A Tensor Framework for Multi-Domain Communication Systems

Adithya Venugopal

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

November 2019

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering (Thesis).

2019/11/14

Abstract

The demand for mobile data is likely to grow at a pace more than envisaged in the coming years. Further, as applications such as the internet of things (IoT) come to fruition, there will be increased diversity in the types of devices demanding internet connectivity and their requirements. Significant increase in data rate requirements are also expected due to sensitive services such as Ultra High Definition (UHD) video streaming and cloud computing. To meet all these demands, physical layer waveform candidates for future generations of communications need to be robust and inherently capable of extending into multiple domains (space, time, frequency, users, transmission media, code etc.) to ensure efficient utilization of resources. Multiple domains can be innately integrated into the design process of modulation schemes by using tensors, which are multi-way arrays.

This thesis introduces a unified tensor framework, which is the foundation for multi-domain communication systems that can be used to represent, design and analyse schemes that span several domains. In our work, transmitted signals are represented by Nth order signal tensors which are coupled, using a system tensor of order $N + M$, with the received signals which are represented by another signal tensor of order M through the contracted convolution. We begin with the continuous time representation of the tensor system model and present both the strict multi-domain generalization of the Nyquist criterion for zero interference (inter-tensor and intra-tensor interference) as well as a relaxation. We present an equivalent discrete time system model and derive tensor based linear and non-linear equalization methods to combat multi-domain interference for criteria such as minimum mean squared error and minimum peak distortion. Lastly, we present the multi-domain generalization of partial response signalling, or correlative coding, where controlled interference is introduced into the design process to improve performance.

Sommaire

La demande de données mobiles devrait croître à un rythme plus rapide que prévu dans les années à venir. En outre, à mesure que des concepts tels que l′Internet des objets (IoT) se concrétiseront, les types d′appareils nécessitant une connectivité Internet et leurs exigences se diversifieront. Une augmentation significative des besoins en débit de données est également attendue en raison de services sensibles tels que le streaming vidéo UHD (Ultra High Definition) et le cloud computing. Pour répondre à toutes ces demandes, les candidats aux formes d′onde de la couche physique pour les futures générations de communications doivent être robustes et capables de s′étendre à de multiples domaines (espace, temps, frquence, utilisateurs, supports de transmission, code, etc.) afin de garantir une utilisation efficace des ressources. Plusieurs domaines peuvent être intégrés de manière innée au processus de conception de schémas de modulation en utilisant des tenseurs, qui sont des tableaux à plusieurs voies.

Dans ce travail, nous introduisons un cadre tenseur unifié, qui constitue la base des systèmes de communication multi-domaines pouvant tre utilisés pour représenter, concevoir et analyser des systémes couvrant plusieurs domaines. Dans notre travail, les signaux transmis sont représentés par des tenseurs de signaux du nième ordre qui sont couplés, à l′aide d′un tenseur de système d′ordre N + M, aux signaux reus qui sont représentés par un autre tenseur de signaux d′ordre M par la convolution contractée. Nous commençons par la représentation temporelle continue du modèle du système tensoriel et présentons à la fois la généralisation multi-domaine stricte du critère de Nyquist pour l′interférence zéro (intra-tenseur et l′inter-tenseur), ainsi qu′une relaxation. Nous présentons un modèle de système à temps discret équivalent et en déduisons des méthodes d′égalisation linéaires et non linéaires basées sur le tenseur pour lutter contre les interférences multi-domaines pour des critères tels que l′erreur quadratique moyenne minimale et la distorsion de crte minimale. Enfin, nous présentons la généralisation multi-domaine de la signalisation à réponse partielle, ou codage corrélatif, dans laquelle une interférence contrôlée est introduite dans le processus de conception pour améliorer les performances.

Acknowledgments

I would like to begin by extending my utmost gratitude towards my supervisor, Prof. Harry Leib, for his guidance and support throughout the duration of my Masters' studies. I would like to thank him for helping me hone my research skills, providing constant and meaningful feedback regarding my work and for guiding me towards the completion of this thesis. I feel that I have grown as a researcher and as an individual over the past few years and am greatful to have had the opportunity to work with him. I would also like to thank Prof. Harry Leib for his financial support from his NSERC grants.

I would like to thank my fellow labmates, especially Divyanshu Pandey and Xiaoxian Hou, for the stimulating discussions, motivation and friendship throughout the duration of my studies. Finally, I would like to thank my father R. Venugopal and my mother Chitra Ratnam Venugopal for constantly pushing me to strive for greater heights throughout my life. None of this would be possible without their unrelenting love and support.

Contents

List of Figures

List of Tables

List of Acronyms

16-QAM	16-point Quadrature Amplitude Modulation
3GPP	Third Generation Partnership Project
MIMO	Multiple Input Multiple Output
OFDM	Orthogonal Frequency Division Multiplexing
FBMC	Filter Bank Multi Carrier
GFDM	Generalized Frequency Division Multiplexing
SISO	Single Input Single Output
TPRS	Tensor Partial Response Signaling

Notation

$(.)^{-1}$	Inverse of the argument
$(.)^{+}$	Pseudo-inverse of the argument
$(.)^{H}$	Conjugate transpose of the argument
$(.)^{T}$	Transpose of the argument
$\|.\|_F$	Frobenius norm of the argument
$\mathbf{a}$	Column vector $\mathbf{a}$
$\mathbf{A}$	Matrix $\mathbf{a}$
$\boldsymbol{\mathcal{A}}$	Tensor $\boldsymbol{\mathcal{A}}$
$\boldsymbol{\mathcal{A}}_{i,j,k}$	Element (i, j, k) of $\boldsymbol{\mathcal{A}}$
$\boldsymbol{\mathcal{A}}(t)$	A signal or system tensor
$\boldsymbol{\mathcal{A}}[k]$	A discrete tensor sequence
$\mathbf{0}_T$	The all zero tensor
$\breve{\boldsymbol{\mathcal{A}}}(f)$	The Fourier transform of $\boldsymbol{\mathcal{A}}(t)$
$\breve{\boldsymbol{\mathcal{A}}}(D)$	The D-transform of $\boldsymbol{\mathcal{A}}[k]$
$\mathbb{C}$	set of complex numbers
$\mathbb{R}$	set of real numbers
$\mathbb{C}_x$	set of complex functions of x
$\Re\{\}$	real part of
$\Im\{\}$	imaginary part of

Chapter 1

Introduction

Wireless communications and the internet have been two of the most disruptive technologies in recent history and the synergetic relationship between them has led to exponential demand for mobile communication services. As presented in the visual network index (VNI) report released by Cisco [1], the amount of wireless data has exploded and it is predicted to continue growing exponentially in the coming years. Significant increase in data rate requirements are expected due to sensitive services such as Ultra High Definition (UHD) video streaming and cloud computing. Hence, future generations of wireless communications will need to provide data rates that are orders of magnitude higher than current 4G technologies. Moreover, with the internet of things (IoT) poised to become a reality, many diverse devices with an eclectic mix of requirements will soon demand wireless connectivity to the internet. In order to service such a vast audience while constrained by radio spectrum scarcity, future communication systems will need to be extremely bandwidth efficient. Given these demands, it is clear that a paradigm shift is required in communication systems of the coming generations (5G and beyond) since incremental improvements on current (4G) systems will not suffice [2].

The use of additional domains in the design process of a communication system is an important means to improve its performance via added robustness from diversity or higher data rates from multiplexing. For example, the addition of the space domain through the utilization of multiple inputs and multiple outputs (MIMO) was the logical successor of single input single output (SISO) systems. MIMO systems boast improved link performance as in the case of space-time coding [3] or higher data rates via spatial multiplexing such as V-BLAST [4]. Multicarrier (MC) systems such as OFDM, GFDM and FBMC are examples of frequency domain utilization and are significant improvements over singlecarrier (SC) systems. The two-dimensional structure of these systems are well represented through the use of matrices. Following this trend, it is crucial that waveform candidates for future generations of wireless communications be natively capable of extending into *multiple* domains (space, time, frequency, and users to name a few) to ensure efficient utilization of resources. The use of tensors, which are multidimensional arrays [5], allows innate integration of several domains into the design process of modulation schemes.

The notion of tensors and tensor decompositions date back to 1927 with the work of Hitchcock [6]. Cattell [7] is credited for introducing the notion of the multi-way model. However, tensors and their decompositions first gained popularity in psychometrics literature through the works of Tucker [8] and Carroll and Chang [9]. Since then, tensors have been extensively used in chemometrics in the food industry, in Fluorescence spectroscopy and flow injection analysis[10, 11, 12]. In the last years, tensor applications have gained significant interest in varied fields such as signal processing [13, 14], data mining [15], graph analysis [16], neuroscience and computer vision [17, 18]. A tensor approach for multidimensional data filtering is presented in [19]. Cumulant-Based Blind Identification of Under-determined Mixtures are explored in [20]. A comprehensive overview of multi-linear algebra, tensor products and their decompositions are provided in [5]. Solution of multi-linear equations using tensor inversion is studied in [21] and a higher-order generalization of the Moore-Penrose pseudo-inverse is derived in [22]. The notion of the various transposes of a tensor is presented in [23].

Matrix decompositions are not unique in general, meaning that a particular matrix may be decomposed in a number of different ways. In order to ensure uniqueness of a matrix decomposition, additional constraints such as positive-definiteness or orthonormality must be imposed. In contrast, such strong constraints are not required for a tensor to offer a unique decomposition due to the use of higher dimensions [24, 25]. This is one of the reasons for the gain in popularity of tensor based approached in wireless communications over recent years. A blind receiver using PARAFAC decompositions for DS-CDMA systems is considered in [26]. Multiple invariance sensor array processing (MI-SAP) is linked to parallel factor (PARAFAC) analysis for both data-domain and subspace formulations in [27]. A blind receiver that uses tensor decompositions for SIMO and MIMO OFDM systems is presented in [28]. A space-time coding model based on a Khatri-Rao product, dubbed KRST, was derived by combining spatial multiplexing and temporal spreading through linear pre-coding and linear post-coding respectively [29]. A tensor based receivers for MIMO communication systems is presented in [30] and [31]. In [32], it is shown that the received signal in oversampled CDMA and OFDM has a multidimensional structure and a constrained Block-PARAFAC model is used for blind equalization where the constraints of the tensor model vary based on the system that is being used. Three dimensional tensors are used to combine space-time coding with spatial multiplexing, dubbed space-time multiplexing (STM) coding, in [33]. Two constrained tensor models dubbed the PARATUCK-(N_1, N) and Tucker-(N_1, N) are introduced in [34], which are then used to derive semi-blind receivers for MIMO OFDM-CDMA systems. A modified alternating least squares (ALS) algorithm for estimating the matrix factors of the Kronecker product is considered in [35], that is used for the design of MIMO wireless communication systems using tensor modelling. Multidimensional Weiner filtering, where the n-mode unfolding of the desired signal is expressed as a weighted combination of orthogonal vectors from the n-mode signal subspace basis is used to determine the theoretical expression of the n-mode Weiner filter, is described in [19].

1.1 Thesis Contribution

This thesis presents a unified tensor framework for multi-domain communication systems. Here, the transmitted signal is represented by an Nth order tensor and the received signal is represented by an Mth order tensor. The transmitted signal tensor is coupled with the received signal tensor by a system tensor of order $M + N$ using either the contracted convolution (continuous time systems) or the contracted product (discrete time systems). Using this framework, we present the foundations for a multi-domain point-to-point communication system that can be used to represent, design and analyse future wireless or wired communication systems. In the formulation of the tensor framework, the mathematical domains of the signal and system tensors are not associated to physical domains. This mapping is instead performed on a per application basis as required. This abstraction makes the framework more general and hence allows a straightforward implementation of a variety of communication systems.

Further, we present both the strict multi-domain generalization of the Nyquist criterion for zero interference (inter-tensor and intra-tensor interference) as well as a relaxation. Tensor based linear and non-linear equalization schemes for multi-domain interference for metrics such as minimum mean squared error and minimum peak distortion are derived. To demonstrate the efficacy of the tensor framework examples such as OFDM, GFDM and FBMC are used to show how this framework can be employed to add additional domains into the design process.

Finally, we present a method to allow a controlled amount of interference in order to achieve improved data rates and spectral shaping. This is a multi-domain generalization of partial response signaling, or correlative coding [36, 37] that is dubbed Tensor Correlative Coding.

1.2 Thesis Outline

Apart from the introduction above and the concluding remarks, this thesis consists of four main chapters. The contents of these chapters are summarized as follows

Chapter 2. This chapter introduces tensors and some relevant tensor based operations. The concept of signal tensors, which are tensors of time functions, are introduced along with their transformations. Using the above preliminaries, the system model in this tensor framework is described. A higher-order generalization of the Nyquist Criterion for zero inter-symbol interference is derived.

Chapter 3. This chapter models some of the existing waveform using the tensor framework such as the 5G selected waveform OFDM, filter bank multicarrier (FBMC) and generalized frequency division multiplexing (GFDM). Insights are provided on how modifications can be made to these waveforms based on the tensor framework.

