues are equal to the optimum energy distribution for the corresponding subchannels. In this case the optimized spectrum $R_{xx}(e^{-\jmath\omega})$ may contain gaps, so by Proposition A.3.1, as $N \to \infty$, there will be eigenvalues equal to zero. Therefore $R_{\mathbf{ww}}$ is singular, and its eigenvalue decomposition is expressed as $R_{\mathbf{ww}} = Q_{IDFT} D Q_{DFT}$, where

$$D = \mathrm{diag}\begin{bmatrix} \mathbf{0} & \mathbf{d}_1 & \mathbf{0} & \mathbf{d}_2 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{d}_L & \mathbf{0} \end{bmatrix} . \tag{A.24}$$

It is assumed that the eigenvalues are arranged in order of increasing frequency. Each of the $N_i \times 1$ vectors $\mathbf{d}_i$ holds the positive eigenvalues corresponding to one of the L bands. The zero vectors correspond to the "gaps" in the optimized spectrum.

The method to remove the unused dimensions is next shown. The following matrix is defined,

$$J = \begin{bmatrix} O_{N_1^z \times N_1} & O_{N_1^z \times N_2} & \dots & O_{N_1^z \times N_L} \\ I_{N_1} & O_{N_1 \times N_2} & \dots & O_{N_1 \times N_L} \\ O_{N_2^z \times N_1} & O_{N_2^z \times N_2} & \dots & O_{N_2^z \times N_L} \\ O_{N_2 \times N_1} & I_{N_2} & \dots & O_{N_2 \times N_L} \\ \vdots & \vdots & \ddots & \vdots \\ O_{N_L \times N_1} & O_{N_L \times N_2} & \dots & I_{N_L} \\ O_{N_{L+1}^z \times N_1} & O_{N_{L+1}^z \times N_2} & \dots & O_{N_{L+1}^z \times N_L} \end{bmatrix} , \tag{A.25}$$

having $N = \sum_{i=1}^{L} N_i + \sum_{i=1}^{L+1} N_i^z$ rows and $N^\star = \sum_{i=1}^{L} N_i$ columns. The rows with all elements equal to zero correspond to dimensions that are assigned no energy and therefore need to be eliminated. Each of the L identity blocks corresponds to a band of the optimized spectrum. Also,

$$Q_{mb} = \mathrm{diag}\left[Q_{DFT,1}\ Q_{DFT,2}\ \dots\ Q_{DFT,L}\right] , \tag{A.26}$$

where each $Q_{DFT,i}$ is an $N_i \times N_i$ DFT matrix corresponding to a band of the optimized spectrum. Setting $A = A_{cp} Q_{IDFT} J Q_{mb}$, the following operation is performed on P,

$$P_{rs} = PA = PA_{cp} Q_{IDFT} J Q_{mb} , \tag{A.27}$$

thus obtaining

$$\mathbf{y} = P_{rs}\mathbf{u} + \mathbf{n}' . \tag{A.28}$$

Note that this operation does not have any consequence on optimum transmission, since the DFT and IDFT operations are known to be one-to-one, and the dimensions eliminated are those which are not useful for communication. Moreover, it serves to "decouple" the bands, and it could be thought of as a generalization of the frequency shift and resampling operations. The technique is illustrated in Figure A.4.

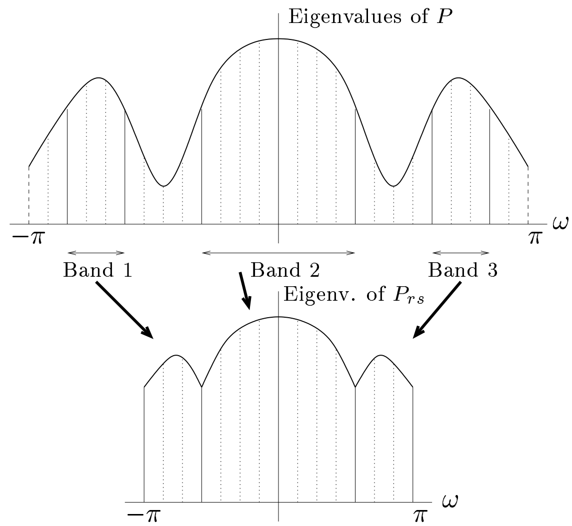

Figure A.4: Illustration of generalized resampling.

It remains to determine the optimum covariance matrix $R_{\mathbf{uu}}$.

Lemma A.4.1 *The optimum covariance matrix $R_{\mathbf{uu}}$ is*

$$R_{\mathbf{uu}} = diag\left[\begin{array}{cccc} R_{\mathbf{uu},1} & R_{\mathbf{uu},2} & \ldots & R_{\mathbf{uu},L} \end{array}\right] \tag{A.29}$$

$$R_{\mathbf{uu},i} = Q_{IDFT,i} D_i Q_{DFT,i} \ , \ i = 1, \ldots, L \ , \tag{A.30}$$

where D_i has diagonal elements equal to the energy distribution in the subchannels of the i-th band.

Proof: Maximizing the mutual information of the channel is equivalent to maximizing

$$|I + P_{rs} R_{\mathbf{uu}} P_{rs}^*| = |I + \Lambda J Q_{mb} R_{\mathbf{uu}} Q_{mb}^* J^* \Lambda^*| \ , \tag{A.31}$$

where $\Lambda = Q_{DFT} \tilde{P} Q_{IDFT}$ is diagonal and its elements equal the DFT values of the channel response. Multiplying Λ by J eliminates the diagonal elements corresponding to frequencies not used for transmission. Defining $\Lambda' = \Lambda J$ with diagonal elements

λ_j, and applying Hadamard's inequality,

$$|I + \Lambda' Q_{mb} R_{\mathbf{uu}} Q_{mb}^* \Lambda'^*| \leq \prod_{j=1}^{N^\star} \left(1 + |\lambda_j|^2 \lambda_{u,j}\right) \ , \tag{A.32}$$

where $\lambda_{u,j}$ are the diagonal elements of $Q_{mb} R_{\mathbf{uu}} Q_{mb}^*$. Equality holds if $Q_{mb} R_{\mathbf{uu}} Q_{mb}^*$ is diagonal, which is true when $R_{\mathbf{uu}}$ is as described by (A.29) and (A.30). The $\lambda_{u,j}$'s are then equal to the non-zero eigenvalues of $R_{\mathbf{uu}}$, and their optimum values are determined by waterfilling in the set of subchannels belonging to the L optimum transmission bands. ■

Finally, the convergence of GDFE to the multiple MMSE-DFE's can be demonstrated:

Theorem A.4.1 *Assume an arbitrary channel, where optimum transmission may require a set of L MMSE-DFE structures operating in parallel. Also, assume the optimum GDFE structure for this channel. The i-th MMSE-DFE consists of the following filters: the transmission optimization filter $G_{x,i}(D)$, the feedforward filter $s_{o,i}^{-1} G_i^{-*}(D^{-*})$ and the feedback filter $1 - G_i(D)$. The optimum GDFE structure is separable into L modules, with each module performing independently of the others. The i-th GDFE module consists of the following matrices: the transmission optimization matrix $G_{u,i}$, the feedforward matrix $S_{o,i}^{-1} G_i^{-*}$ and the feedback matrix $I - G_i$.*

When $N \to \infty$, the rows of the matrices of the i-th GDFE module become identical and the elements of a row are equal to the terms of the corresponding filter of the i-th MMSE-DFE.

Proof: First, it is shown that the GDFE is separable into L modules. Deriving the Cholesky factorization, one obtains $R_{\mathbf{uu}} = G_u R_{\mathbf{vv}} G_u^*$, where G_u is upper triangular and monic. By Lemma A.4.1, G_u must also be block diagonal with its diagonal blocks being upper triangular matrices, and each block corresponding to a band. Then,

$$R_f = G_u^* Q_{mb}^* J^* \Lambda^* \Lambda J Q_{mb} G_u = G_u^* Q_{mb}^* \Lambda'^* \Lambda' Q_{mb} G_u \ , \tag{A.33}$$

where $\Lambda = Q_{DFT} P A_{cp} Q_{IDFT}$ is diagonal and its elements equal the DFT values of the channel response. The operation $\Lambda' = \Lambda J$ eliminates the diagonal elements corresponding to unused frequencies, and $\Lambda'^{*}\Lambda'$ is diagonal. From there, it is straightforward to see that R_f is block diagonal.

Therefore, $R_b^{-1} = R_f + R_{\mathbf{vv}}^{-1}$ must also be block diagonal. After deriving its Cholesky factorization $R_b^{-1} = G^* S_o G$, one observes that the matrix G is block diagonal. Clearly, the feedforward section, $S_o^{-1} G^{-*}$, is block diagonal as well (with its diagonal block matrices being lower triangular), and the same is true for the feedback section $I - G$.

It has now been shown that G_u, $S_o^{-1} G^{-*}$ and $I - G$ are block diagonal with the same block sizes. Then, it is easy to observe that each device of the GDFE (the transmit optimization matrix at the transmitter, and the feedforward and feedback matrices at the receiver) can be broken into L independent components corresponding to the L bands. Combining the corresponding components from all devices, L independent GDFE modules are obtained, one for each band.

Making use of Theorem A.3.2, it is found that as the number of dimensions grows to infinity, the GDFE module of each band approaches the corresponding MMSE-DFE system. ■

A.5 Example

The convergence results are now demonstrated for the channel $P(D) = 1 + D + D^2$. The noise has variance $\sigma^2 = 0.3$, and the input energy per dimension is $\mathcal{E}_x = 1$. Given these parameters, the frequency response magnitude and the optimum transmission spectrum determined by water-filling are shown in Figure A.5. Clearly, optimum MMSE-DFE transmission in this channel requires the operation of three independent MMSE-DFE systems in three disjoint bands.

Assuming the use of generalized resampling, the GDFE structure for this channel is separable into 3 independent modules. The magnitude of the DFT of the center row of each component is next plotted for N equal to 128, 256 and 512, together with the DFT magnitude of the corresponding MMSE-DFE filter. Figures A.6 and A.7

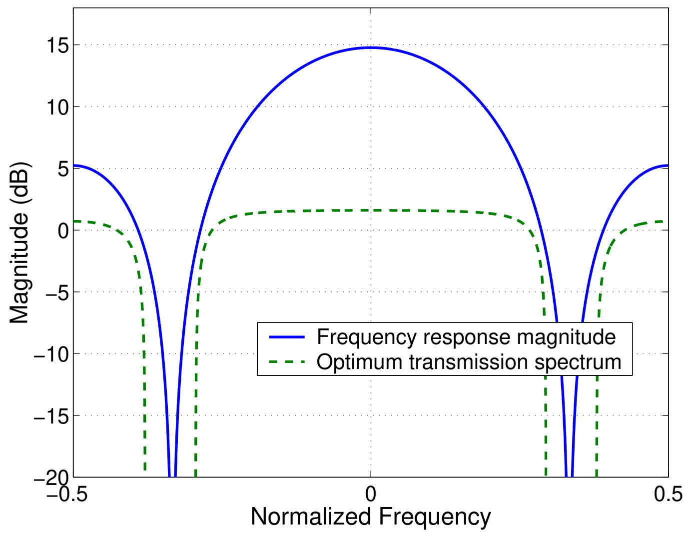

Figure A.5: Frequency response magnitude and optimum transmission spectrum.

depict the GDFE matrices and MMSE-DFE filters corresponding to bands 2 and 3. (Plots for band 1 are symmetric to those of band 3.) It is observed that convergence holds for all matrices in all bands.

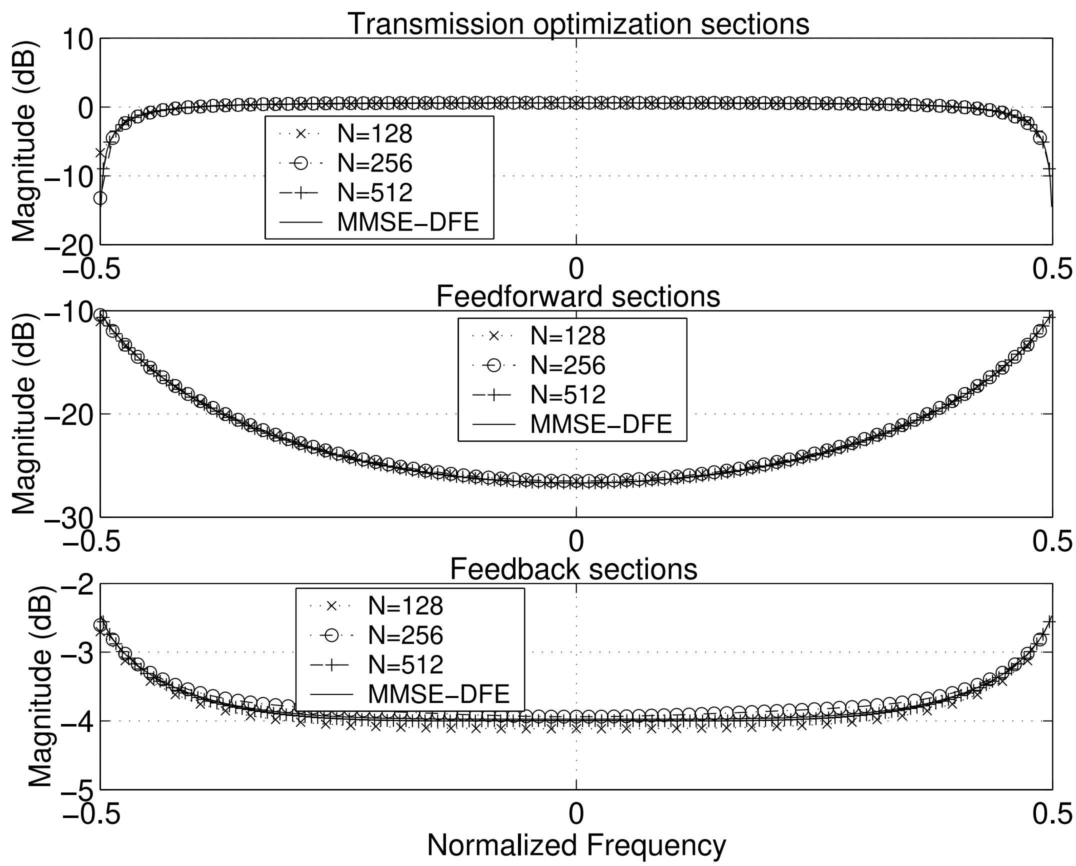

Figure A.6: Illustration of convergence of GDFE in band 2.

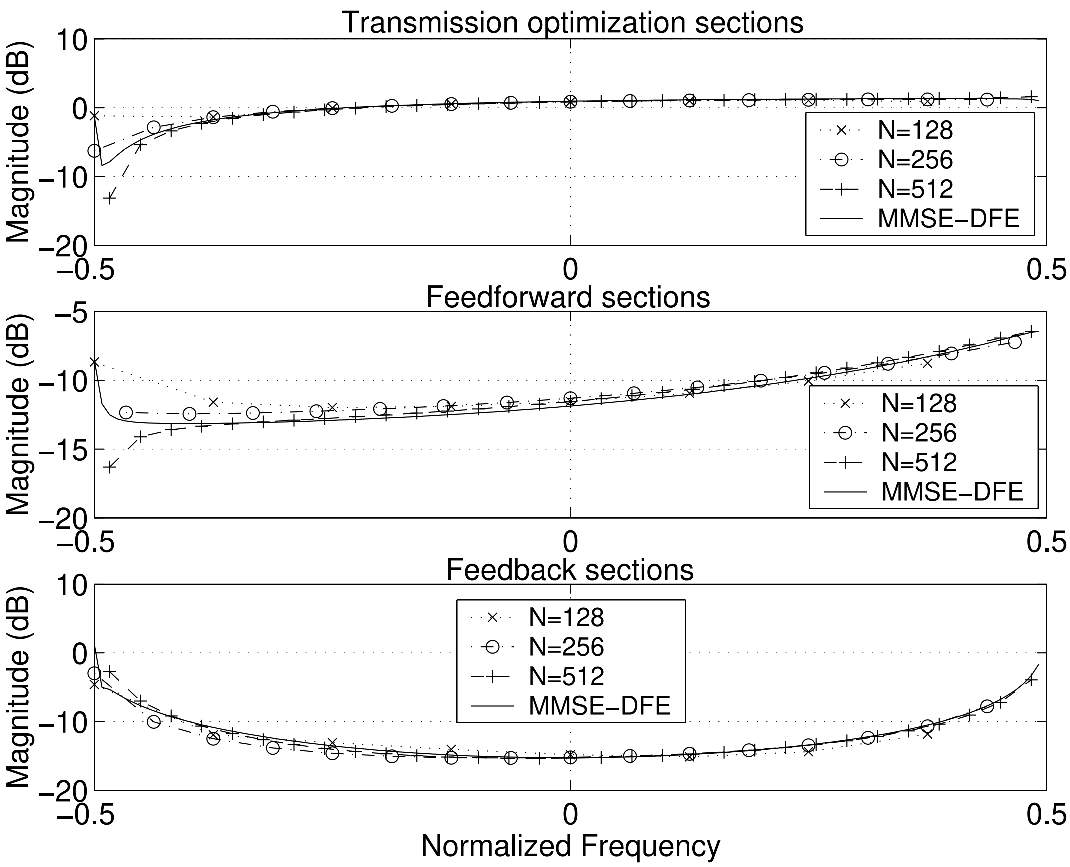

Figure A.7: Illustration of convergence of GDFE in band 3.

Appendix B

Relation of V-BLAST with the Generalized DFE

THIS APPENDIX PROVES that the V-BLAST [40, 47] receiver processing can be interpreted as a special case of the GDFE [42]. Adopting this alternative view has certain benefits: It gives further insight regarding the functions of V-BLAST, provides a link between V-BLAST and the vast accumulated knowledge about DFE's, and easily produces a decomposition into parallel sub-channels that simplifies SNR and capacity calculations.

Section B.1 describes the GDFE algorithm for a multiple antenna channel both with a ZF and with a MMSE criterion. Section B.2 describes the V-BLAST algorithm both for an interference-nulling (ZF) and for a MMSE criterion. Last, Section B.3 proves that the corresponding operations of the GDFE and the V-BLAST algorithms are equivalent.