Chapter 4. This chapter introduces the discrete time equivalent system for the continuous time tensor framework described in chapter 2. Using this equivalent model, different equalization schemes are studied such as zero forcing, minimum mean squared error equalization and decision feedback equalization (DFE) for both finite and infinite tensor taps. Performance results are presented for these equalizers. Further, some performance results from literature are reproduced using the tensor framework for the purpose of confirming the correct operation of the simulation software.

Chapter 5. This chapter describes tensor based correlative coding where controlled intersymbol interference is allowed to increase data rates and for spectrum shaping. Partial Response equalization is described where the equalizers defined in the previous chapter are used to cancel only part of the effects of the transmission channel.

Chapter 2

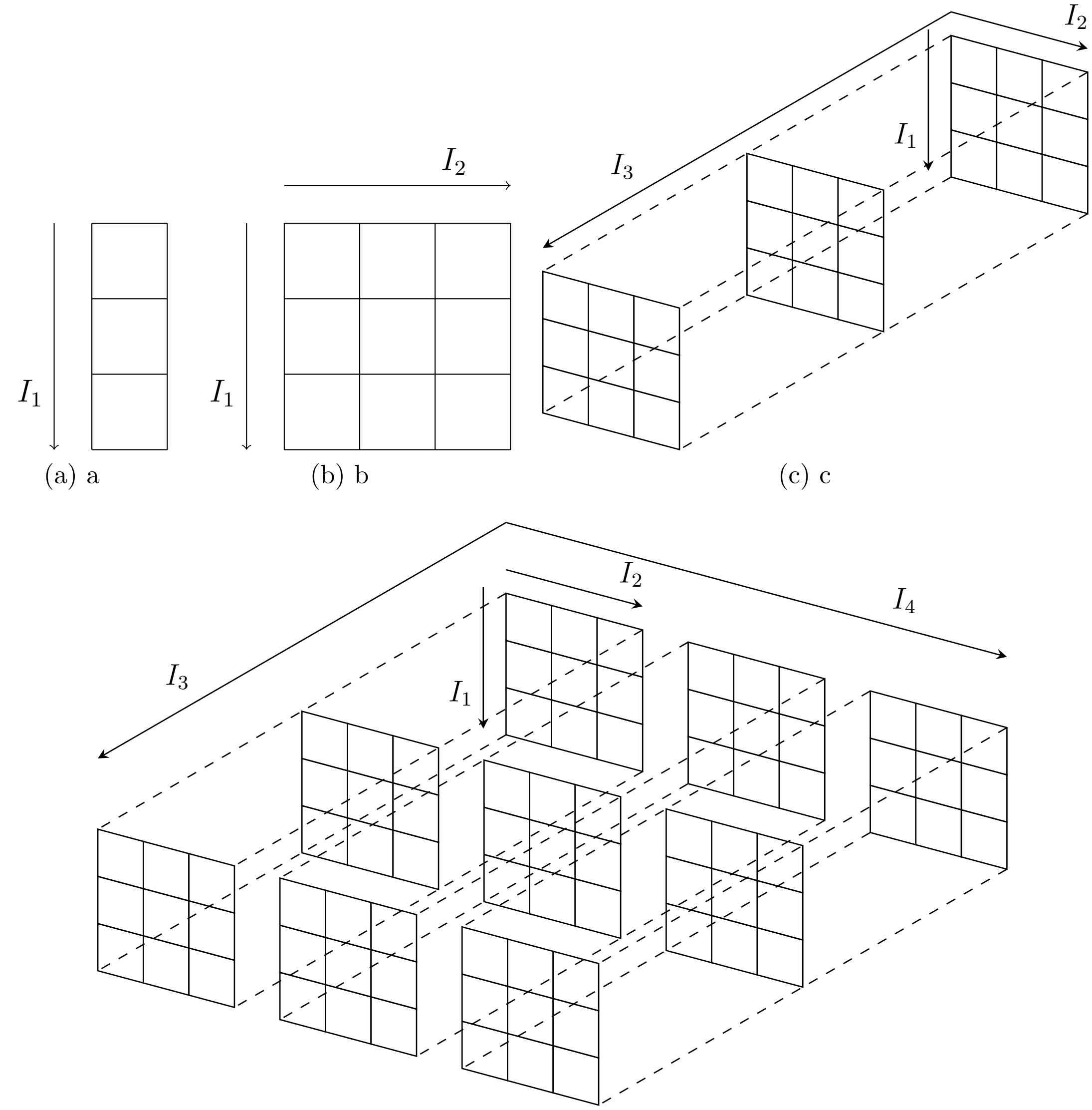
Preliminary Definitions and System Model

This chapter introduces tensors and some of their important properties. The notion of a tensor of functions is defined, along with specific types of function tensors such as signal and system tensors and their transformations. Using these definitions, the tensor framework for multi-domain communication, in its most generic form is defined. Finally, a higher order generalization of the scalar Nyquist's criterion for zero inter-symbol interference is presented with examples using different number of domains.

2.1 Tensors

A tensor is a multi-dimensional array of data [5]. The order of a tensor is the number of dimensions. A vector is a tensor of order one, a matrix is a tensor of order two and tensors of order greater than two are known as higher order tensors. Figure 2.1 shows the structure of tensors of order 1,2,3, and 4.

Definition 2.1.1. *The Contracted Product:* The *contracted product* over K dimensions, or modes, of an Nth order tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and an Mth order tensor $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_M}$ where $I_1 = J_1, \ldots, I_K = J_K$ with $K \leq \min(N, M)$ is a $(N + M - 2K)$th order tensor



(a) a (b) b (c) c

(d) d

Fig. 2.1 (a) a first order tensor (vector) (b) a second order tensor (matrix) (c) a third order tensor of size $3 \times 3 \times 3$ (d) a fourth order tensor of size $3 \times 3 \times 3 \times 3$

$\boldsymbol{\mathfrak{C}} \in \mathbb{C}^{I_{K+1} \times I_{K+2} \times \ldots \times I_N \times J_{K+1} \times J_{K+2} \times \ldots J_M}$ defined as [5]

$$\boldsymbol{\mathfrak{C}} = \{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathfrak{B}}\}_{(1,\ldots,K;1,\ldots,K)} \tag{2.1}$$

where

$$\boldsymbol{\mathcal{C}}_{i_{K+1},\ldots,i_N,j_{K+1},\ldots,j_M} = \sum_{i_1} \ldots \sum_{i_K} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_K,i_{K+1},\ldots,i_N} \boldsymbol{\mathcal{B}}_{i_1,\ldots,i_K,j_{K+1},\ldots,j_M}. \tag{2.2}$$

In (2.1), the modes of contraction are the first K modes of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$. However, it should be noted that the modes of contraction do not have to be the same in both tensors, since any two modes of same size can be contracted. For example, the first and second modes of tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{3\times4\times5}$ and the second and third modes of tensor $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{2\times3\times4}$ can be contracted to give a tensor

$$\boldsymbol{\mathcal{X}} = \{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}_{(1,2;2,3)} \tag{2.3}$$

where

$$\boldsymbol{\mathcal{X}}_{i_3,j_1} = \sum_{i_1=1}^{3} \sum_{i_2=1}^{4} \boldsymbol{\mathcal{A}}_{i_1,i_2,i_3} \boldsymbol{\mathcal{B}}_{j_1,i_1,i_2}. \tag{2.4}$$

A contraction that appears commonly throughout this thesis is one where the modes of contraction appear at the end of the first tensor and the beginning of the second. Consider a $(P+N)$th order tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1\times\ldots\times I_P\times J_1\times\ldots\times J_N}$ and a $(N+Q)$th order tensor $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1\times\ldots\times J_N\times K_1\times\ldots\times K_Q}$. The contracted product over the last N modes of $\boldsymbol{\mathcal{A}}$ and the first N modes of $\boldsymbol{\mathcal{B}}$ is a $(P+Q)$th order tensor $\boldsymbol{\mathcal{C}}$

$$\boldsymbol{\mathcal{C}} = \{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}_{(P+1,\ldots,P+N;1,\ldots,N)} \tag{2.5}$$

with components

$$\boldsymbol{\mathcal{C}}_{i_1,\ldots,i_P,k_1,\ldots,k_Q} = \sum_{j_1} \ldots \sum_{j_N} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_P,j_1,\ldots,j_N} \boldsymbol{\mathcal{B}}_{j_1,\ldots,j_N,k_1,\ldots,k_Q} \tag{2.6}$$

In the rest of this thesis, for the sake of brevity, we use the shorthand notation

$$\boldsymbol{\mathcal{C}} = \{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}_{(N)} \tag{2.7}$$

to denote the contraction in (2.5).

Next, we explore a special case where the contracted product is associative. All tensor product chains that appear in this work take the form described in this derivation and hence are associative.

Theorem 1. For tensors $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times \ldots \times I_M \times J_1 \times \ldots \times J_N}$, $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1 \times \ldots \times J_N \times K_1 \times \ldots \times K_P}$ and $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{K_1 \times \ldots \times K_P \times L_1 \times \ldots \times L_Q}$, we have

$$\{\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}_{(M+1,\ldots,M+N;1,\ldots,N)}, \boldsymbol{\mathcal{C}}\}_{(M+1,\ldots,M+P;1,\ldots,P)} = \{\boldsymbol{\mathcal{A}}, \{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}}\}_{(N+1,\ldots,N+P;1,\ldots,P)}\}_{(M+1,\ldots,M+N;1,\ldots,N)} \tag{2.8}$$

Proof. Let

$$\boldsymbol{\mathcal{X}} = \{\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}_{(M+1,\ldots,M+N;1,\ldots,N)}, \boldsymbol{\mathcal{C}}\}_{(M+1,\ldots,M+P;1,\ldots,P)} \tag{2.9}$$

and

$$\boldsymbol{\mathcal{Y}} = \{\boldsymbol{\mathcal{A}}, \{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}}\}_{(N+1,\ldots,N+P;1,\ldots,P)}\}_{(M+1,\ldots,M+N;1,\ldots,N)} \tag{2.10}$$

with components

$$\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_M,l_1,\ldots,l_Q} = \sum_{k_1} \cdots \sum_{k_P} \Big(\sum_{j_1} \cdots \sum_{j_N} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_M,j,1\ldots,j_N} \boldsymbol{\mathcal{B}}_{j_1,\ldots,j_N,k_1,\ldots,k_P} \Big) \boldsymbol{\mathcal{C}}_{k_1,\ldots,k_P,l_1,\ldots,l_Q} \tag{2.11}$$

and

$$\boldsymbol{\mathcal{Y}}_{i_1,\ldots,i_M,l_1,\ldots,l_Q} = \sum_{j_1} \cdots \sum_{j_N} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_M,j_1,\ldots,j_N} \Big(\sum_{k_1} \cdots \sum_{k_P} \boldsymbol{\mathcal{B}}_{j_1,\ldots,j_N,k_1,\ldots,k_P} \boldsymbol{\mathcal{C}}_{k_1,\ldots,k_P,l_1,\ldots,l_Q} \Big) \tag{2.12}$$

Notice that (2.11) can be re-written after removing the inner parenthesis as

$$\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_M,l_1,\ldots,l_Q} = \sum_{k_1} \cdots \sum_{k_P} \sum_{j_1} \cdots \sum_{j_N} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_M,j_1,\ldots,j_N} \boldsymbol{\mathcal{B}}_{j_1,\ldots,j_N,k_1,\ldots,k_P} \boldsymbol{\mathcal{C}}_{k_1,\ldots,k_P,l_1,\ldots,l_Q} \tag{2.13}$$

Changing the order of summation in (2.13) we get

$$\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_M,l_1,\ldots,l_Q} = \sum_{j_1} \cdots \sum_{j_N} \sum_{k_1} \cdots \sum_{k_P} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_M,j_1,\ldots,j_N} \boldsymbol{\mathcal{B}}_{j_1,\ldots,j_N,k_1,\ldots,k_P} \boldsymbol{\mathcal{C}}_{k_1,\ldots,k_P,l_1,\ldots,l_Q} \tag{2.14}$$

Factoring out $\boldsymbol{\mathcal{A}}$ from the inner summation over $k_1, k_2, \ldots, k_P$ gives

$$\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_M,l_1,\ldots,l_Q} = \sum_{j_1} \cdots \sum_{j_N} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_M,j_1,\ldots,j_N} \Big(\sum_{k_1} \cdots \sum_{k_P} \boldsymbol{\mathcal{B}}_{j_1,\ldots,j_N,k_1,\ldots,k_P} \boldsymbol{\mathcal{C}}_{k_1,\ldots,k_P,l_1,\ldots,l_Q} \Big) \tag{2.15}$$

which is the same as (2.12) □

Definition 2.1.2. *Outer Product:* The outer product of two tensors $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \ldots \times I_N}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1 \times J_2 \ldots \times J_M}$ is denoted by $\boldsymbol{\mathcal{A}} \circ \boldsymbol{\mathcal{B}} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_M}$ and can be represented as a specific case of the contracted product

$$\mathcal{A} \circ \mathcal{B} = \{\mathcal{A}, \mathcal{B}\}_{(0,0)} \tag{2.16}$$