B.1 GDFE Algorithm

The sampled output of a narrowband MIMO system with an equal number of receive and transmit elements satisfies

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}' \ , \tag{B.1}$$

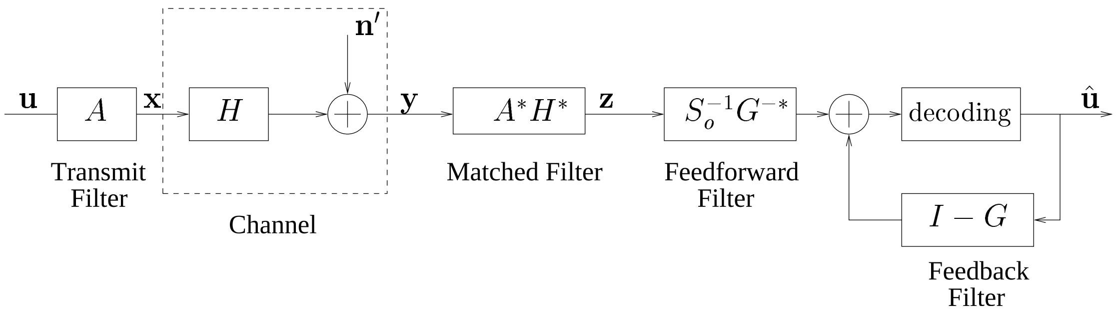

Figure B.1: GDFE equalization structure.

where $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{n}'$ are vectors of the output samples, the transmitted symbols and the noise samples, respectively. Noise is assumed white, Gaussian with variance $N_o/2$. Also, H is an $N \times N$ non-singular matrix with the element h_{ij} denoting the complex gain from transmitter element j to receiver element i.

The GDFE equalization structure is shown in Figure B.1. Although the GDFE in general includes transmit filtering, it is here assumed that $A = I$, so that $\mathbf{u} = \mathbf{x}$ and $\hat{\mathbf{u}} = \hat{\mathbf{x}}$.

B.1.1 ZF-GDFE

The ZF-GDFE is derived in Subsection 3.1.3, where S_o and G are obtained from the Cholesky decomposition of $R_f = H^*H$. However, further simplification is possible using the QR factorization $H = QR$, where Q is unitary and R is upper triangular. Then, $R_f = R^*R$, and $S_o^{1/2} = \operatorname{diag}(r_{11}, r_{22}, \ldots, r_{NN})$, where r_{ij} are the elements of R. Also, $G = S_o^{-1/2}R$, and the elements of G are $g_{ij} = \frac{r_{ij}}{r_{ii}}$. The combined matched-feedforward filter becomes

$$S_o^{-1}G^{-*}H^* = S_o^{-1}S_o^{1/2}R^{-*}R^*Q^* = S_o^{-1/2}Q^* \ , \tag{B.2}$$

which can be interpreted as a rotation/reflection operation performed by Q^*, followed by a scaling operation performed by $S_o^{-1/2}$. The equalization procedure is described as follows:

ZF-GDFE Algorithm

$\tilde{\mathbf{z}} := S_o^{-1/2} Q^* \mathbf{y}$

`for k = 0 to N-1`

$\quad \hat{x}_{N-k} := \texttt{decoding}\left(\tilde{z}_{N-k} - \sum_{i=1}^{k} g_{N-k,N-k+i}\hat{x}_{N-k+i}\right)$

`end`

where $\hat{x}_i$, $\tilde{z}_i$ are the elements of $\hat{\mathbf{x}}$ and $\tilde{\mathbf{z}}$ correspondingly.

B.1.2 MMSE-GDFE

The MMSE-GDFE is derived in Subsection 3.1.2, where S_o and G are obtained from the Cholesky decomposition of $R_b^{-1} = R_f + \frac{N_o}{2} R_{\mathbf{xx}}^{-1}$. In this case, the structure cannot be simplified using the QR factorization. Then, the algorithm becomes:

MMSE-GDFE Algorithm

$\tilde{\mathbf{z}} := S_o^{-1} G^{-*} H^* \mathbf{y}$

`for k = 0 to N-1`

$\quad \hat{x}_{N-k} := \texttt{decoding}\left(\tilde{z}_{N-k} - \sum_{i=1}^{k} g_{N-k,N-k+i}\hat{x}_{N-k+i}\right)$

`end`

The MMSE-GDFE achieves a larger geometric SNR than the ZF-GDFE [42].

B.2 V-BLAST Algorithm

With V-BLAST [40, 47], the symbols are detected sequentially. Prior to the detection of a symbol, interference from previously detected symbols is subtracted, and the received vector elements are linearly weighted to null (or reduce) interference from the yet undetected symbols.

Assume that the channel is again described by (B.1), and $H = [\mathbf{h}_1\, \mathbf{h}_2\, \ldots\, \mathbf{h}_N]$, where $\mathbf{h}_i$ is the i-th column of H. The detection algorithm is described as follows:

V-BLAST Algorithm

$\mathbf{v}_1 := \mathbf{y}$
`for k = 0 to N-1`
  $\hat{x}_{N-k} := \texttt{decoding}\left(\mathbf{w}^*_{N-k}\mathbf{v}_{k+1}\right)$
  $\mathbf{v}_{k+2} = \mathbf{v}_{k+1} - \mathbf{h}_{N-k}\hat{x}_{N-k}$
`end`

Note that reordering operations can be taken into account by rearranging the rows and columns of H. The vector $\mathbf{w}_{N-k}$, performing the linear weighting of the elements of $\mathbf{v}_{k+1}$, is chosen to satisfy either a ZF or an MMSE criterion.

B.2.1 ZF V-BLAST

With the ZF criterion, interference from the yet undetected symbols is nulled. This is compactly expressed by

$$H^*_{N-k}\mathbf{w}_{N-k} = \mathbf{c}_{N-k} \ , \ k = 0, \ldots, N-1 \ , \tag{B.3}$$

where $H_{N-k} = [\mathbf{h}_1\, \mathbf{h}_2\, \ldots\, \mathbf{h}_{N-k}]$ (full-rank for any k), and $\mathbf{c}_i$ is an $i \times 1$ column vector with its i-th element equal to 1, and all other elements 0. The least-norm (and thus noise enhancement minimizing) solution of (B.3) is

$$\mathbf{w}_{N-k} = H_{N-k}\left(H^*_{N-k}H_{N-k}\right)^{-1}\mathbf{c}_{N-k} \ . \tag{B.4}$$

Further simplification is possible with the QR factorization. Then, $H_{N-k} = QR_{N-k}$, $R_{N-k} = [\mathbf{r}_1\ \mathbf{r}_2\ \ldots\ \mathbf{r}_{N-k}]$ and the upper triangular structure of R implies that

$$R_{N-k} = \begin{bmatrix} I_{N-k} \\ O_{k\times(N-k)} \end{bmatrix} R'_{N-k} \ , \tag{B.5}$$

where R'_{N-k} (the top $N-k$ rows of R_{N-k}) is invertible. Equation (B.4) becomes

$$\begin{array}{l} \mathbf{w}_{N-k} = \\ Q\left[I\ O\right]^{T} R'_{N-k}\left(R'^{*}_{N-k}R'_{N-k}\right)^{-1}\mathbf{c}_{N-k} = \\ \left[\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_{N-k}\right] R'^{-*}_{N-k}\mathbf{c}_{N-k} = \\ \mathbf{q}_{N-k}\frac{1}{r^{*}_{N-k,N-k}}\ , \end{array} \tag{B.6}$$

where $\mathbf{q}_i$ is the i-th column of Q.

B.2.2 MMSE V-BLAST

With the MMSE criterion, the objective is to minimize the variance of the error between the transmitted symbol and the receiver's estimate, which leads to the following orthogonality requirement,

$$E\left(x_{N-k}\mathbf{v}^{*}_{k+1}\right) = \mathbf{w}^{*}_{N-k}E\left(\mathbf{v}_{k+1}\mathbf{v}^{*}_{k+1}\right)\ . \tag{B.7}$$

Assuming that decisions are correct with a very high probability, $\hat{x}_{N-k} = x_{N-k}$ and $\mathbf{v}_{k+1} = \sum_{j=1}^{N-k}\mathbf{h}_j x_j + \mathbf{n}'$. With the transmitted symbols being white with variance $\mathcal{E}_x$,

$$E\left(x_{N-k}\mathbf{v}^{*}_{k+1}\right) = \mathcal{E}_x\mathbf{h}^{*}_{N-k}\ , \tag{B.8}$$

$$E\left(\mathbf{v}_{k+1}\mathbf{v}^{*}_{k+1}\right) = \mathcal{E}_x\sum_{j=1}^{N-k}\mathbf{h}_j\mathbf{h}^{*}_j + \frac{N_o}{2}I\ , \tag{B.9}$$

and finally,

$$\mathbf{w}^{*}_{N-k} = \mathbf{h}^{*}_{N-k}\left(\sum_{j=1}^{N-k}\mathbf{h}_j\mathbf{h}^{*}_j + \frac{1}{SNR}I\right)^{-1}\ , \tag{B.10}$$

where $SNR = \frac{\mathcal{E}_x}{No/2}$.