The components of the outer product between $\mathcal{A}$ and $\mathcal{B}$ are

$$\left(\mathcal{A} \circ \mathcal{B}\right)_{i_1,\ldots,i_N,j_1\ldots,j_M} = \mathcal{A}_{i_1,\ldots,i_N}\mathcal{B}_{j_1,\ldots,j_M} \tag{2.17}$$

The tensor outer product is a generalization of the outer product between two vectors (tensor of order one) resulting in a matrix (tensor of order two). For example, consider vectors $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{y} \in \mathbb{C}^M$ then (2.16) becomes

$$\mathbf{x} \circ \mathbf{y} = \{\mathbf{x}, \mathbf{y}\}_{(0,0)} = \mathbf{x}\mathbf{y}^T \tag{2.18}$$

The tensor outer product is distributive and associative. It is not in general commutative. For tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N}$, $\mathcal{B} \in \mathbb{C}^{I_1\times\ldots\times I_N}$ and $\mathcal{C} \in \mathbb{C}^{J_1\times\ldots\times J_M}$ we have

$$\left(\mathcal{A} + \mathcal{B}\right) \circ \mathcal{C} = \mathcal{A} \circ \mathcal{C} + \mathcal{B} \circ \mathcal{C} \tag{2.19}$$

Proof.

$$\begin{aligned}
\left[\left(\mathcal{A} + \mathcal{B}\right) \circ \mathcal{C}\right]_{i_1\ldots i_N k_1\ldots k_P} &= \left(\mathcal{A} + \mathcal{B}\right)_{i_1\ldots i_N}\mathcal{C}_{k_1\ldots k_P} \\
&= \left(\mathcal{A}_{i_1,\ldots,i_N} + \mathcal{B}_{i_1,\ldots,i_N}\right)\mathcal{C}_{k_1,\ldots,k_P} \\
&= \mathcal{A}_{i_1,\ldots,i_N}\mathcal{C}_{k_1,\ldots,k_P} + \mathcal{B}_{i_1,\ldots,i_N}\mathcal{C}_{k_1,\ldots,k_P} \\
&= \left[\mathcal{A} \circ \mathcal{C} + \mathcal{B} \circ \mathcal{C}\right]_{i_1,\ldots,i_N,k_1,\ldots,k_P} \qquad \square
\end{aligned}$$

Similarly, For tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\ldots\times I_N}$,$\mathcal{B} \in \mathbb{C}^{J_1\times\ldots\times J_M}$ and $\mathcal{C} \in \mathbb{C}^{K_1\times\ldots\times K_P}$ we have

$$\left(\mathcal{A} \circ \mathcal{B}\right) \circ \mathcal{C} = \mathcal{A} \circ \left(\mathcal{B} \circ \mathcal{C}\right) \tag{2.20}$$

Proof.

$$\begin{aligned}
\left[\left(\mathcal{A} \circ \mathcal{B}\right) \circ \mathcal{C}\right]_{i_1,\ldots,i_N,j_1,\ldots,j_M,k_1,\ldots,k_P} &= \left(\mathcal{A} \circ \mathcal{B}\right)_{i_1,\ldots,i_N,j_1,\ldots,j_M}\mathcal{C}_{k_1,\ldots,k_P} \\
&= \mathcal{A}_{i_1,\ldots,i_N}\mathcal{B}_{j_1,\ldots,j_M}\mathcal{C}_{k_1,\ldots,k_P} \\
&= \mathcal{A}_{i_1,\ldots,i_N}\left(\mathcal{B} \circ \mathcal{C}\right)_{j_1,\ldots,j_M,k_1,\ldots,k_P} \\
&= \left[\mathcal{A} \circ \left(\mathcal{B} \circ \mathcal{C}\right)\right]_{i_1,\ldots,i_N,j_1,\ldots,j_M,k_1,\ldots,k_P} \qquad \square
\end{aligned}$$

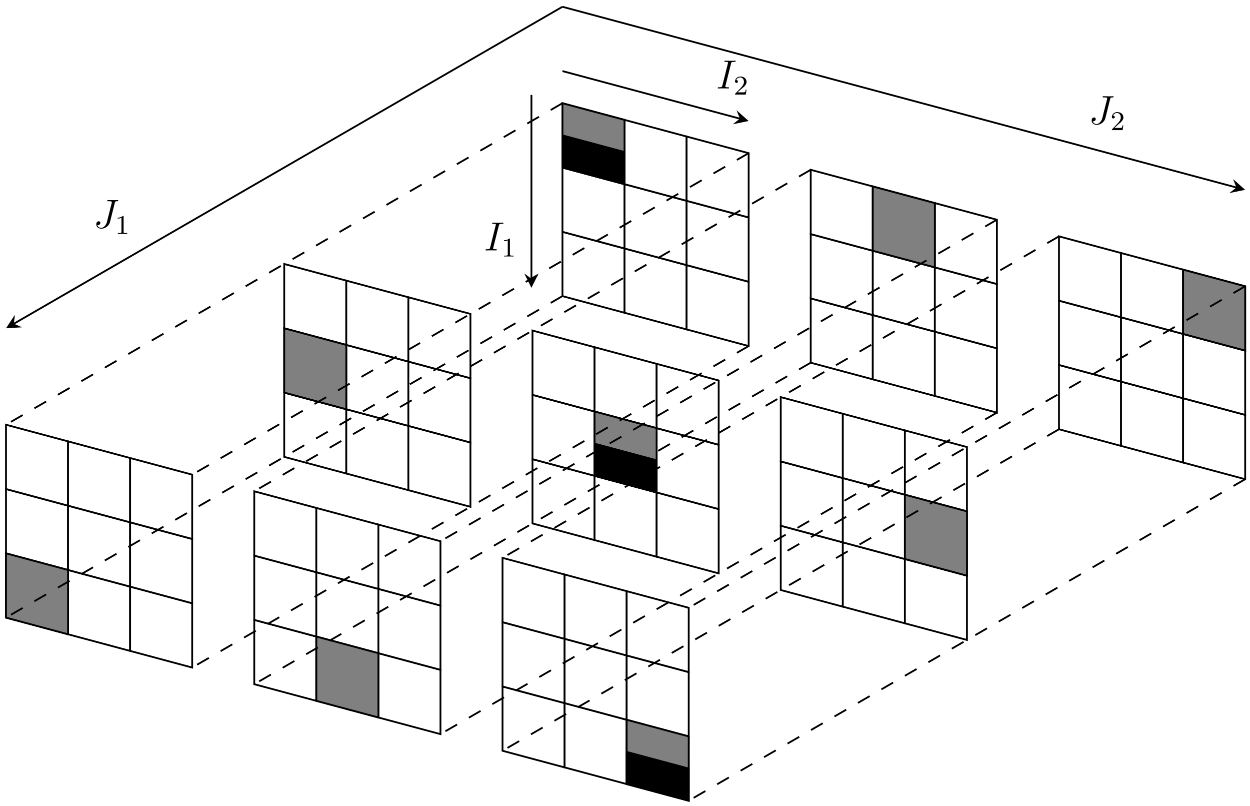


Fig. 2.2 Pseudo-diagonal (gray) and diagonal (black) elements of a tensor of size $3 \times 3 \times 3 \times 3$

Definition 2.1.3. *Diagonal and Pseudo-diagonal tensors:* A tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \ldots \times I_N}$ is *diagonal* if

$$\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N} = \begin{cases} k_{i_1,\ldots,i_N} & \text{if } i_1 = i_2 = \ldots i_N \\ 0 & \text{otherwise} \end{cases} \tag{2.21}$$

where $k_{i_1 \ldots i_N}$ is an arbitrary scalar. A pseudo-diagonal tensor is a tensor $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ with components

$$\boldsymbol{\mathcal{B}}_{i_1,\ldots,i_N,j_1,\ldots,j_N} = \begin{cases} k_{i_1,\ldots,i_N,j_1,\ldots,j_N} & \text{if } i_1 = j_1,\ i_2 = j_2,\ \ldots\ i_N = j_N \\ 0 & \text{otherwise} \end{cases} \tag{2.22}$$

Authors in [21] and [22] define tensors of the form in (2.22) as diagonal tensors. However, given the stricter, more prevalent, definition of a diagonal tensor [5] the notion of pseudo-diagonality is used in this thesis. The non-zero entries of a pseudo-diagonal tensor are known as its pseudo-diagonal entries. Figure 2.2 shows a fourth order tensor of size $J_1 \times J_2 \times I_1 \times I_2$ with $I_1 = I_2 = J_1 = J_2 = 3$ with the pseudo-diagonal elements highlighted in gray and diagonal elements highlighted in black. We can see that all the diagonal elements are also pseudo-diagonal elements and hence a diagonal tensor is a pseudo-diagonal tensor with zeroes in some of its pseudo-diagonal entries.

Definition 2.1.4. *Identity tensor:* We define an *identity tensor* of order $2N$ as a pseudo-diagonal tensor $\mathcal{I}_N \in \mathbb{C}^{I_1 \times I_2 \ldots \times I_N \times I_1 \times I_2 \ldots \times I_N}$ with entries

$$\mathcal{I}_{N_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} = \delta_{i_1,i'_1} \cdots \delta_{i_N,i'_N} \tag{2.23}$$

where $\delta_{x,y}$ is the kronecker delta defined as

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases} \tag{2.24}$$

The sub-script N is used to denote the order of the identity tensor. For example, an identity tensor $\mathcal{I}_N$ is of order $2N$ while $\mathcal{I}_M$ is of order $2M$. For a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \ldots \times I_N \times J_1 \times J_2 \ldots \times J_M}$ we have

$$\{\mathcal{X}, \mathcal{I}_M\}_{(M)} = \{\mathcal{I}_N, \mathcal{X}\}_{(N)} = \mathcal{X} \tag{2.25}$$

Definition 2.1.5. *Inner product and Frobenius norm of a tensor:* The *inner product* of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \{\mathcal{A}, \mathcal{B}\}_{(1,\ldots,N;1,\ldots,N)} = \sum_{i_1} \ldots \sum_{i_N} \mathcal{A}_{i_1,\ldots,i_N} \mathcal{B}_{i_1,\ldots,i_N} \tag{2.26}$$

The *Frobenium norm* of a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ is defined as

$$\|\mathcal{X}\|_F = \left(\sum_{i_1} \ldots \sum_{i_N} \left| \mathcal{X}_{i_1,\ldots,i_N} \right|^2 \right)^{\frac{1}{2}} \tag{2.27}$$

Definition 2.1.6. *Transpose and Hermitian of a Tensor:* A matrix has two indices and the transpose of a matrix is a permutation of these two indices. Since there are several dimensions in a tensor, there are many permutations of its indices and hence there are several ways to write the transpose of a tensor. Authors in [23] define the transpose of a tensor using permutations.

Assume the set $S_N = \{1, 2, \ldots, N\}$ and σ is a permutation of S_N. We denote $\sigma(j) = i_j$ for $j = 1, 2, \ldots, N$ where $\{i_1, i_2, \ldots, i_N\} = \{1, 2, \ldots, N\} = S_N$. Since S_N is a finite set with N elements, it has $N!$ different permutations. Hence, discounting the identity permutation $\sigma(j) = [1, 2, \ldots, N]$, there are $N! - 1$ different transposes for a tensor with N dimensions or modes.

For a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \ldots \times I_N}$ we define its transpose associated with a certain permutation σ as $\mathcal{A}^{T\sigma} \in \mathbb{C}^{I_{\sigma(1)} \times I_{\sigma(2)} \ldots \times I_{\sigma(N)}}$ with entries

$$\mathcal{A}^{T\sigma}_{i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(N)}} = \mathcal{A}_{i_1, i_2, \ldots, i_N} \tag{2.28}$$

Similarly, the Hermitian of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \ldots \times I_N}$ associated with a permutation σ is defined as the conjugate of its transpose and is denoted as $\mathcal{A}^{H\sigma} = (\mathcal{A}^{T\sigma})^* \in \mathbb{C}^{I_{\sigma(1)} \times I_{\sigma(2)} \ldots \times I_{\sigma(N)}}$ with entries

$$\mathcal{A}^{H\sigma}_{i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(N)}} = (\mathcal{A}^{T\sigma}_{i_{\sigma(1)}, i_{\sigma(2)}, \ldots, i_{\sigma(N)}})^* = (\mathcal{A}_{i_1, i_2, \ldots, i_N})^* \tag{2.29}$$

For example, a transpose of a third order tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ such that its third mode is transposed with the first can be written as $\mathcal{X}^{T\sigma}$ where $\sigma = [3, 2, 1]$ with components $\mathcal{X}^{T\sigma}_{i_3, i_2, i_1} = \mathcal{X}_{i_1, i_2, i_3}$. For two tensors $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \ldots \times I_N}$ and $\mathcal{B} \in \mathbb{C}^{I_1 \times J_2 \ldots \times I_N}$ we have [23]

$$\langle \mathcal{A}, \mathcal{B} \rangle = \langle \mathcal{A}^{T\sigma}, \mathcal{B}^{T\sigma} \rangle \tag{2.30}$$

and

$$\|\mathcal{A}\|_F = \left\|\mathcal{A}^{T\sigma}\right\|_F \tag{2.31}$$

Consider a tensor $\mathcal{Y} \in \mathbb{C}^{I_1 \times \ldots \times I_N J_1 \times \ldots \times J_M}$ with a transposition such that the final M modes are swapped with the first N modes can be represented by a permutation function $\sigma = [(N+1), \ldots (N+M), 1, \ \ldots \ N]$ such that $\mathcal{Y}^{T\sigma}_{j_1, \ldots, j_M, i_1, \ldots i_N} = \mathcal{Y}_{i_1, \ldots, i_N, j_1, \ldots, j_M}$. Since this type of transposition is most commonly used throughout this work, we drop the superscript σ for ease of representation and represent such a transpose by $\mathcal{Y}^T$ and its conjugate by $\mathcal{Y}^H$.

For tensors $\mathcal{A} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_M}$ and $\mathcal{B} \in \mathbb{C}^{J_1 \times \ldots \times J_M \times K_1 \times \ldots \times K_P}$ we have

$$\left(\{\mathcal{A}, \mathcal{B}\}_{(M)}\right)^H = \{\mathcal{B}^H, \mathcal{A}^H\}_{(M)} \tag{2.32}$$

Proof.

$$\begin{aligned}
\left(\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\}_{(M)}\right)^{H}_{k_1,\ldots,k_P,i_1,\ldots,i_N} &= \left(\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\}_{(M)}\right)_{i_1,\ldots,i_N,k_1,\ldots,k_P} \\
&= \sum_{j_1}\cdots\sum_{j_M}\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}\boldsymbol{\mathcal{B}}_{j_1,\ldots,j_M,k_1,\ldots,k_P} \\
&= \sum_{j_1}\cdots\sum_{j_M}\boldsymbol{\mathcal{A}}^{H}_{j_1,\ldots,j_M,i_1,\ldots,i_N}\boldsymbol{\mathcal{B}}^{H}_{k_1,\ldots,k_P,j_1,\ldots,j_M} \\
&= \sum_{j_1}\cdots\sum_{j_M}\boldsymbol{\mathcal{B}}^{H}_{k_1,\ldots,k_P,j_1,\ldots,j_M}\boldsymbol{\mathcal{A}}^{H}_{j_1,\ldots,j_M,i_1,\ldots,i_N} \\
&= \left(\{\boldsymbol{\mathcal{B}}^{H},\boldsymbol{\mathcal{A}}^{H}\}_{(M)}\right)_{k_1,\ldots,k_P,i_1,\ldots,i_N} \qquad (2.33)
\end{aligned}$$

□

For the case of matrices (order-two tensors), (2.32) reduces to the familiar relation

$$\left(\mathbf{A}\mathbf{B}\right)^{H} = \mathbf{B}^{H}\mathbf{A}^{H} \qquad (2.34)$$

where $\mathbf{A} \in \mathbb{C}^{I\times J}$ and $\mathbf{B} \in \mathbb{C}^{J\times K}$ are two matrices.

For tensors $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1\times\ldots\times I_N}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1\times\ldots\times J_M\times I_1\times\ldots\times I_N}$ we have

$$\{\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{A}}\}_{(N)} = \{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}^{T}\}_{(N)} \qquad (2.35)$$

Proof.

$$\begin{aligned}
\left[\{\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{A}}\}_{(N)}\right]_{j_1,\ldots,j_M} &= \sum_{i_1}\cdots\sum_{i_N}\boldsymbol{\mathcal{B}}_{j_1,\ldots,j_M,i_1,\ldots i_N}\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N} \\
&= \sum_{i_1}\cdots\sum_{i_N}\boldsymbol{\mathcal{B}}^{T}_{i_1,\ldots i_N,j_1,\ldots,j_M}\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N} \\
&= \sum_{i_1}\cdots\sum_{i_N}\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N}\boldsymbol{\mathcal{B}}^{T}_{i_1,\ldots i_N,j_1,\ldots,j_M} \\
&= \left[\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}^{T}\}_{(N)}\right]_{j_1,\ldots,j_M} \qquad (2.36)
\end{aligned}$$

□

Definition 2.1.7. *Tensor to Matrix Transformation:* For a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$, we define a transformation f_N that transforms $\boldsymbol{\mathcal{A}}$ to a matrix $\mathbf{A} \in \mathbb{C}^{I_1\ldots I_N\times J_1\ldots J_M}$ such that

$f(\boldsymbol{\mathcal{A}}) = \mathbf{A}$. Component-wise we have

$$\boldsymbol{\mathcal{A}}_{i_1,i_2,\ldots,i_N,j_1,\ldots,j_M} \xrightarrow{f_N} \mathbf{A}_{(i_1+\sum\limits_{k=2}^{N}(i_k-1)\prod\limits_{l=1}^{k-1}I_l),(j_1+\sum\limits_{k=2}^{M}(j_k-1)\prod\limits_{l=1}^{k-1}J_l)} \tag{2.37}$$

The subscript in f_N denotes a partition of the modes of the tensor being transformed. The product of the first N modes of the tensor becomes the number of rows of the matrix and the product of the remaining modes of the tensor becomes the number of columns of the matrix. For example, consider a tensor $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{2\times3\times4\times5\times6}$ and a transformation f_3 such that $f_3(\boldsymbol{\mathcal{B}}) = \mathbf{B}$. The size of $\mathbf{B}$ is $(2 \cdot 3 \cdot 4) \times (5 \cdot 6)$ and

$$\boldsymbol{\mathcal{B}}_{i,j,k,l,m} \xrightarrow{f_3} \mathbf{B}_{(i+2(j-1)+6(k-1)),(l+5(m-1))} \tag{2.38}$$

These transformations are called column or row major formats in many computer languages and essentially represents a particular type of matrix unfolding of a tensor. It is shown in [21], for the case of fourth-order tensors of the form $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I\times J\times I\times J}$, that the above transformation function is a bijection with a bijective inverse mapping f_N^{-1} to convert the matrix $\mathbf{A}$ back into the original tensor $\boldsymbol{\mathcal{A}}$. Authors in [38] extend this result to the case of tensors of any order.

Definition 2.1.8. *Tensor Inverse:* The group of invertible $N \times N$ matrices with matrix multiplication is called the general linear group denoted by $\mathbb{M}_{N,N}(\mathbb{C})$ [21]. Denote

$$\mathbb{T}_{I_1,I_2,\ldots,I_N,I_1,I_2,\ldots,I_N}(\mathbb{C}) = \{\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N} : det(f_N(\boldsymbol{\mathcal{A}})) \neq 0\} \tag{2.39}$$

Authors in [38] and [21] (for the special case of fourth order tensors) have shown that the set $\mathbb{T}_{I_1,I_2,\ldots,I_N,I_1,I_2,\ldots,I_N}(\mathbb{C})$ forms a group equipped with the contraction $\{\}_{(N)}$ as defined in (2.7) and the transformation f_N is an isomorphism between $\mathbb{T}_{I_1,I_2,\ldots,I_N,I_1,I_2,\ldots,I_N}(\mathbb{C})$ and $\mathbb{M}_{(I_1I_2\cdots I_N),(I_1I_2\cdots I_N)}(\mathbb{C})$. This indicates that for any tensor $\boldsymbol{\mathcal{A}} \in \mathbb{T}_{I_1,I_2,\ldots,I_N,I_1,I_2,\ldots,I_N}(\mathbb{C})$ there exists a tensor $\boldsymbol{\mathcal{B}} \in \mathbb{T}_{I_1,I_2,\ldots,I_N,I_1,I_2,\ldots,I_N}(\mathbb{C})$ such that [38]:

$$\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}_{(N)} = \{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{A}}\}_{(N)} = \boldsymbol{\mathcal{I}}_N. \tag{2.40}$$

where $\boldsymbol{\mathcal{I}}_N$ is the identity tensor. $\boldsymbol{\mathcal{B}}$ is called the inverse of $\boldsymbol{\mathcal{A}}$ and is denoted by $\boldsymbol{\mathcal{A}}^{-1}$.

The *Moore-Penrose inverse* of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1\times I_2\ldots\times I_N\times J_1\times J_2\ldots\times J_N}$,which is a generalization of the matrix Moore-Penrose inverse, is a tensor $\boldsymbol{\mathcal{A}}^{+} \in \mathbb{C}^{J_1\times J_2\ldots\times J_N\times I_1\times I_2\ldots\times I_N}$ that satisfies

[22]:

$$\{\{\mathcal{A}, \mathcal{A}^+\}_{(N)}, \mathcal{A}\}_{(N)} = \mathcal{A}$$
$$\{\{\mathcal{A}^+, \mathcal{A}\}_{(N)}, \mathcal{A}^+\}_{(N)} = \mathcal{A}^+$$
$$\{\mathcal{A}, \mathcal{A}^+\}_{(N)}^H = \{\mathcal{A}, \mathcal{A}^+\}_{(N)}$$
$$\{\mathcal{A}^+, \mathcal{A}\}_{(N)}^H = \{\mathcal{A}^+, \mathcal{A}\}_{(N)}$$

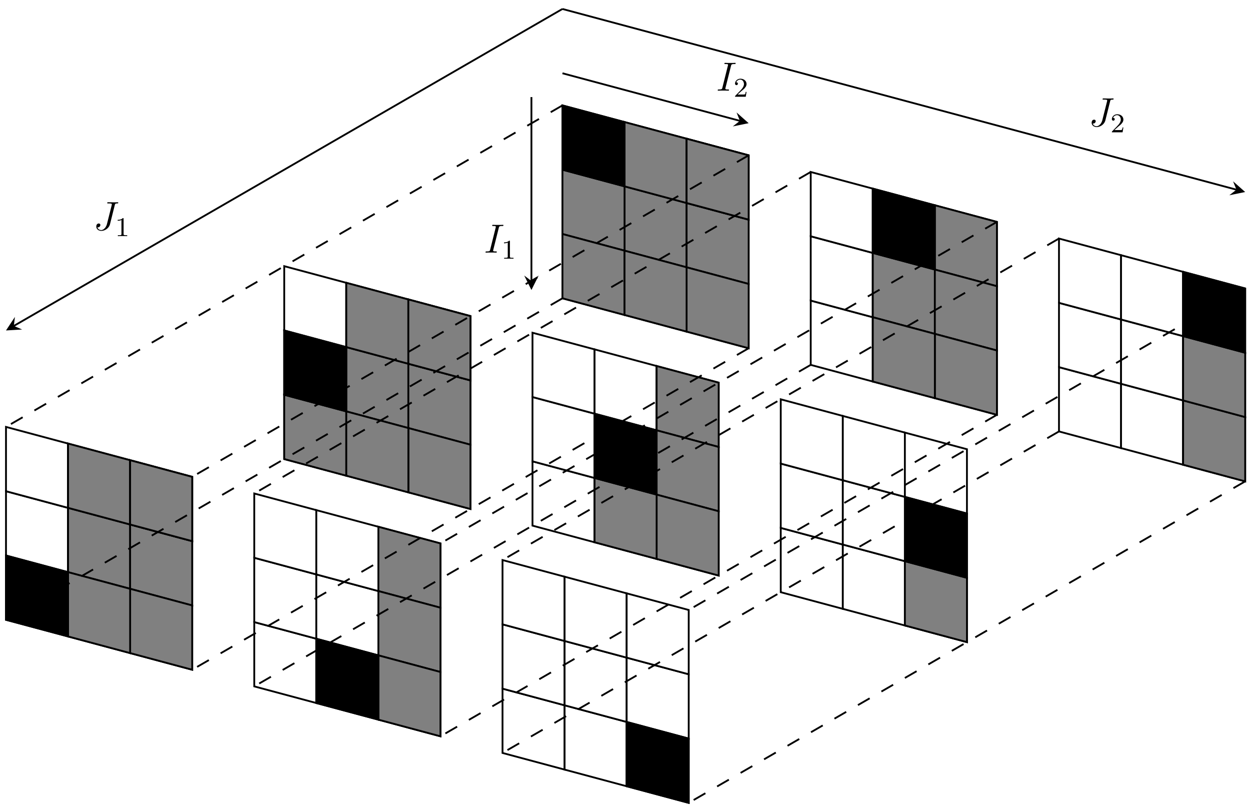


Fig. 2.3 Pseudo-Lower Triangular Tensor

Definition 2.1.9. *Pseudo-Triangular Tensor:* A tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ is defined to be pseudo-lower triangular if

$$\mathcal{A}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} = \begin{cases} 0 \quad \text{if } (i'_1 + \sum_{k=2}^{N} (i'_k - 1) \prod_{l=1}^{k-1} I_l) \geq (i_1 + \sum_{k=2}^{N} (i_k - 1) \prod_{l=1}^{k-1} I_l) \\ a_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} \quad \text{otherwise} \end{cases} \tag{2.41}$$