B.3 Equivalence between V-BLAST and GDFE Receiver Operations

The equivalence of the V-BLAST and GDFE receiver processing is demonstrated by proving that the operations performed are identical, and assuming that the ordering of the received substreams is the same for both algorithms. The proofs differ depending on whether the ZF or the MMSE criterion is used.

B.3.1 ZF Case

Employing the fact that $\mathbf{v}_{k+1} = \mathbf{y} - \sum_{i=1}^{k} \mathbf{h}_{N-k+i}\hat{x}_{N-k+i}$, the fundamental operation of the V-BLAST algorithm is

$$\mathbf{w}^*_{N-k}\mathbf{v}_{k+1} = \mathbf{w}^*_{N-k}\left(\mathbf{y} - \sum_{i=1}^{k} \mathbf{h}_{N-k+i}\hat{x}_{N-k+i}\right) \; . \tag{B.11}$$

Recalling that $\tilde{z}_{N-k} = \frac{1}{r_{N-k,N-k}}\mathbf{q}^*_{N-k}\mathbf{y}$, and substituting with (B.6), equation (B.11) becomes

$$\begin{array}{l} \mathbf{w}^*_{N-k}\mathbf{v}_{k+1} = \\ \tilde{z}_{N-k} - \sum_{i=1}^{k} \frac{1}{r_{N-k,N-k}}\mathbf{q}^*_{N-k}\mathbf{h}_{N-k+i}\hat{x}_{N-k+i} \end{array} \; . \tag{B.12}$$

Given the result

$$\mathbf{q}^*_{N-k}\mathbf{h}_{N-k+i} = \mathbf{q}^*_{N-k}Q\mathbf{r}_{N-k+i} = r_{N-k,N-k+i} \; , \tag{B.13}$$

and that $g_{ij} = \frac{r_{ij}}{r_{ii}}$, it is deduced that

$$\mathbf{w}^*_{N-k}\mathbf{v}_{k+1} = \tilde{z}_{N-k} - \sum_{i=1}^{k} g_{N-k,N-k+i}\hat{x}_{N-k+i} \; . \tag{B.14}$$

This equation proves that both ZF V-BLAST and ZF-GDFE make decisions with exactly the same inputs. Therefore, the methods are equivalent and the corresponding post-detection SNR's are equal.

B.3.2 MMSE Case

Starting with (B.10) and making use of $H_{N-k} = H\left[I_{N-k}\ O_{k\times(N-k)}\right]^T$, one finds

$$\begin{array}{l}\mathbf{w}^*_{N-k} = \\ \mathbf{h}^*_{N-k}\left(H_{N-k}H^*_{N-k} + (1/SNR)I\right)^{-1} = \\ \mathbf{c}^T_{N-k}H^*\left\{H\left[I\ O\right]^T\left[I\ O\right]H^* + (1/SNR)I\right\}^{-1} = \\ \mathbf{c}^T_{N-k}\left\{H^*H\left[I\ O\right]^T\left[I\ O\right] + (1/SNR)I\right\}^{-1}H^*\end{array}, \tag{B.15}$$

where $\mathbf{c}_{N-k}$ has the same interpretation as before, but now has size $N \times 1$. Utilizing the fact that $R_b^{-1} = R_f + \frac{1}{SNR}I \Rightarrow G^*S_oG = H^*H + \frac{1}{SNR}I$, (B.15) becomes

$$\begin{array}{l}\mathbf{w}^*_{N-k} = \\ \mathbf{c}^T_{N-k}\left\{\left(G^*S_oG - \frac{1}{SNR}I\right)\begin{bmatrix} I & O \\ O & O\end{bmatrix} + \frac{1}{SNR}I\right\}^{-1}H^*\end{array}. \tag{B.16}$$

Define $S_o = \mathrm{diag}(S_{o,1}, S_{o,2})$, and

$$G = \begin{bmatrix} G_{11} & G_{12} \\ O_{k\times(N-k)} & G_{22}\end{bmatrix}, \tag{B.17}$$

where $S_{o,1}$, $S_{o,2}$ are diagonal with dimensions $(N-k)\times(N-k)$ and $k\times k$ respectively, G_{11}, G_{22} are upper triangular with dimensions $(N-k)\times(N-k)$ and $k\times k$ respectively, and G_{12} is $(N-k)\times k$. Then, the following is derived

$$\begin{array}{l}\mathbf{w}^*_{N-k} = \\ \mathbf{c}^T_{N-k}\left\{G^*\begin{bmatrix} S_{o,1}G_{11} & O \\ O & O\end{bmatrix} + \begin{bmatrix} O & O \\ O & \frac{1}{SNR}I\end{bmatrix}\right\}^{-1}H^* = \\ \mathbf{c}^T_{N-k}\begin{bmatrix} G^*_{11}S_{o,1}G_{11} & O \\ G^*_{12}S_{o,1}G_{11} & \frac{1}{SNR}I\end{bmatrix}^{-1}H^*\end{array}. \tag{B.18}$$

Employing the *block matrix inversion* formula [89, Appendix A],

$$\begin{bmatrix} G_{11}^* S_{o,1} G_{11} & O \\ G_{12}^* S_{o,1} G_{11} & \frac{1}{SNR} I \end{bmatrix}^{-1} = \begin{bmatrix} G_{11}^{-1} S_{o,1}^{-1} G_{11}^{-*} & O \\ -SNR \cdot G_{12}^* G_{11}^{-*} & SNR \cdot I \end{bmatrix} . \tag{B.19}$$

Equation (B.18) now becomes

$$\begin{array}{l} \mathbf{w}_{N-k}^* = \\ \mathbf{c}_{N-k}^T \begin{bmatrix} G_{11}^{-1} S_{o,1}^{-1} G_{11}^{-*} & O \\ -SNR \cdot G_{12}^* G_{11}^{-*} & SNR \cdot I \end{bmatrix} H^* = \\ \mathbf{c}_{N-k}^T \begin{bmatrix} G_{11}^{-1} S_{o,1}^{-1} G_{11}^{-*} & O \\ O & O \end{bmatrix} H^* \end{array} . \tag{B.20}$$

Since G_{11}^{-1} is monic, it is deduced that

$$\mathbf{w}_{N-k}^* = \mathbf{c}_{N-k}^T S_o^{-1} G^{-*} H^* . \tag{B.21}$$

The fundamental operation of MMSE V-BLAST is next examined. Knowing that $\tilde{\mathbf{z}} = S_o^{-1} G^{-*} H^* \mathbf{y}$,

$$\begin{array}{l} \mathbf{w}_{N-k}^* \mathbf{v}_{k+1} = \\ \mathbf{c}_{N-k}^T S_o^{-1} G^{-*} H^* \left(\mathbf{y} - \sum_{i=1}^k \mathbf{h}_{N-k+i} \hat{x}_{N-k+i} \right) = \\ \mathbf{c}_{N-k}^T \left(\tilde{\mathbf{z}} - \sum_{i=1}^k S_o^{-1} G^{-*} H^* \mathbf{h}_{N-k+i} \hat{x}_{N-k+i} \right) = \\ \mathbf{c}_{N-k}^T \left(\tilde{\mathbf{z}} - \sum_{i=1}^k S_o^{-1} G^{-*} H^* H \mathbf{c}_{N-k+i} \hat{x}_{N-k+i} \right) \end{array} . \tag{B.22}$$

Substituting $H^* H$, one finds

$$\begin{array}{l} \mathbf{w}_{N-k}^* \mathbf{v}_{k+1} = \\ \mathbf{c}_{N-k}^T \left[\tilde{\mathbf{z}} - \sum_{i=1}^k S_o^{-1} G^{-*} \left(G^* S_o G - \frac{1}{SNR} I \right) \mathbf{c}_{N-k+i} \hat{x}_{N-k+i} \right] = \\ \mathbf{c}_{N-k}^T \left[\tilde{\mathbf{z}} - \sum_{i=1}^k \left(G - \frac{1}{SNR} S_o^{-1} G^{-*} \right) \mathbf{c}_{N-k+i} \hat{x}_{N-k+i} \right] = \\ \tilde{z}_{N-k} - \sum_{i=1}^k \mathbf{c}_{N-k}^T \left(G - \frac{1}{SNR} S_o^{-1} G^{-*} \right) \mathbf{c}_{N-k+i} \hat{x}_{N-k+i} = \\ \tilde{z}_{N-k} - \sum_{i=1}^k g_{N-k,N-k+i} \hat{x}_{N-k+i} \end{array} . \tag{B.23}$$

Therefore, it becomes evident that the decoder inputs for both MMSE V-BLAST and MMSE-GDFE are exactly the same. The techniques are equivalent and result in the same post-detection SNR's.

It is not hard to show that both techniques (with either the ZF or the MMSE criterion) have approximately the same complexity. Assuming that all system parameters have been pre-computed (as for a stationary channel), the number of flops required for the detection of each block is $\mathcal{O}(N^2)$. If the system parameters need to be re-computed for each transmitted block (as for a block-stationary channel), then for both methods the complexity is $\mathcal{O}(N^3)$.

Appendix C

Code Analysis

SEVERAL OF THE CODES EMPLOYED for communication can be described with a trellis. Such codes include convolutional codes [90], trellis codes [91] and codes defined by partial response (PR) channels [92]. The trellis representation allows for the computation of a number of parameters that are useful for characterizing the performance of a code. This appendix presents an algorithm for the computation of the parameters of a linear code. The underlying method is a forward search on the trellis describing the code.