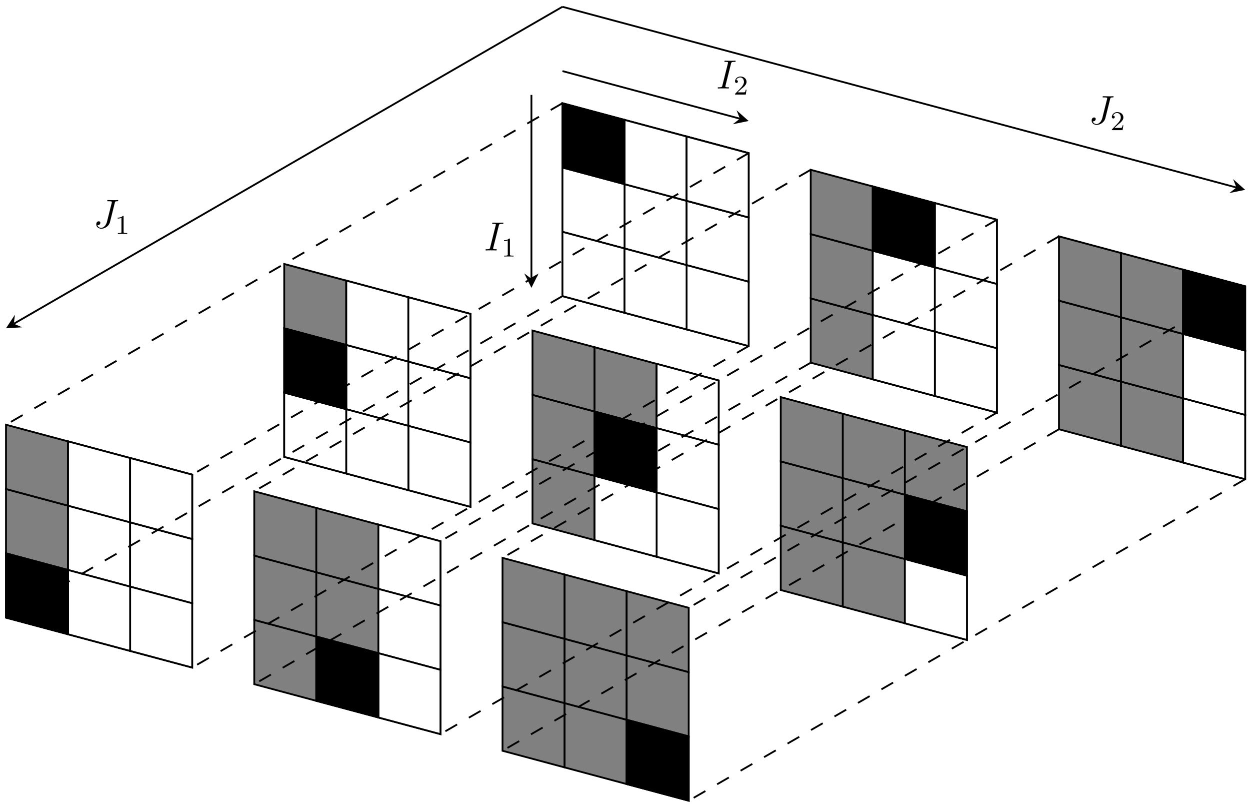


Fig. 2.4 Pseudo-Upper Triangular Tensor

where $a_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}$ are arbitrary scalars. Similarly, the tensor is said to be pseudo-upper triangular if

$$\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} = \begin{cases} 0 & \text{if } (i'_1 + \sum\limits_{k=2}^{N}(i'_k - 1)\prod\limits_{l=1}^{k-1} I_l) \leq (i_1 + \sum\limits_{k=2}^{N}(i_k - 1)\prod\limits_{l=1}^{k-1} I_l) \\ a_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} & \text{otherwise} \end{cases} \tag{2.42}$$

Shown in figures 2.3 and 2.4 are two tensors of size $J_1 \times J_2 \times I_1 \times I_2$ with $I_1 = I_2 = J_1 = J_2 = 3$ with their pseudo-lower and pseudo-upper triangular elements highlighted in gray along with their pseudo-diagonal elements shown in black. It can be readily seen that a lower triangular tensor becomes a lower triangular matrix under the tensor to matrix transformation defined in (2.37) and a pseudo-upper triangular tensor becomes an upper triangular matrix.

Proof. Consider a pseudo-lower triangular tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ with components

$\mathcal{A}_{i_1,\dots,i_N,i'_1,\dots,i'_N}$. The indices of the non-zero elements of this tensor have the following relation by definition:

$$(i'_1 + \sum_{k=2}^{N}(i'_k - 1)\prod_{l=1}^{k-1} I_l) < (i_1 + \sum_{k=2}^{N}(i_k - 1)\prod_{l=1}^{k-1} I_l) \tag{2.43}$$

Let the matrix transformation of this tensor be $f_N(\mathcal{A}) = \mathbf{A}$. The components of this matrix are

$$\mathbf{A}_{(i_1+\sum_{k=2}^{N}(i_k-1)\prod_{l=1}^{k-1} I_l),(i'_1+\sum_{k=2}^{N}(i'_k-1)\prod_{l=1}^{k-1} I_l)} \tag{2.44}$$

Let $x = i_1 + \sum_{k=2}^{N}(i_k - 1)\prod_{l=1}^{k-1} I_l$ and $y = i'_1 + \sum_{k=2}^{N}(i'_k - 1)\prod_{l=1}^{k-1} I_l$. Under the inequalities in (2.43) we have

$$\begin{aligned} x &= i_1 + \sum_{k=2}^{N}(i_k - 1)\prod_{l=1}^{k-1} I_l \\ &< i'_1 + \sum_{k=2}^{N}(i'_k - 1)\prod_{l=1}^{k-1} I_l \\ &= y \end{aligned} \tag{2.45}$$

This implies that all non-zero elements of the tensor are present either on or below the diagonal of the matrix $\mathbf{A}$. A similar proof shows that a pseudo-upper triangular tensor transforms into an upper triangular matrix. □

2.2 Signal and System Tensors

Definition 2.2.1. *Function Tensor:* A function tensor $\mathcal{A}(x)$ is a tensor whose components are functions of x. Using a third order function tensor as an example, each component of $\mathcal{A}(x)$ is written as $\mathcal{A}_{i,j,k}(x)$.

Definition 2.2.2. *Multivariate Function tensor:* A generalization of the previous definition would be the multivariate function tensor $\mathcal{A}(x_1, \dots, x_p)$, which is a tensor whose components are functions of the variables $x_1, \dots, x_p$. Using the same example of a third order tensor, each component can be written as $\mathcal{A}_{i,j,k}(x_1, x_2, \dots, x_p)$. For example, a tensor $\mathcal{A}(t, u)$ is a multivariate function tensor whose components are functions of t and u.

Definition 2.2.3. *Signal Tensor and System Tensors:* A *signal tensor* $\boldsymbol{\mathcal{X}}(t)$ is a function tensor whose components are functions of time. A *system tensor* $\boldsymbol{\mathcal{H}}(t,u)$, used to describe linear time varying multidomain systems, is a tensor of order $N+M$ that couples two signal tensors of orders N and M respectively through a contracted linear functional. For example, let $\boldsymbol{\mathcal{H}}(t,u) \in \mathbb{C}_t^{Y_1 \times Y_2 \ldots \times Y_M \times X_1 \times X_2 \ldots \times X_N}$ be a system tensor that couples $\boldsymbol{\mathcal{X}}(t) \in \mathbb{C}_t^{X_1 \times X_2 \ldots \times X_N}$ with $\boldsymbol{\mathcal{Y}}(t) \in \mathbb{C}_t^{Y_1 \times Y_2 \ldots \times Y_M}$. Here, $\mathbb{C}_t^{A \times B}$ is used to denote the set of tensors of size $A \times B$ whose components are complex functions of t. The output tensor $\boldsymbol{\mathcal{Y}}(t)$ has components

$$\boldsymbol{\mathcal{Y}}_{y_1 y_2 \ldots y_M}(t) = \sum_{x_1 x_2 \ldots x_N} \int_{-\infty}^{\infty} \boldsymbol{\mathcal{H}}_{y_1 y_2 \ldots y_M x_1 x_2 \ldots x_N}(t,u) \boldsymbol{\mathcal{X}}_{x_1 x_2 \ldots x_N}(u) du. \tag{2.46}$$

Definition 2.2.4. *The Contracted Convolution and Time Invariant System Tensor:* A linear time invariant system tensor $\boldsymbol{\mathcal{H}}(t)$ is a tensor of order $N+M$ that couples two signal tensors of orders N and M respectively. Extending the contracted product to a contracted convolution allows us to define the coupling of the input and output signal tensors by a linear time invariant system tensor. Consider a signal tensor $\boldsymbol{\mathcal{X}}(t) \in \mathbb{C}_t^{X_1 \times X_2 \ldots \times X_N}$ and a system tensor $\boldsymbol{\mathcal{H}}(t) \in \mathbb{C}_t^{Y_1 \times Y_2 \ldots \times Y_M \times X_1 \times X_2 \ldots \times X_N}$. The contracted convolution of tensor $\boldsymbol{\mathcal{X}}(t)$ and tensor $\boldsymbol{\mathcal{H}}(t)$ is a tensor $\boldsymbol{\mathcal{Y}}(t) \in \mathbb{C}_t^{Y_1 \times Y_2 \ldots \times Y_M}$ defined as

$$\boldsymbol{\mathcal{Y}}(t) = \{\boldsymbol{\mathcal{H}}(t) * \boldsymbol{\mathcal{X}}(t)\}_{(M+1,\ldots,M+N;1,2,\ldots N)}, \tag{2.47}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_{y_1,\ldots,y_M}(t) &= \sum_{x_1,\ldots,x_N} \boldsymbol{\mathcal{H}}_{y_1,\ldots,y_M,x_1,\ldots,x_N}(t) * \boldsymbol{\mathcal{X}}_{x_1,\ldots,x_N}(t) \\ &= \sum_{x_1,\ldots,x_N} \int_{-\infty}^{\infty} \boldsymbol{\mathcal{H}}_{y_1,\ldots,y_M,x_1,\ldots,x_N}(t-\tau) \boldsymbol{\mathcal{X}}_{x_1,\ldots,x_N}(\tau) d\tau.\end{aligned} \tag{2.48}$$

Let $\boldsymbol{\mathcal{X}}(t) \in \mathbb{C}_t^{X_1 \times X_2 \ldots \times X_N}$ be a signal tensor and $\boldsymbol{\mathcal{H}}(t) \in \mathbb{C}_t^{Y_1 \times Y_2 \ldots \times Y_M \times X_1 \times X_2 \ldots \times X_N}$ and $\boldsymbol{\mathcal{G}}(t) \in \mathbb{C}_t^{Z_1 \times Z_2 \ldots \times Z_P \times Y_1 \times Y_2 \ldots \times Y_M}$ be system tensors. It can readily be seen that the conditions for associativity from (2.8) are also valid for function tensors since the only change is that multiplications are replaced by convolutions and scalars with functions. If the output of the cascade of these two systems is denoted by $\boldsymbol{\mathcal{Z}}(t) \in \mathbb{C}_t^{Z_1 \times Z_2 \ldots \times Z_P}$, we have

$$\begin{aligned}\boldsymbol{\mathcal{Z}}(t) &= \{\{\boldsymbol{\mathcal{G}}(t) * \boldsymbol{\mathcal{H}}(t)\}_{(P+1,\ldots,P+M;1,\ldots,M)} * \boldsymbol{\mathcal{X}}(t)\}_{(P+1,\ldots,P+N;1,\ldots,N)} \\ &= \{\boldsymbol{\mathcal{G}}(t) * \{\boldsymbol{\mathcal{H}}(t) * \boldsymbol{\mathcal{X}}(t)\}_{(M+1,\ldots,M+N;1,\ldots,N)}\}_{(P+1,\ldots,P+M;1,\ldots,M)}\end{aligned} \tag{2.49}$$

Definition 2.2.5. *The Fourier Transform of Signal Tensors:* The Fourier transform of a signal tensor is a tensor of the Fourier transforms of its individual components. If the Fourier transforms of all the individual functions exist, then $\breve{\boldsymbol{\mathcal{X}}}(f)$ the Fourier transform of $\boldsymbol{\mathcal{X}}(t) \in \mathbb{C}_t^{I_1 \times \ldots \times I_N}$ has components

$$\breve{\boldsymbol{\mathcal{X}}}_{x_1,\ldots,x_N}(f) = \int_{-\infty}^{\infty} \boldsymbol{\mathcal{X}}_{x_1,\ldots,x_N}(t) e^{-j2\pi f t} dt = \mathcal{F}\big[\boldsymbol{\mathcal{X}}_{x_1,\ldots,x_N}(t)\big] \tag{2.50}$$