A linear code is defined as a code where the sum of any two codewords produces some other valid codeword. Convolutional codes and trellis codes belong to this category, but PR codes do not. The linearity property allows the simplification of the search procedure, since one only needs to identify the error event paths with respect to some reference codeword (e.g. the all 0's codeword for convolutional codes). The error event paths with respect to any other codeword are easily derived by the error paths for the reference codeword.

Section C.1 gives a description of the trellis search algorithm. The algorithm produces the following set of parameters:

- Minimum distance between trellis paths: d_{min}
- Number of neighboring paths at minimum distance: N_e
- Number of neighboring paths at next to minimum distances: N_1,N_2,N_3,...

- Number of bit errors corresponding to minimum distance error events: N_b
- Number of bit errors corresponding to next to minimum distance error events: $N_{b,1}$,$N_{b,2}$,$N_{b,3}$,...
- Maximum path length corresponding to a minimum distance error event: L_D
- Maximum path lengths corresponding to next to minimum distance error events: $L_{D,1}$,$L_{D,2}$,$L_{D,3}$,...

where the number of parameters corresponding to error events for next to minimum distances can be set arbitrarily.

This whole set of parameters is described as the "distance spectrum" of the code. Section C.2 includes tables of distance spectra of several popular convolutional codes.

C.1 Algorithm Description

The algorithm performs a forward search on the trellis, until either all the parameters of interest have been determined, or some bound on the number of forward steps has been exceeded.

A matrix L is maintained with rows corresponding to the number of trellis states, and columns corresponding to the parameters:

$$d_{min}, N_e, N_1, N_2, N_3, \ldots, N_b, N_{b,1}, N_{b,2}, N_{b,3}, \ldots$$

This matrix stores information about paths starting from state 0, diverging at the first time instant, and eventually reaching the specified state, under the constraint that the intermediate states are different than 0. The entries contain information with respect to *paths having length equal to the total number of forward steps.* Each time the search progresses forward by one step, L is updated.

Additionally, a vector L_1 is maintained with information about paths starting from state 0, diverging at the first time instant, and reaching again state 0. The entries represent the parameters:

$$d_{min}, N_e, N_1, N_2, N_3, \ldots, N_b, N_{b,1}, N_{b,2}, N_{b,3}, \ldots, L_D, L_{D,1}, L_{D,2}, L_{D,3}, \ldots$$

However, the difference is that these parameters now refer to *all possible paths with length shorter than the number of forward steps*. As before, each time the search moves forward by one step, L_1 is updated.

The algorithm is first described in general terms: The first step, corresponding to the first "divergence" event on the trellis, involves the initialization of L. Then, forward "propagation" steps are performed, until the maximum number of steps has been exceeded, or some condition is satisfied indicating that all quantities of interest have already been determined, and that searching further is useless. At each such propagation step, L is updated for all states. Then, L_1 is also updated making use of the information stored in L regarding state 0. In other words, information regarding paths of length shorter than the current number of steps, that started at state 0, diverged, and then converged back to state 0 is updated by using the corresponding information for paths of length equal to the current number of steps. An example of the algorithm operation is shown in Figure C.1.

It is worthwhile to discuss in more detail the methods used to compute the path parameters. For each state, a polynomial representation can be adopted to convey path information. In order to represent number of neighbors information, the polynomial is defined as

$$g_n(x) = N_e x^{d_{min}} + N_1 x^{d_{min}+1} + N_2 x^{d_{min}+2} + N_3 x^{d_{min}+3} + \ldots \tag{C.1}$$

For representing number of bit errors information, the form of the polynomial becomes

$$g_b(x) = N_b x^{d_{min}} + N_{b,1} x^{d_{min}+1} + N_{b,2} x^{d_{min}+2} + N_{b,3} x^{d_{min}+3} + \ldots \tag{C.2}$$

Observe that the exponents do not actually have to be integers and the exponent increments can be different than 1 (e.g. in trellis codes).

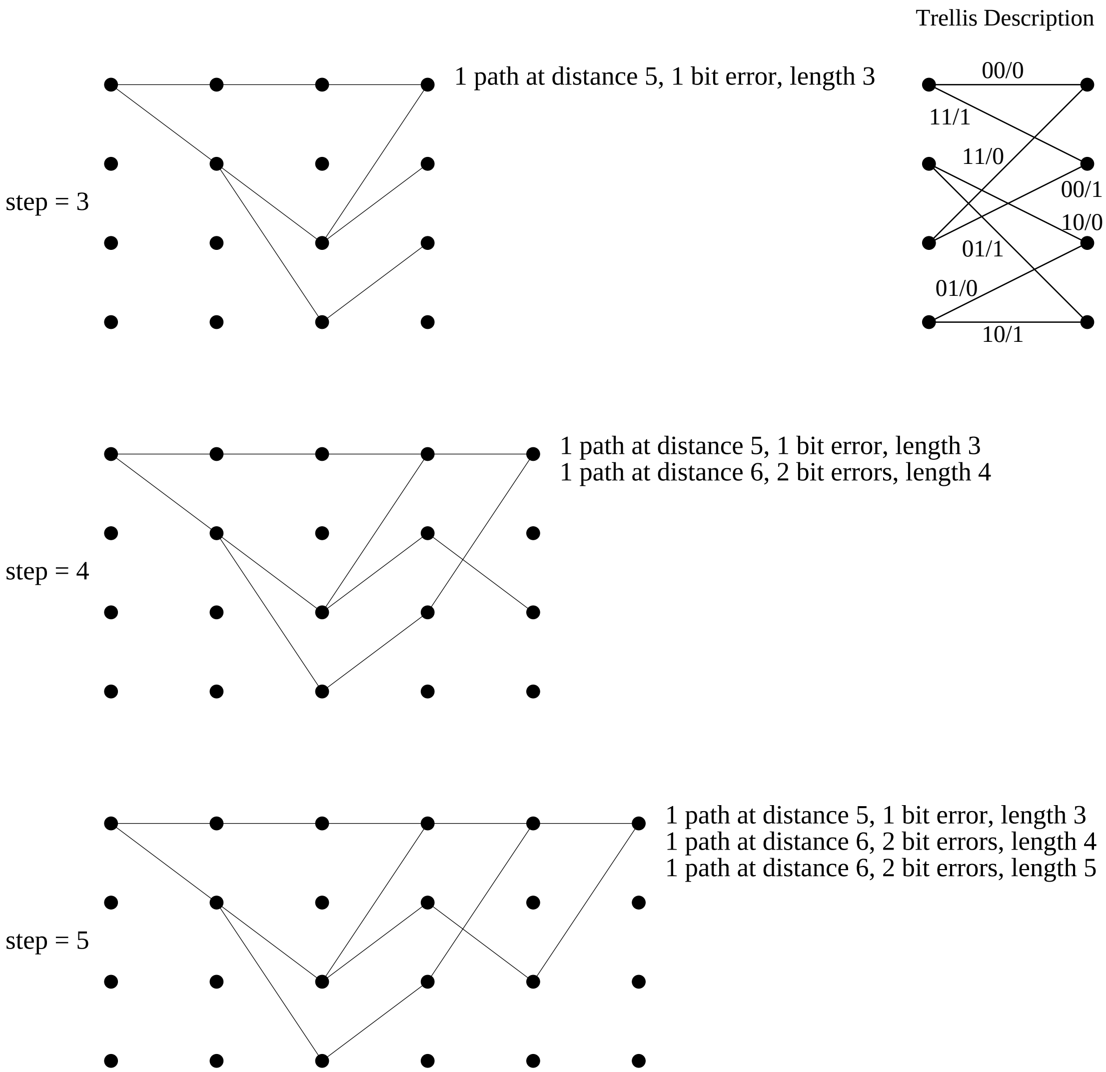

Figure C.1: Example of the algorithm operation (only minimum and next to minimum distance paths are shown). As forward propagation progresses, more paths are discovered.

The advantage of this representation is the ease with which the "embedded" information can be updated. As an example, suppose that the algorithm discovers a new path of distance d with n_b bit errors. Then, the number of neighbors information is updated by adding the term x^d to $g_n(x)$, while the number of bit errors information is updated by adding the term $n_b x^d$ to $g_b(x)$. Also, since only a few lowest order terms are of actual interest, one needs to store only the first exponent and few coefficients, which will correspond to minimum distance and next to minimum distances error events.

Some more algorithmic details can now be provided: The vector L_1 is initialized to all 0's, except for the d_{min} entry, which is set to infinity (or some very large value). Similarly, the matrix L is also initialized to all 0's, except for the d_{min} entries of all states, which are set to infinity. For the first step, it is assumed that the initial state is 0. Then, for each possible input m_0 leading to some state $j \neq 0$, the distance $d_{0,j}$ and the number of bits in error n_{0,j,m_0} can be computed with respect to the path leading to state 0. Each such input corresponds to a branch, and the relevant number of neighbors (or number of bit errors) information can be expressed as a monomial. Adding up all such monomials produces polynomials for each target state $j \neq 0$,

$$g_{n,j}(x) = \sum_{m_0 \in B(0,j)} x^{d_{0,j}} \ , \tag{C.3}$$

and

$$g_{b,j}(x) = \sum_{m_0 \in B(0,j)} n_{0,j,m_0} x^{d_{0,j}} \ , \tag{C.4}$$

where $B(0,j)$ is the set of inputs which lead from state 0 to state j. Accordingly, L is updated.

During the "propagation" step, L is first updated. For each state $i \neq 0$, the branches corresponding to each possible input m_i are considered. Assume that the initial polynomials for state i are $g_{n,i}(x)$ and $g_{b,i}(x)$, and that the branch leading from state i to state j with input m_i has cost $d_{i,j}$ and n_{i,j,m_i} bits in error. Then, the updated polynomials for any state j can be expressed as

$$g_{n,j}^{new}(x) = \sum_{i \neq 0} \sum_{m_i \in B(i,j)} g_{n,i}(x) x^{d_{i,j}} \ , \tag{C.5}$$

and

$$g_{b,j}^{new}(x) = \sum_{i \neq 0} \sum_{m_i \in B(i,j)} \left(g_{b,i}(x) + n_{i,j,m_i} g_{n,i}(x) \right) x^{d_{i,j}} \ , \tag{C.6}$$

where $B(i,j)$ is the set of inputs which lead from state i to state j. Note that multiplying by $x^{d_{i,j}}$ implies a distance increase for all error events, which is the consequence of traversing a branch. Also, adding $g_{b,i}(x)$ with $n_{i,j,m_i} g_{n,i}(x)$ updates the number of bits in error count. Thus, L is updated.

Once the new L has been determined, a new L_1 can be derived fairly easily. Suppose that $g_n(x)$, $g_b(x)$ are the original polynomials expressing the information in L_1, and $g_{n,0}(x)$, $g_{b,0}(x)$ are the polynomials expressing the information about state 0 in the updated L. Then, the entries of L_1 can be updated as dictated by the expressions

$$g_n^{new}(x) = g_n(x) + g_{n,0}(x) \ , \tag{C.7}$$

and

$$g_b^{new}(x) = g_b(x) + g_{b,0}(x) \ . \tag{C.8}$$

Thus, L_1 conveys information about all paths with lengths up to the current number of steps.

The maximum path length information can lastly be determined. If during the last update of L_1, the minimum distance changed, then all lengths are set to the current number of steps. Otherwise, if one of the $N_e, N_1, N_2, N_3, \ldots$ quantities of L_1 changed, then the corresponding distance (one of $L_D, L_{D,1}, L_{D,2}, L_{D,3}, \ldots$) is set to the current number of steps.

If during the update of L, it is observed that the d_{min} for some state i is significantly larger than the currently known d_{min} for state 0, then state i can be ignored for the polynomial calculations. If it is realized that *all* states satisfy this condition, then the forward search can be terminated. The interpretation of "significantly larger" depends on how many next to minimum distances are of interest.

Finally, if the number of forward steps exceeds some preset upper bound, then again the algorithm terminates. This event may be an indication that the code is "catastrophic", which can be interpreted as the case where minimum distance error events have arbitrarily large path lengths.

C.2 Distance Spectra of Convolutional Codes

This section gives the distance spectra of a number of convolutional codes as computed by the code analysis algorithm. Tables C.1, C.2 and C.3 give the generator matrices and the corresponding parameters of maximum free distance codes with rates 1/2, 1/3 and 1/4, where non-systematic implementations are assumed. Tables C.4 and C.5 give the parity matrices and the corresponding parameters of maximum free distance codes with rates 2/3 and 3/4, where systematic implementations are assumed. The generator and parity matrices of the codes have been obtained from [90].

The notation of the tables is briefly outlined. The encoder output is expressed as

$$\mathbf{u}_{out}(D) = \mathbf{u}_{in}(D)G(D) \ , \tag{C.9}$$

where $\mathbf{u}_{out}(D)$, $\mathbf{u}_{in}(D)$ are respectively the output and input vector bit sequences, which are expressed as polynomials. Also, $G(D)$ is the $k \times n$ generator matrix corresponding to a k/n convolutional code. For the non-systematic, rate $1/n$ codes the generator matrices are defined as

$$G(D) = [g_{11}(D) \ \ldots \ g_{1n}(D)] \ , \tag{C.10}$$

where $g_{1i}(D)$, $i = 1, \ldots, n$ are polynomials defined in Tables C.1, C.2 and C.3. For the systematic, rate $(n-1)/n$ codes the generator matrix is defined as

$$G(D) = \begin{bmatrix} 1 & \cdots & 0 & h'_1(D) \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & h'_{n-1}(D) \end{bmatrix} \ , \tag{C.11}$$

and the parity matrix is

$$H(D) = \left[h_1'(D) \ \ldots \ h_{n-1}'(D) \ 1 \right] \ , \tag{C.12}$$

where $h_i'(D) = h_i(D)/h_n(D)$, $i = 1, \ldots, n-1$, and $h_i(D)$, $i = 1, \ldots, n$ are polynomials defined in Tables C.4 and C.5. The polynomials are specified in octal form. For example, 53 translates to 101011, which corresponds to $D^5 + D^3 + D + 1$.

states	$g_{11}(D)$	$g_{12}(D)$	d_{min}	N_e	N_1	N_2	N_3	N_b	$N_{b,1}$	$N_{b,2}$	$N_{b,3}$	L_D	$L_{D,1}$	$L_{D,2}$	$L_{D,3}$
4	7	5	5	1	2	4	8	1	4	12	32	3	5	7	9
8	17	13	6	1	3	5	11	2	7	18	49	5	7	9	11
16	23	35	7	2	3	4	16	4	12	20	72	8	12	13	15
32	77	51	8	2	3	8	15	4	11	36	83	8	15	14	21
64	163	135	10	12	0	53	0	46	0	332	0	16	-	26	-
128	323	275	10	1	6	13	20	6	22	56	136	14	16	21	24
256	457	755	12	10	9	30	51	40	33	196	357	18	26	27	32
512	1337	1475	12	1	8	8	31	2	36	34	187	11	18	20	23
1024	2457	2355	14	19	0	80	0	82	0	530	0	22	-	30	-
2048	6133	5745	14	1	10	25	46	4	46	156	366	19	25	31	35
4096	17663	11271	15	2	10	29	55	6	46	177	386	18	27	30	35
8192	26651	36477	16	5	15	21	56	26	91	128	416	28	28	35	38
16384	46253	77361	17	3	16	44	62	17	78	276	484	27	34	41	41
32768	114727	176121	18	5	15	45	86	26	95	308	650	37	31	38	39
65536	330747	207225	19	9	16	48	112	55	80	344	914	33	33	43	42
131072	507517	654315	20	6	31	58	125	30	183	472	1115	27	34	43	47

Table C.1: Rate 1/2 maximum free distance non-systematic convolutional codes.

states	$g_{11}(D)$	$g_{12}(D)$	$g_{13}(D)$	d_{min}	N_e	N_1	N_2	N_3	N_b	$N_{b,1}$	$N_{b,2}$	$N_{b,3}$	L_D	$L_{D,1}$	$L_{D,2}$	$L_{D,3}$
4	5	7	7	8	2	0	5	0	3	0	15	0	4	-	7	-
8	15	13	17	10	3	0	2	0	6	0	6	0	6	-	7	-
16	25	33	37	12	5	0	3	0	12	0	12	0	8	-	9	-
32	71	65	57	13	1	3	6	4	1	8	26	20	6	10	13	16
64	171	165	133	15	3	3	6	9	7	8	22	44	11	12	13	16
128	365	353	227	16	1	5	2	6	2	13	6	26	10	12	12	16
256	561	325	747	17	1	2	6	7	1	4	20	34	9	11	14	18
512	1735	1063	1257	20	7	0	19	0	19	0	97	0	15	-	21	-
1024	3645	2133	3347	21	4	1	4	14	12	2	16	64	16	12	18	19
2048	6531	5615	7523	22	3	0	9	0	8	0	35	0	15	-	22	-
4096	13471	15275	10637	24	2	8	10	15	6	30	44	71	17	23	23	25
8192	32671	27643	22617	26	7	0	23	0	19	0	113	0	20	-	30	-
16384	47371	51255	74263	27	6	4	6	21	22	18	28	112	24	25	25	28
32768	151711	167263	134337	28	1	6	5	17	2	24	24	103	17	23	24	28
65536	166255	321143	227277	28	1	0	10	0	4	0	44	0	22	-	27	-

Table C.2: Rate 1/3 maximum free distance non-systematic convolutional codes.