2.3 The Tensor Framework

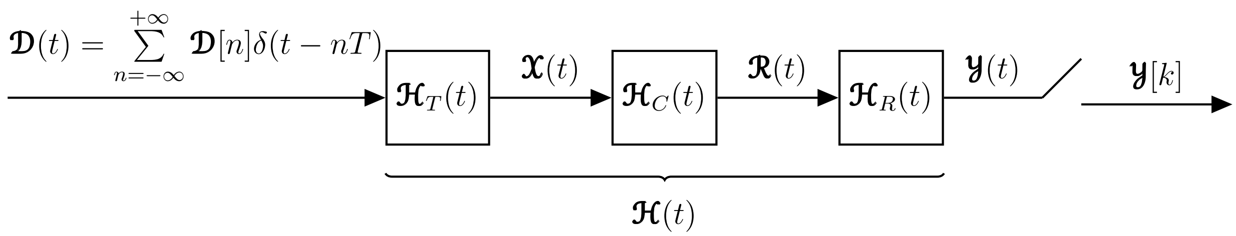


Fig. 2.5 System Model

Consider a tensor communication system where $\boldsymbol{\mathcal{D}}[n] \in \mathbb{C}_n^{I_1 \times \ldots \times I_N}$ represents the data to be transmitted by the nth tensor symbol. The components of $\boldsymbol{\mathcal{D}}[n]$ may be constellation mapped data symbols or may be precoded data symbols. An example of the latter is Tensor Partial Response Signalling (TPRS), which is detailed in Chapter 5. Let the symbol period be T, i.e., a data tensor is transmitted at intervals of time T. Then we can represent such a data symbol by $\boldsymbol{\mathcal{D}}[n]\delta(t - nT)$ where $\delta(t)$ is Dirac's delta function. Let the transmit filters, the channel and the receive filters be represented by three system tensors $\boldsymbol{\mathcal{H}}_T(t) \in \mathbb{C}_t^{J_1 \times \ldots \times J_P \times I_1 \times \ldots \times I_N}$, $\boldsymbol{\mathcal{H}}_C(t) \in \mathbb{C}_t^{K_1 \times \ldots \times K_Q \times J_1 \times \ldots \times J_P}$ and $\boldsymbol{\mathcal{H}}_R(t) \in \mathbb{C}_t^{I_1 \times \ldots \times I_N \times K_1 \times \ldots \times K_Q}$. The overall system model in the absence of noise is presented in Figure 2.5. The input to the transmit filter is $\sum_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{D}}[n]\delta(t - nT)$. The dimension of the transmit signal tensor being different from the data tensor allows a unifying representation of various schemes. For example, P would be greater than N if the same data symbol may be sent on multiple components of $\boldsymbol{\mathcal{X}}(t)$. Similarly, when multiple data symbols are sent on a single component of $\boldsymbol{\mathcal{X}}(t)$ then P would be smaller than N. If there is a one-to-one mapping between the symbols and waveforms then P would be equal to N.

The transmit signal tensor of order P is

$$\boldsymbol{\mathcal{X}}(t) = \sum_{n=-\infty}^{+\infty} \left\{\boldsymbol{\mathcal{H}}_T(t) * \boldsymbol{\mathcal{D}}[n]\delta(t-nT)\right\}_{(N)} \tag{2.51}$$

$$= \sum_{n=-\infty}^{+\infty} \left\{\boldsymbol{\mathcal{H}}_T(t-nT), \boldsymbol{\mathcal{D}}[n]\right\}_{(N)} \tag{2.52}$$

The effects of the channel on the transmit signal tensor is represented by a contraction with a channel system tensor $\boldsymbol{\mathcal{H}}_C(t)$ of order $(Q+P)$. The received signal $\boldsymbol{\mathcal{R}}(t) \in \mathbb{C}_t^{K_1 \times \ldots \times K_Q}$ is

$$\boldsymbol{\mathcal{R}}(t) = \{\boldsymbol{\mathcal{H}}_C(t) * \boldsymbol{\mathcal{X}}(t)\}_{(P)} \tag{2.53}$$

The receive system tensor $\boldsymbol{\mathcal{H}}_R(t)$ of order $(N+Q)$ transforms the received signal tensor $\boldsymbol{\mathcal{R}}(t)$ into a signal tensor of the same size as the data tensor. The output of the receive filter tensor $\boldsymbol{\mathcal{Y}}(t) \in \mathbb{C}_t^{I_1 \times I_2 \ldots \times I_N}$ is

$$\boldsymbol{\mathcal{Y}}(t) = \{\boldsymbol{\mathcal{H}}_R(t) * \boldsymbol{\mathcal{R}}(t)\}_{(Q)} \tag{2.54}$$

From (2.54),(2.53) and (2.51) we get

$$\begin{aligned}
\boldsymbol{\mathcal{Y}}(t) &= \{\boldsymbol{\mathcal{H}}_R(t) * \{\boldsymbol{\mathcal{H}}_C(t) * \boldsymbol{\mathcal{X}}(t)\}_{(P)}\}_{(Q)} \\
&= \left\{\boldsymbol{\mathcal{H}}_R(t) * \left\{\boldsymbol{\mathcal{H}}_C(t) * \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}_T(t) * \boldsymbol{\mathcal{D}}[n]\delta(t-nT)\}_{(N)}\right\}_{(P)}\right\}_{(Q)} \\
&= \sum_{n=-\infty}^{+\infty} \left\{\boldsymbol{\mathcal{H}}_R(t) * \left\{\boldsymbol{\mathcal{H}}_C(t) * \{\boldsymbol{\mathcal{H}}_T(t) * \boldsymbol{\mathcal{D}}[n]\delta(t-nT)\}_{(N)}\right\}_{(P)}\right\}_{(Q)}
\end{aligned}$$

Using the associativity property (2.8) we have

$$\begin{aligned}
\boldsymbol{\mathcal{Y}}(t) &= \sum_{n=-\infty}^{+\infty} \left\{\left\{\boldsymbol{\mathcal{H}}_R(t) * \{\boldsymbol{\mathcal{H}}_C(t) * \boldsymbol{\mathcal{H}}_T(t)\}_{(P)}\right\}_{(Q)} * \boldsymbol{\mathcal{D}}[n]\delta(t-nT)\right\}_{(N)} \\
&= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(t) * \boldsymbol{\mathcal{D}}[n]\delta(t-nT)\}_{(N)} \\
&= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(t-nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)} \qquad (2.55)
\end{aligned}$$

where $\boldsymbol{\mathcal{H}}(t) = \left\{\boldsymbol{\mathcal{H}}_R(t) * \{\boldsymbol{\mathcal{H}}_C(t) * \boldsymbol{\mathcal{H}}_T(t)\}_{(P)}\right\}_{(Q)}$ is the overall system tensor of order $2N$

that couples the input data stream with the output of the receiver.

Sampling the received signal tensor $\boldsymbol{\mathcal{Y}}(t)$ at a rate of $\frac{1}{T}$ gives us the estimate of the data tensor

$$\begin{aligned}\boldsymbol{\mathcal{Y}}[k] &= \boldsymbol{\mathcal{Y}}(kT) \\ &= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(kT - nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)}\end{aligned} \tag{2.56}$$

It is important to note that no physical meaning has been attached to the mathematical domains of the signal and system tensors in the tensor framework. This abstraction makes the framework more general and the mapping from mathematical domains to physical domains is done on a per application basis. The main aim of the tensor framework is to serve as a unifying foundation that can be used to represent and design a variety of different communication systems using several different domains. The task of mapping physical domains to their mathematical counterparts is trivial as the basic structure of the tensor framework remains the same for different systems. Some examples of this are detailed in Chapter 3.

Multidimensional communication systems that exploit several domains have gained traction in recent years. For instance, multidomain index modulation in the context of vehicle-to-infrastructure (V2I), vehicle-to-vehicle (V2V), and high speed train communication systems is discussed in [39]. Here, the transmitted data is mapped to the indices of the various available domains. The domains used for transmission include the indices of transmit antennas, receive antennas, code type, channel impulse response taps and many more (listed in detail in [39]). Apart from this, conventional digital modulation is also used simultaneously to improve performance. Such a system would contain several domains and is well suited for a tensor based representation.

Inter-domain communications for in-house networks where a single transmission scheme can be used for multiple types of wires are gaining popularity in the literature [40]. The International Telecommunication Union (ITU) G.hn standard identified the classical in-house mediums such as power lines, twisted-pairs and coax, to enable broadband data communication [41]. However, a rigorous mathematical model has yet to be created for the study and design of such systems. Moreover, the concept of inter-domain communications can be extended beyond indoor environments. A system that uses different transmission media, multiple sub carriers, multiple time slots and caters to multiple users at once has input and outputs that have five different domains. Using the tensor framework, such a system can be modelled using fifth order tensors for the input and output and a tenth order tensor for the channel between the two.

2.4 Tensor Nyquist Criterion

The Nyquist criterion for distortionless transmission for the scalar case is well known. A waveform $x(t)$ is said to satisfy the Nyquist criterion for signal interval T if

$$x(nT) = \delta_n \tag{2.57}$$

where δ_n is the delta function defined as

$$\delta_n = \begin{cases} 0 & \text{if } n \neq 0 \\ 1 & \text{if } n = 0 \end{cases} \tag{2.58}$$

Denoting the Fourier transform of $x(t)$ by $X(f)$, and using the Poisson Sum Formula [42]

$$\frac{1}{T}\sum_{k=-\infty}^{+\infty} X(f - \frac{k}{T}) = \sum_{k=\infty}^{+\infty} x(nT)e^{-j2\pi f nT} \tag{2.59}$$

we have

$$\frac{1}{T}\sum_{k=-\infty}^{+\infty} X(f - \frac{k}{T}) = 1 \tag{2.60}$$

For the matrix-vector case, a generalized Nyquist criterion has been derived in the literature by authors of [43, 44, 45]. In this section we derive a generalization of (2.57) and (2.60), called the *Tensor Nyquist Criterion*, for the multi-domain case with higher-order signal and system tensors. We then show that the existing generalizations are specific cases of the *Tensor Nyquist Criterion*.

2.4.1 The Tensor Poisson Sum Formula and Nyquist's Criterion for Zero Inter-Symbol Interference

To find the multi-domain criterion for zero inter-symbol interference we begin by generalizing the ordinary Poisson sum formula [42]. Consider a signal tensor $\boldsymbol{\mathcal{A}}(t) \in \mathbb{C}_t^{I_1 \times I_2 \ldots \times I_K}$. Define

$$\boldsymbol{\mathcal{A}}_s(t) = \boldsymbol{\mathcal{A}}(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT) \tag{2.61}$$

$$= \sum_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{A}}(nT)\delta(t - nT) \tag{2.62}$$

Taking the Fourier transform of (2.61) we get

$$\breve{\boldsymbol{\mathcal{A}}}_s(f) = \mathcal{F}[\boldsymbol{\mathcal{A}}(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT)]$$

$$= \breve{\boldsymbol{\mathcal{A}}}(f) * \mathcal{F}[\sum_{n=-\infty}^{+\infty} \delta(t - nT)]$$

$$= \breve{\boldsymbol{\mathcal{A}}}(f) * (\frac{1}{T} \sum_{n=-\infty}^{+\infty} \delta(f - \frac{n}{T})) \tag{2.63}$$

$$= \frac{1}{T} \sum_{n=-\infty}^{+\infty} \breve{\boldsymbol{\mathcal{A}}}(f - \frac{n}{T}) \tag{2.64}$$

Taking the Fourier transform of (2.62) we get

$$\breve{\boldsymbol{\mathcal{A}}}_s(f) = \mathcal{F}[\sum_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{A}}(nT)\delta(t - nT)]$$

$$= \sum_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{A}}(nT)e^{-j2\pi fnT} \tag{2.65}$$

Equating (2.64) and (2.65) we get the Tensor Poisson Sum Formula:

$$\frac{1}{T} \sum_{n=-\infty}^{+\infty} \breve{\boldsymbol{\mathcal{A}}}(f - \frac{n}{T}) = \sum_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{A}}(nT)e^{-j2\pi fnT} \tag{2.66}$$

Expanding (2.56) we get

$$\boldsymbol{\mathcal{Y}}[k] = \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(kT - nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)}$$

$$= \{\boldsymbol{\mathcal{H}}(0), \boldsymbol{\mathcal{D}}[k]\}_{(N)} + \sum_{n=-\infty n\neq k}^{+\infty} \{\boldsymbol{\mathcal{H}}(kT-nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)} \tag{2.67}$$

A sufficient condition to get zero interference between symbols is

$$\boldsymbol{\mathcal{H}}(iT) = \begin{cases} \boldsymbol{\mathcal{H}}(0) & \text{if } i = 0 \\ \boldsymbol{0}_T & \text{if } i \neq 0 \end{cases} \tag{2.68}$$

where $\boldsymbol{0}_T$ is the all zero tensor. Using the Tensor Poisson's sum formula (2.66) we obtain the *Tensor Nyquist Criterion* for zero inter-symbol interference:

$$\frac{1}{T}\sum_{n=-\infty}^{+\infty} \breve{\boldsymbol{\mathcal{H}}}(f - \frac{n}{T}) = \boldsymbol{\mathcal{H}}(0) = \boldsymbol{\mathcal{K}} \tag{2.69}$$

where $\boldsymbol{\mathcal{K}}$ is a non-zero tensor.

Assuming that (2.68) is satisfied we have

$$\boldsymbol{\mathcal{Y}}[k] = \{\boldsymbol{\mathcal{H}}(0), \boldsymbol{\mathcal{D}}[k]\}_{(N)} \tag{2.70}$$

whose elements are

$$\begin{aligned} \boldsymbol{\mathcal{Y}}_{l_1,\ldots,l_N}[k] &= \sum_{i_1,\ldots,i_N} \boldsymbol{\mathcal{H}}_{l_1,\ldots,l_N,i_1,\ldots,i_N}(0)\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] \\ &= \boldsymbol{\mathcal{H}}_{l_1,\ldots,l_N,l_1,\ldots,l_N}(0)\boldsymbol{\mathcal{D}}_{l_1,\ldots,l_N}[k] + \sum_{\substack{i_1,\ldots,i_N \\ i_1\neq l_1,\ldots,i_N\neq l_N}} \boldsymbol{\mathcal{H}}_{l_1,\ldots,l_N,i_1,\ldots,i_N}(0)\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] \end{aligned} \tag{2.71}$$

We see that the first term in (2.71) is a scaled version of the required data symbol and the second term represents intra-symbol interference from other data symbols within the same data tensor. A rather strict condition which will ensure that we are able to retrieve the transmitted data from $\boldsymbol{\mathcal{Y}}[k]$ without any interference is

$$\boldsymbol{\mathcal{H}}_{l_1,\ldots,l_N,i_1,\ldots,i_N}(0) = \begin{cases} 1 & \text{if } i_1 = l_1,\ \ldots,\ i_N = l_N \\ 0 & \text{otherwise} \end{cases} \tag{2.72}$$

This means that the tensor $\boldsymbol{\mathcal{H}}(0)$ is an identity tensor. Combining (2.68) and (2.72) we get

$$\boldsymbol{\mathcal{H}}_{l_1,\ldots,l_N,i_1,\ldots,i_N}(iT) = \delta_{i,0}\prod_{k=1}^{N}\delta_{i_k,l_k} \tag{2.73}$$

where

$$\delta_{n,m} = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \tag{2.74}$$

Using this in (2.69) we get

$$\frac{1}{T}\sum_{n=-\infty}^{+\infty} \breve{\boldsymbol{\mathcal{H}}}(f - \frac{n}{T}) = \boldsymbol{\mathcal{H}}(0) = \boldsymbol{\mathcal{I}}_N \tag{2.75}$$

Even if the strict criterion is not met, it is still possible to recover the transmitted data from $\boldsymbol{\mathcal{Y}}[k]$ if certain conditions are met. Assuming that (2.68) holds then recovering the data reduces to solving the multi-linear tensor system (2.70) for $\boldsymbol{\mathcal{D}}[k]$. If the inverse of $\boldsymbol{\mathcal{H}}(0)$ exists, we have

$$\boldsymbol{\mathcal{D}}[k] = \{\boldsymbol{\mathcal{H}}^{-1}(0), \boldsymbol{\mathcal{Y}}[k]\}_{(N)} \tag{2.76}$$

where $\boldsymbol{\mathcal{H}}^{-1}(0)$ can be approximated by using iterative algorithms, such as the biconjugate gradient or Jacobi methods using tensor computations [21].

If the inverse does not exist, pseudo-inversion can be used to find the solution to the multi-linear system (2.70). The tensors $\hat{\boldsymbol{\mathcal{D}}}[k]$ minimising $\left\|\{\boldsymbol{\mathcal{H}}(0), \boldsymbol{\mathcal{D}}[k]\}_{(N)} - \boldsymbol{\mathcal{Y}}[k]\right\|_F^2$ are called the least-square solutions of (2.70) and $\widetilde{\boldsymbol{\mathcal{D}}}[k] = \min_{\hat{\boldsymbol{\mathcal{D}}}[k]} \left\|\hat{\boldsymbol{\mathcal{D}}}[k]\right\|_F^2$ is called the minimum-norm least square solution of (2.70) [22]. If $\{\boldsymbol{\mathcal{H}}^H(0), \boldsymbol{\mathcal{H}}(0)\}_{(N)}$ is invertible then the least-square solution has a unique minimiser and the multilinear system is solved as [21]

$$\boldsymbol{\mathcal{D}}[k] = \{\{\{\boldsymbol{\mathcal{H}}^H(0), \boldsymbol{\mathcal{H}}(0)\}^{-1}_{(N)}, \boldsymbol{\mathcal{H}}^H(0)\}_{(N)}, \boldsymbol{\mathcal{Y}}[k]\}_{(N)} \tag{2.77}$$

Finally, if such an inversion does not exist, then the minimum-norm least square solution of (2.70) is

$$\boldsymbol{\mathcal{D}}[k] = \{\boldsymbol{\mathcal{H}}^{+}(0), \boldsymbol{\mathcal{Y}}[k]\}_{(N)} \tag{2.78}$$

where $\boldsymbol{\mathcal{H}}^{+}(0)$ is the Moore-Pensore pseudoinverse of $\boldsymbol{\mathcal{H}}(0)$ [22].

2.4.2 Comparison with Existing Generalisations of Nyquist's Criterion

This section surveys existing generalisations of the Nyquist Criterion and compares them to the Tensor Nyquist Criterion presented in this paper. The problem of interference in a multi-carrier system is considered in [43] and [44], that propose a constraint on the overall system impulse response to simultaneously eliminate both ISI and cross-talk. Based on previous work, [45] presents a multidimensional Nyquist criterion.

The general system considered in [43, 44, 45] is a multi-carrier system specified as

$$b_r(t) = \sum_{n=-\infty}^{+\infty} \sum_{m=1}^{M} a_{nm} v_{mr}(t - nT) \quad r = 1, 2, \ldots, M \tag{2.79}$$

where a_{nm} is the data transmitted on the mth sub-carrier during the nth symbol and $v_{mr}(t)$ represents the overall system impulse response consisting of the mth transmit filter, the channel and the rth receive filter, $v_{mr}(t) = p_m(t) * b(t) * r_r(t)$ where $p_m(t)$ is the mth transmit filter, $b(t)$ is the channel and $r_r(t)$ is the rth receive filter. Representing (2.79) in vector matrix form we get

$$\mathbf{b}(t) = \sum_n \mathbf{V}(t - nT)\mathbf{a}_n \tag{2.80}$$

where

$$\mathbf{b}(t) = \begin{bmatrix} b_1(t) \\ b_2(t) \\ \vdots \\ b_M(t) \end{bmatrix} \mathbf{V}(t) = \begin{bmatrix} v_{11}(t) & v_{21}(t) & \ldots & v_{M1}(t) \\ v_{12}(t) & v_{22}(t) & \ldots & v_{M2}(t) \\ \vdots & \vdots & \vdots & \vdots \\ v_{1M}(t) & v_{2M}(t) & \ldots & v_{MM}(t) \end{bmatrix} \mathbf{a}_n = \begin{bmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{nM} \end{bmatrix}$$

Sampling $\mathbf{b}(t)$ at rate $\frac{1}{T}$ we get [43]

$$\mathbf{b}(kT) = \sum_n \mathbf{V}(kT - nT)\mathbf{a}_n \tag{2.82}$$

For no interference, it is required that $\mathbf{b}(kT) = \mathbf{a}_k$. The generalised Nyquist criterion to achieve this is [43, 44]

$$\mathbf{V}(iT) = \delta_{i,0}\mathbf{I} \tag{2.83}$$

where $\mathbf{I}$ is the identity matrix. Taking the Fourier transform of $\mathbf{V}(t)$ and using the Poisson's sum formula we get the frequency domain conditions for zero interference as [44]

$$\frac{1}{T}\sum_{n}\breve{\mathbf{V}}(f-\frac{n}{T}) = \frac{1}{T}\sum_{n}\breve{\mathbf{p}}(f-\frac{n}{T})\breve{b}(f-\frac{n}{T})\breve{\mathbf{r}}^{T}(f-\frac{n}{T}) = \sum_{n=-\infty}^{+\infty}\mathbf{V}(nT)e^{-j2\pi fnT} = \delta_{n,0}\mathbf{I} \tag{2.84}$$

where $\breve{\mathbf{p}}(f) = [\breve{p}_1(f)\ \breve{p}_2(f)\ \ \ldots\ \ \breve{p}_M(f)]^T$, $\breve{b}(f)$ and $\breve{\mathbf{r}}^T(f) = [\breve{r}_1(f)\ \breve{r}_2(f)\ \ \ldots\ \ \breve{r}_M(f)]^T$ are the Fourier transforms of the transmit filters, channel and receive filters respectively. Comparing (2.84) with (2.75) we can see that it is a special case of the generalised Nyquist criterion where the overall system tensor is of size $M \times M$ with components $\mathbf{H}_{i_1,i_2}(t) = p_{i_2}(t) * b(t) * r_{i_1}(t)$ and the data tensor is of order one (a vector of size M) with components $\mathbf{d}[n] = \mathbf{a}_n$.

Chapter 3

Current Systems Modelled Using the Tensor Framework

This chapter presents models for selected waveforms using the tensor framework as examples. Representation using the tensor framework preserves the multi-domain structure of the transmitted data. The purpose of this chapter is to show that different communication systems can be modelled accurately using the tensor framework and that the framework allows extensions to higher domains, such as MIMO, of waveforms that were not originally designed for this. Besides this, it is important to note that our tensor framework is backward compatible and can be used to model systems where the transmitted and received signals have only one (vector) or two (matrix) domains. To this end, we begin with the treatment of OFDM systems where the transmitted data is a vector and is coupled with the received data by a matrix. We then move on to FBMC, where the transmitted data is a matrix and is coupled with the received data by a fourth order system tensor. Finally, we show the representation of GFDM, where the transmitted data is a third order tensor and the system is a sixth order tensor.

The tensor framework presented in this thesis consists of signal and system tensors that are coupled using contracted convolutions or contracted products. Other tensor based approaches to represent specific waveforms can be found in the literature. For example, MIMO OFDM is modeled using a constrainted Block-PARAFAC model in [32]. A similar PARAFAC based model is used for joint channel estimation and data detection in FBMC/OQAM systems in [46] where the received signal is coupled with the transmitted data using a Khatri-Rao product. A tensor model of the GFDM transmit signal using the PARATUCK2 decomposition is presented in [47].

3.1 OFDM

In Release 15 of 3GPP [48] it was agreed that an Orthogonal Frequency Division Multiplexing (OFDM) based scheme will be used for the 5G New Radio (NR) uplink and downlink as the main candidate with Discrete Fourier Transform Spread OFDM (DFT-S-OFDM) being used in some cases. OFDM is a multi-carrier transmission technique which uses F orthogonal sub-carriers simultaneously to transmit data. By making all the sub-channels narrowband, they experience almost flat fading, making equalization very simple. The orthogonality of the sub-carriers ensures that there is no intrinsic inter-carrier interference (ISI). There are F sub-carriers with spacing $F_0 = \dfrac{1}{T}$ that carry data in one OFDM symbol. The transmit signal is

$$x(t) = \sum_{n=-\infty}^{+\infty} \sum_{\kappa_d=1}^{F} d_{n,\kappa_d} w(t-nT) e^{j2\pi(\kappa_d-1)F_0 t} = \sum_{n=-\infty}^{+\infty} \sum_{\kappa_d=1}^{F} d_{n,\kappa_d} p_{T_{\kappa_d}}(t-nT) \tag{3.1}$$

where d_{n,κ_d} is the complex data symbol transmitted during the nth OFDM symbol on the κ_dth ($\kappa_d = 1, \ldots, F$) sub-carrier. The filter $p_{T_{\kappa_d}}(t) = w(t)e^{j2\pi(\kappa_d-1)F_0 t}$ where $w(t)$ is a rectangular window of duration T defined as

$$w(t) = \begin{cases} 1 & \text{for } -\frac{T}{2} \leq t \leq \frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

Under the assumption of an ideal channel, the received signal $r(t)$ is the same as the transmitted signal $x(t)$. The receive filter, for the κ_yth subcarrier, $p_{R_{\kappa_y}}(t) = p^*_{T_{\kappa_y}}(-t)$ is matched to the transmit filter, i.e., $p_{R_{\kappa_y}}(t) = w(-t)e^{j2\pi(\kappa_y-1)F_0 t}$. The output of the receive filter $p_{R_{\kappa_y}}(t)$ is

$$y_{\kappa_y}(t) = x(t) * p_{R_{\kappa_y}}(t)$$

$$= \sum_{n=-\infty}^{+\infty} \sum_{\kappa_d=1}^{F} \int_{-\infty}^{+\infty} d_{n,\kappa_d} w(\tau - nT) e^{j2\pi F_0(\kappa_d-1)\tau} w(-(t-\tau)) e^{j2\pi F_0(\kappa_y-1)(t-\tau)} d\tau \quad (3.3)$$

Sampling $y_{\kappa_y}(t)$ at intervals of T and using $w(t) = w(-t)$ gives

$$y_{\kappa_y}(kT) = \sum_{n=-\infty}^{+\infty} \sum_{\kappa_d=1}^{F} \int_{-\infty}^{+\infty} d_{n,\kappa_d} w(\tau - nT) e^{j2\pi F_0(\kappa_d-1)(\tau)} w(kT-\tau) e^{j2\pi F_0(\kappa_y-1)(kT-\tau)} d\tau \quad (3.4)$$

$$= \sum_{n=-\infty}^{+\infty} \sum_{\kappa_d=1}^{F} \int_{-\infty}^{+\infty} d_{n,\kappa_d} w(\tau - nT) w(\tau - kT) e^{j2\pi F_0(\kappa_d-\kappa_y)(\tau)} d\tau \quad (3.5)$$

Since $w(t)$ is a rectangular window of duration T we have $y_{\kappa_y}(kT) = d_{k,\kappa_y}$.