states	$g_{11}(D)$	$g_{12}(D)$	$g_{13}(D)$	$g_{14}(D)$	d_{min}	N_e	N_1	N_2	N_3	N_b	$N_{b,1}$	$N_{b,2}$	$N_{b,3}$	L_D	$L_{D,1}$	$L_{D,2}$	$L_{D,3}$
4	7	7	7	5	10	1	1	1	3	2	1	4	9	4	4	7	6
8	17	15	13	13	13	2	1	0	3	4	2	0	10	6	5	-	8
16	37	35	33	25	16	4	0	2	0	8	0	7	0	7	-	8	-
32	73	65	57	47	18	3	0	5	0	6	0	17	0	8	-	11	-
64	163	147	135	135	20	10	0	0	0	37	0	0	0	16	-	-	-
128	367	323	275	271	22	1	4	3	2	2	10	10	8	9	11	13	12
256	751	575	633	627	24	1	3	4	7	2	9	14	29	10	13	14	16
512	0671	1755	1353	1047	26	3	0	4	0	6	0	10	0	12	-	13	-
1024	3321	2365	3643	2277	28	4	0	5	0	9	0	14	0	16	-	18	-
2048	7221	7745	5223	6277	30	4	0	4	0	9	0	14	0	15	-	18	-
4096	15531	17435	05133	17627	32	4	0	9	0	13	0	30	0	17	-	21	-
8192	23551	25075	26713	37467	34	1	0	11	0	3	0	38	0	18	-	22	-
16384	66371	50575	56533	51447	37	3	5	6	10	7	18	26	44	19	22	24	23
32768	176151	123175	135233	156627	39	5	7	10	4	17	26	48	18	21	26	25	24
65536	247631	264335	235433	311727	41	3	7	7	10	7	24	33	46	20	24	27	26

Table C.3: Rate 1/4 maximum free distance non-systematic convolutional codes.

states	$h_1(D)$	$h_2(D)$	$h_3(D)$	d_{min}	N_e	N_1	N_2	N_3	N_b	$N_{b,1}$	$N_{b,2}$	$N_{b,3}$	L_D	$L_{D,1}$	$L_{D,2}$	$L_{D,3}$
8	17	15	13	4	1	5	24	71	3	15	98	331	2	4	7	9
16	05	23	27	5	7	23	59	240	26	91	276	1329	7	9	11	15
32	53	51	65	6	9	19	80	276	38	94	432	1711	9	11	15	19
64	121	113	137	6	1	17	47	145	6	78	241	865	5	13	15	18
128	271	257	265	7	6	26	105	378	27	141	635	2528	11	15	19	22
256	563	601	475	8	8	40	157	598	44	226	1034	4323	9	18	22	30
512	1405	1631	1333	8	1	20	75	243	6	120	488	1815	7	15	20	22
1024	1347	3641	3415	9	9	45	166	593	56	313	1199	4755	15	19	26	30
2048	5575	7377	4033	10	29	0	473	0	200	0	3759	0	18	-	29	-
4096	14107	13125	11203	10	4	34	127	450	27	248	1010	3848	14	19	24	27
8192	20535	31637	27773	11	9	72	222	824	66	564	1949	7652	16	27	27	36
16384	66147	41265	57443	12	58	0	847	0	469	0	7857	0	26	-	31	-
32768	100033	167575	155377	12	25	0	462	0	193	0	4258	0	23	-	30	-
65536	353431	300007	267063	12	7	26	120	437	56	227	1098	4344	22	25	33	37

Table C.4: Rate 2/3 maximum free distance systematic convolutional codes.

states	$h_1(D)$	$h_2(D)$	$h_3(D)$	$h_4(D)$	d_{min}	N_e	N_1	N_2	N_3	N_b	$N_{b,1}$	$N_{b,2}$	$N_{b,3}$	L_D	$L_{D,1}$	$L_{D,2}$	$L_{D,3}$
32	45	71	63	51	5	4	28	161	713	16	126	875	4362	3	8	10	14
64	151	121	177	155	6	24	61	528	1313	112	321	3118	9383	8	10	15	17

Table C.5: Rate 3/4 maximum free distance systematic convolutional codes.

Bibliography

[1] G. A. Miller, "The magical number 7, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, vol. 63, pp. 81–97, Mar. 1956.

[2] ANSI T1.413-1998, *Network to Customer Installation Interfaces - Asymmetric Digital Subscriber Line (ADSL) Metallic Interface*.

[3] J. M. Cioffi, V. Oksman, J.-J. Werner, T. Pollet, P. M. P. Spruyt, J. S. Chow, and K. S. Jacobsen, "Very-high-speed digital subscriber lines," *IEEE Commun. Mag.*, vol. 37, pp. 72–79, Apr. 1999.

[4] ANSI T1E1.4 Trial Use Standard, *Interface between Networks and Customer Installation - Very High Bit-rate Digital Subscriber Line (VDSL) Metallic Interface, Part 1: Functional Requirements and Common Specification*.

[5] IEEE Std 802.3-2000, *Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*.

[6] IEEE Std 802.3ab-1999, *Physical Layer Parameters and Specification for 1000Mb/s Operation Over 4-Pair of Category 5 Balanced Copper Cabling, Type 1000BASE-T*.

[7] W. J. Dally and J. Poulton, "Transmitter equalization for 4-Gbps signaling," *IEEE Micro*, vol. 17, pp. 48–56, Jan. 1997.

[8] R. Farjad-Rad, C.-K. K. Yang, and M. A. Horowitz, "A 0.3μm CMOS 8Gs/s 4-PAM serial link transceiver," *IEEE J. Solid State Circ.*, vol. 35, pp. 757–764, May 2000.

[9] J. L. Zerbe, P. S. Chau, C. W. Werner, W. F. Stonecypher, H. J. Liaw, G. J. Yeh, T. P. Thrush, S. C. Best, and K. S. Donnelly, "A 2GS/s/pin 4-PAM parallel bus interface with transmit crosstalk cancellation, equalization, and integrating receivers," in *IEEE Solid State Circ. Conf., Dig. of Tech. Papers*, (San Francisco, CA), pp. 66–67, Feb. 2001.

[10] T. Koyama, J. Peng, and P. Cohen, "A multi-dimensional equalizer for Gigabit Ethernet," in *Proc. IEEE Int. Conf. Commun.*, (Helsinki, Finland), pp. 509–512, June 2001.

[11] V. Stojanovic, G. Ginis, and M. A. Horowitz, "Transmit pre-emphasis for high-speed time-division-multiplexed serial-link transceiver," in *Proc. IEEE Int. Conf. Commun.*, (New York, NY), May 2002.

[12] T. Starr, J. M. Cioffi, and P. J. Silvermann, *Understanding Digital Subscriber Line Technology.* Upper Saddle River, NJ: Prentice-Hall, 1999.

[13] F. Sjöberg, M. Isaaksson, R. Nilsson, P. Ödling, S. K. Wilson, and P. O. Börjesson, "Zipper: A duplex method for VDSL based on DMT," *IEEE Trans. Commun.*, vol. 47, pp. 1245–1252, Aug. 1999.

[14] K. J. Kerpez, "Near-end crosstalk is almost gaussian," *IEEE Trans. Commun.*, vol. 41, pp. 670–672, Jan. 1993.

[15] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, (Denver, CO), pp. 964–967, Apr. 1980.

[16] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE Trans. Commun. Tech.*, vol. 19, pp. 626–634, Oct. 1971.

[17] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[18] ANSI T1.417-2001, *Spectrum Management for Loop Transmission Systems.*

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, NY: Wiley, 1991.

[20] S. Kasturia, *Vector Coding for Digital Communication on Spectrally Shaped Channels.* Ph.D. Thesis, Stanford University, Mar. 1989.

[21] J. T. Aslanis, *Coding for Communications Channels with Memory.* Ph.D. Thesis, Stanford University, Nov. 1989.

[22] S. Kasturia, J. Aslanis, and J. M. Cioffi, "Vector coding for partial-response channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 741–762, July 1990.

[23] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multi-input-multi-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, pp. 3221–3231, Dec. 1994.

[24] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inform. Theory (Tsahkadsor, Armenian S.S.R.)*, pp. 23–52, Prague, 1971. Publishing House of the Hungarian Academy of Sciences.

[25] S. Verdu, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Allerton Conf. Commun., Contr. and Comput.*, (Allerton, IL), pp. 1025–1034, Oct. 1986.

[26] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative waterfilling for Gaussian vector multiple access channels," in *Proc. IEEE Int. Symp. Inform. Theory*, (Washington, DC), p. 322, June 2001.

[27] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2524–2530, Oct. 1998.

[28] G. Caire and S. Shamai, "On achievable rates in a multi-access Gaussian broadcast channel," in *Proc. IEEE Int. Symp. Inform. Theory*, (Washington, DC), p. 147, June 2001.

[29] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with applications to DSL systems," in *Proc. Asilomar Conf. Signals, Systems, Comp.*, (Pacific Grove, CA), pp. 1627–1631, Oct. 2000.

[30] G. Ginis and J. M. Cioffi, "Vectored-DMT: A FEXT canceling modulation scheme for coordinating users," in *Proc. IEEE Int. Conf. Commun.*, (Helsinki, Finland), pp. 305–309, June 2001.

[31] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," in *Proc. IEEE Int. Symp. Inform. Theory and Appl.*, (Honolulu, HI), pp. 681–684, Nov. 2000.

[32] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," in *Proc. IEEE Int. Symp. Inform. Theory*, (Lausanne, Switzerland), July 2002.

[33] A. Duel-Hallen, "Equalizer for multiple input/multiple output channels and PAM systems with cyclostationary input sequences," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 630–639, Apr. 1992.

[34] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1334–1347, Sept. 1994.

[35] N. Al-Dhahir and A. H. Sayed, "The finite length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Proc.*, vol. 48, pp. 2921–2936, Oct. 2000.

[36] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous code-division multiple-access channel," *IEEE Trans. Commun.*, vol. 41, pp. 285–290, Feb. 1993.

[37] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian

multiple-access channel," in *Proc. Asilomar Conf. Signals, Systems, Comp.*, (Pacific Grove, CA), pp. 1405–1409, Nov. 1997.

[38] M. K. Varanasi, "Decision feedback multiuser detection: a systematic approach," *IEEE Trans. Inform. Theory*, vol. 45, pp. 219–240, Jan. 1999.

[39] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antenna," *Bell Syst. Tech. J.*, vol. 1, pp. 41–59, Autumn 1996.

[40] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1841–1852, Nov. 1999.

[41] N. Al-Dhahir and J. M. Cioffi, "Block transmission over dispersive channels: Transmit filter optimization and realization, and MMSE-DFE receiver performance," *IEEE Trans. Inform. Theory*, vol. 42, pp. 137–160, Jan. 1996.

[42] J. M. Cioffi and G. D. Forney, *Communication, Computation, Control and Signal Processing*, ch. Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise, pp. 79–127. Boston, MA: Kluwer, 1997. Editors: A. Paulraj and V. Roychowdhury and C. Shaper.

[43] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. 33, pp. 665–75, July 1985.

[44] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding - Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, pp. 2582–2594, Oct. 1995.

[45] G. Ginis and J. M. Cioffi, "On the convergence of the generalized DFE to the MMSE-DFE," in *Proc. IEEE GlobeCom*, (San Antonio, TX), pp. 1327–1335, Nov. 2001.

[46] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[47] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high bit rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Systems, Electronics*, (Pisa, Italy), pp. 295–300, Oct. 1998.

[48] G. Ginis and J. M. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Comm. Lett.*, vol. 5, pp. 364–366, Sept. 2001.

[49] A. Agrawal, G. Ginis, and J. M. Cioffi, "Channel diagonalization through orthogonal space-time coding," in *Proc. IEEE Int. Conf. Commun.*, (New York, NY), May 2002.

[50] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 744–765, July 1999.

[51] V. Joshi and D. D. Falconer, "Sequence estimation techniques for digital subscriber loop transmission with crosstalk interference," *IEEE Trans. Commun.*, vol. 38, pp. 1367–1374, Sept. 1990.

[52] M. Abdulrahman and D. D. Falconer, "Cyclostationary crosstalk suppression by decision feedback equalization on digital subscriber loops," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 640–649, Apr. 1992.

[53] G. D. Golden, J. E. Mazo, and J. Salz, "Transmitter design for data transmission in the presence of a data-like interferer," *IEEE Trans. Commun.*, vol. 43, pp. 837–850, Feb. 1995.

[54] P. S. Kumar and S. Roy, "Optimization for crosstalk suppression with noncoordinating users," *IEEE Trans. Commun.*, vol. 44, pp. 894–905, July 1996.

[55] M. L. Honig, K. Steiglitz, and B. Gopinath, "Multichannel signal processing for data communications in the presence of crosstalk," *IEEE Trans. Commun.*, vol. 38, pp. 551–558, Apr. 1990.

[56] M. L. Honig, P. Crespo, and K. Steiglitz, "Suppression of near- and far-end crosstalk by linear pre- and post-filtering," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 614–629, Apr. 1992.

[57] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE J. Select. Areas Commun.*, vol. 20, June 2002.

[58] B. R. Vojčić and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. Commun.*, vol. 46, pp. 1346–1355, Oct. 1998.

[59] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *IEEE Electron. Lett.*, vol. 7, pp. 138–139, Mar. 1971.

[60] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774–780, Aug. 1972.

[61] G. D. Forney and M. V. Eyuboglu, "Combined equalization and coding using precoding," *IEEE Commun. Mag.*, vol. 29, pp. 25–34, Dec. 1991.

[62] G. D. Forney, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, pp. 281–300, Mar. 1992.

[63] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE GlobeCom*, (San Antonio, TX), pp. 1344–1348, Nov. 2001.

[64] G. H. Golub and C. F. V. Loan, *Matrix Computations.* Johns Hopkins University Press, 3 ed., 1996.

[65] M. Barton and M. L. Honig, "Optimization of discrete multitone to maintain spectrum compatibility with other transmission systems on twisted copper pairs," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1558–1563, Dec. 1995.

[66] A. Sendonaris, V. V. Veeravalli, and B. Aazhang, "Joint signaling strategies for approaching the capacity of twisted-pair channels," *IEEE Trans. Commun.*, vol. 46, pp. 673–685, May 1998.

[67] W. Yu, G. Ginis, and J. M. Cioffi, "An adaptive multiuser power control algorithm for VDSL," in *Proc. IEEE GlobeCom*, (San Antonio, TX), pp. 394–398, Nov. 2001.

[68] W. Yu, G. Ginis, and J. M. Cioffi, "An adaptive multiuser power control algorithm for VDSL," *IEEE J. Select. Areas Commun.*, vol. 20, June 2002.

[69] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding - Part II: Coding results," *IEEE Trans. Commun.*, vol. 43, pp. 2595–2604, Oct. 1995.

[70] R. Price, "Nonlinearly feedback equalized PAM versus capacity for noisy filter channels," in *Proc. IEEE Int. Conf. Commun.*, (Philadelphia, PA), pp. 22.12–22.17, June 1972.

[71] K. Jacobsen, "Methods of upstream power backoff on very high-speed digital subscriber lines," *IEEE Commun. Mag.*, vol. 39, pp. 210–216, Mar. 2001.

[72] C. Zeng, C. Aldana, A. Salvekar, and J. M. Cioffi, "Crosstalk identification in xDSL systems," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1488–1496, Aug. 2001.

[73] C. H. Aldana and J. M. Cioffi, "Channel tracking for multiple input, single output systems using EM algorithm," in *Proc. IEEE Int. Conf. Commun.*, (Helsinki, Finland), pp. 586–590, June 2001.

[74] C. Y. Wong, R. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1747–1758, Oct. 1999.

[75] W. Yu and J. M. Cioffi, "FDMA capacity of the gaussian multiple access channel with ISI," *IEEE Trans. Commun.*, vol. 50, pp. 102–111, Jan. 2002.

[76] L. M. C. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Asymptotic FDMA capacity region for broadcast channels with ISI," in *Proc. IEEE Int. Conf. Commun.*, (Helsinki, Finland), pp. 1648–1653, June 2001.

[77] ANSI T1E1.4 Trial Use Standard, *Interface between Networks and Customer Installation - Very High Bit-rate Digital Subscriber Line (VDSL) Metallic Interface, Part 3: Technical Specification for Multi-Carrier Modulation (MCM) Transceivers.*

[78] A. Leke and J. M. Cioffi, "Impact of imperfect channel knowledge on the performance of multicarrier systems," in *Proc. IEEE GlobeCom*, (Sydney, Australia), pp. 951–955, Nov. 1998.

[79] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming.* Duxbury Press/Brooks/Cole Publishing Company, 1993.

[80] B. A. Murtagh and M. A. Saunders, "MINOS 5.5 User's Guide," Tech. Rep. SOL 83-20R, Dept. of Operations Research, Stanford University, July 1998.

[81] D. S. Chen and S. Roy, "An adaptive multiuser receiver for CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 808–816, June 1994.

[82] P. B. Rapajic and S. Vucetic, "Adaptive receiver structures for asynchronous CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 685–697, May 1994.

[83] P. B. Rapajic, M. L. Honig, and G. Woodward, "Multiuser decision-feedback detection: Performance bounds and adaptive algorithms," in *Proc. IEEE Int. Symp. Inform. Theory*, (Cambridge, MA), p. 34, Aug. 1998.

[84] A. M. Tehrani, B. Hassibi, and J. M. Cioffi, "An estimation-based approach to multiple-input multiple-output (MIMO) channel equalization," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, (Cambridge, MA), pp. 27–31, Mar. 2000.

[85] G. Ginis, Y. Bar-Ness, and J. M. Cioffi, "Blind adaptive MIMO decision feedback equalization using Givens rotations," in *Proc. IEEE Int. Conf. Commun.*, (New York, NY), May 2002.

[86] Y. Bar-Ness and J. B. Punt, "Adaptive bootstrap CDMA multi-user detector," *Wireless Personal Communications*, vol. 3, pp. 55–71, 1996.

[87] H. Ge and Y. Bar-Ness, "Comparative study of the linear minimum mean squared error (LMMSE) and the adaptive bootstrap multi-user detectors for CDMA," in *Proc. IEEE Int. Conf. Commun.*, (Dallas, TX), pp. 78–82, June 1996.

[88] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. 18, pp. 725–730, Nov. 1972.

[89] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation.* Upper Saddle River, NJ: Prentice Hall, 2000.

[90] R. Johannesson and K. S. Sigangirov, *Fundamentals of Convolutional Coding.* New York NY: IEEE Press, 1999.

[91] E. Biglieri, D. Divsalar, P. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications.* New York, NY: McMillan, 1991.

[92] H. Kobayashi, "Application of probabilistic decoding to digital magnetic recording systems," *IBM J. Res. Develop.*, vol. 15, pp. 64–74, Jan. 1971.