Using the tensor framework, the complex data to be transmitted on the nth OFDM symbol $\mathcal{D}[n] \in \mathbb{C}_n^F$ is

$$\mathcal{D}[n] = [d_{n,1}, d_{n,2}, \ldots, d_{n,F}]^T \quad (3.6)$$

with components $\mathcal{D}_i[n] = d_{n,i}$ for $i = 1, \ldots, F$. The transmit system tensor $\mathcal{H}_T(t) \in \mathbb{C}_t^{1\times F}$ is

$$\mathcal{H}_T(t) = [p_{T_1}(t), p_{T_2}(t), \ldots, p_{T_F}(t)] \quad (3.7)$$

and has components $\mathcal{H}_{T_{1,i}}(t) = p_{T_i}(t)$ for $i = 1, \ldots, F$. The transmit tensor $\mathcal{H}_T(t)$ converts $\mathcal{D}[n]$ into a signal $\mathcal{X}(t) \in \mathbb{C}_t$. We can write (3.1) using the components of (3.6) and (3.7) as $\mathcal{X}(t) = \sum_{n=-\infty}^{+\infty} \sum_{i=1}^{F} \mathcal{H}_{T_{1,i}}(t - nT)\mathcal{D}_i[n]$ which, in tensor notation, becomes

$$\mathcal{X}(t) = \sum_{n=-\infty}^{+\infty} \{\mathcal{H}_T(t - nT), \mathcal{D}[n]\}_{(1)} \quad (3.8)$$

Under the assumption of an ideal channel the channel tensor is $\mathcal{H}_C(t) \in \mathbb{C}_t = \delta(t)$ and the received signal $\mathcal{R}(t) = \mathcal{X}(t)$. The receive system tensor $\mathcal{H}_R(t) \in \mathbb{C}_t^{F\times 1}$ is

$$\mathcal{H}_R(t) = [p_{R_1}(t), p_{R_2}(t), \ldots, p_{R_F}(t)]^T \quad (3.9)$$

Using (3.7) and (3.9), we get the overall system tensor $\mathcal{H}(t) \in \mathbb{C}_t^{F\times F}$ as $\mathcal{H}(t) = \{\mathcal{H}_R(t) * \mathcal{H}_T(t)\}_{(1)}$ with components

$$\mathcal{H}_{\kappa_y,\kappa_d}(t) = \mathcal{H}_{R_{\kappa_y}}(t) * \mathcal{H}_{T_{\kappa_d}}(t) = p_{R_{\kappa_y}}(t) * p_{T_{\kappa_d}}(t). \quad (3.10)$$

Collecting the outputs of the received filters from (3.3) for all sub-carriers into a vector we may now write the received signal tensor $\boldsymbol{\mathcal{Y}}(t) \in \mathbb{C}_t^F$ as $\boldsymbol{\mathcal{Y}}(t) = \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(t-nT), \boldsymbol{\mathcal{D}}[n]\}_{(1)}$ with components

$$\boldsymbol{\mathcal{Y}}_{\kappa_y}(t) = \sum_{n=-\infty}^{+\infty} \sum_{\kappa_d} \boldsymbol{\mathcal{H}}_{\kappa_y,\kappa_d}(t-nT) \boldsymbol{\mathcal{D}}_{\kappa_d}[n] \tag{3.11}$$

Comparing $\boldsymbol{\mathcal{Y}}(t)$ with (2.55), we can see that this is a specific case when $N = 1$.

3.2 FBMC

Filter Bank Multi-carrier is a scheme considered for 5G. There are two types of FBMC schemes, Staggered Multitone(SMT) and Cosine-Modulated Multitone(CMT) [49]. This section describes SMT, also known as OQAM/OFDM. The number of sub-carriers K is assumed to be even ($K = 2M$) and for two consecutive sub-carriers, the time offset is applied to the imaginary part of the QAM symbol on one sub-carrier while it is applied to the real part of the QAM symbol on the other sub-carrier. The transmitted signal is [50]

$$\begin{aligned} x(t) = \sqrt{2} \sum_{n=-\infty}^{+\infty} \sum_{m=0}^{M-1} & \left(c_{2m,n}^R p(t-nT) + j c_{2m,n}^I p(t - \frac{T}{2} - nT) \right) e^{j2\pi(2m)Ft} \\ & + \left(j c_{2m+1,n}^I p(t-nT) + c_{2m+1,n}^R p(t - \frac{T}{2} - nT) \right) e^{j2\pi(2m+1)Ft} \end{aligned} \tag{3.12}$$

where T is the signalling interval, $F = \frac{1}{T}$ is the sub-carrier spacing, $c_{m,n}^R$ and $c_{m,n}^I$ are the real and imaginary components of the QAM symbol $c_{m,n}$ to be transmitted on the mth sub-carrier during the nth multi-carrier symbol, and $p(t)$ is a real symmetric prototype filter of duration KT where K is the overlapping factor that denotes the number of multi-carrier symbols that overlap in time. We introduce the following notations, as in [50], to simplify the expression in (3.12):

$$d_{2m,2n} = c_{2m,n}^R, \quad d_{2m,2n+1} = c_{2m,n}^I, \quad d_{2m+1,2n} = c_{2m+1,n}^I, \quad d_{2m+1,2n+1} = c_{2m+1,n}^R \tag{3.13}$$

$$\psi_{2m,2n} = 0, \quad \psi_{2m,2n+1} = \frac{\pi}{2}, \quad \psi_{2m+1,2n} = \frac{\pi}{2}, \quad \psi_{2m+1,2n+1} = 0 \tag{3.14}$$

Using (3.13) and (3.14) in (3.12) we get

$$x(t) = \sqrt{2}\sum_{n=-\infty}^{+\infty}\sum_{m=0}^{M-1}\Big(d_{2m,2n}p(t-(2n)\frac{T}{2})e^{\psi_{2m,2n}} + d_{2m,2n+1}p(t-(2n+1)\frac{T}{2})e^{\psi_{2m,2n+1}}\Big)e^{j2\pi(2m)Ft}$$
$$+\Big(d_{2m+1,2n}p(t-(2n)\frac{T}{2})e^{\psi_{2m+1,2n}} + d_{2m+1,2n+1}p(t-(2n+1)\frac{T}{2})e^{\psi_{2m+1,2n+1}}\Big)e^{j2\pi(2m+1)Ft} \tag{3.15}$$

Defining $\lambda_{m,n}(t) = \sqrt{2}p(t-n\frac{T}{2})e^{j2\pi mFt}e^{\psi_{m,n}}$ and substiting in (3.15) we get

$$x(t) = \Big(\sum_{n=-\infty}^{+\infty}\sum_{m=0}^{M-1} d_{2m,2n}\lambda_{2m,2n}(t) + d_{2m,2n+1}\lambda_{2m,2n+1}(t)$$
$$+ d_{2m+1,2n}\lambda_{2m+1,2n}(t) + d_{2m+1,2n+1}\lambda_{2m+1,2n+1}(t)\Big) \tag{3.16}$$

Since $\sum_{k=-P}^{k=Q}(x_{2k}+x_{2k+1}) = \sum_{k=-2P}^{k=2Q+1} x_k$, we have from (3.16)

$$x(t) = \sum_{n=-\infty}^{+\infty}\sum_{m=0}^{2M-1}(d_{m,2n}\lambda_{m,2n}(t) + d_{m,2n+1}\lambda_{m,2n+1}(t)) = \sum_{n=-\infty}^{+\infty}\sum_{m=0}^{2M-1} d_{m,n}\lambda_{m,n}(t)$$
$$= \sum_{n=-\infty}^{+\infty}\sum_{m=0}^{2M-1} d_{m,n}p_m(t-n\frac{T}{2})e^{j\psi_{m,n}} \tag{3.17}$$

where $p_m(t) = \sqrt{2}p(t)e^{j2\pi mFt}$. The received signal is passed through an analysis filter bank (AFB) to separate the data from different sub-carriers. The receive filter for the rth sub-carrier, $p_{R_r}(t)$ is matched to the transmit filter for the rth sub-carrier $p_r(t)$. i.e., we have $p_{R_r}(t) = p_r^*(-t) = \sqrt{2}p(-t)e^{j2\pi mFt}$ ($p^*(-t) = p(-t)$ as the prototype filter $p(t)$ is real). Let

$$\langle\lambda_{m,n}(t), \lambda_{p,q}(t)\rangle = \int_{-\infty}^{+\infty} p_m(t-n\frac{T}{2})e^{j\psi_{m,n}}p_p^*(t-q\frac{T}{2})e^{-j\psi_{p,q}}dt \tag{3.18}$$

where $\langle,\rangle$ denotes the inner product. In a distortion free channel, perfect reconstruction of the data is obtained if the transmit and receive filters satisfy the real orthogonality condition [51]:

$$\Re\{\langle\lambda_{m,n}(t), \lambda_{p,q}(t)\rangle\} = \Re\{\int_{-\infty}^{+\infty} p_m(t-n\frac{T}{2})e^{j\psi_{m,n}}p_p^*(t-q\frac{T}{2})e^{-j\psi_{p,q}}dt\} = \delta_{m,p}\delta_{n,q} \tag{3.19}$$

In other words, this means that for $(m,n) \neq (p,q)$, $\langle\lambda_{m,n}(t), \lambda_{p,q}(t)\rangle$ can be purely imaginary or zero. The received signal after passing through the receive filter corresponding to the

rth sub-carrier is

$$y_r(t) = x(t) * p_r^*(-t) = \sum_{n=-\infty}^{+\infty} \sum_{m=0}^{2M-1} d_{m,n} e^{j\psi_{m,n}} \int_{-\infty}^{+\infty} p_m(\tau - n\frac{T}{2}) p_r^*(\tau - t) d\tau \tag{3.20}$$

Using the tensor framework, the data to be transmitted on the nth multicarrier symbol is a tensor $\mathcal{D}[n] \in \mathbb{C}_n^{2M}$ with components $\mathcal{D}_m[n] = d_{(m-1),n} e^{j\psi_{(m-1),n}}$ and the overall channel is $\mathcal{H}(t) \in \mathbb{C}_t^{2M\times 2M}$ with components $\mathcal{H}_{r,m}(t) = \int_{-\infty}^{+\infty} p_{(m-1)}(\tau) p_{(r-1)}^*(\tau - t) d\tau$. We may now re-write (3.20) in tensor form as

$$\mathcal{Y}_r(t) = \sum_{n=-\infty}^{+\infty} \sum_{m=1}^{2M} \mathcal{H}_{r,m}(t - n\frac{T}{2}) \mathcal{D}_m[n] \tag{3.21}$$

where $\mathcal{Y}_r(t) = y_{(r-1)}(t)$. In tensor notation (3.21) becomes $\mathcal{Y}(t) = \sum_{n=-\infty}^{+\infty} \{\mathcal{H}(t-n\frac{T}{2}), \mathcal{D}[n]\}_{(1)}$. The signal $y_r(t)$ is sampled at intervals of $k\frac{T}{2}$ and multiplied by the phase term $e^{-j\psi_{r,k}}$. This gives

$$\begin{aligned} y_{r,k} = y_r(k\frac{T}{2})e^{-j\psi_{r,k}} &= \sum_n \sum_m d_{m,n} e^{j\psi_{m,n}} e^{-j\psi_{r,k}} \int_{-\infty}^{+\infty} p_m(\tau - n\frac{T}{2}) p_r^*(\tau - k\frac{T}{2}) d\tau \\ &= \sum_n \sum_m d_{m,n} \langle \lambda_{m,n}(t), \lambda_{r,k}(t) \rangle \\ &= \sum_n \sum_m d_{m,n} \Re\{\langle \lambda_{m,n}, (t) \lambda_{r,k}(t) \rangle\} + \sum_n \sum_m d_{m,n} \Im\{\langle \lambda_{m,n}(t), \lambda_{r,k}(t) \rangle\} \end{aligned} \tag{3.22}$$

Using (3.19) we get

$$y_{r,k} = d_{r,k} + j\left(\sum_n \sum_m d_{m,n} \Im\{\langle \lambda_{m,n}(t), \lambda_{r,k}(t) \rangle\} \right) \tag{3.23}$$

Since the interference in (3.23) is imaginary, the estimate of the transmitted data is $\hat{d}_{r,k} = \Re\{y_{r,k}\}$. Using tensor notation, this becomes $\hat{\mathcal{D}}[k] = \Re\{\mathcal{Y}(k\frac{T}{2})\}$.

Next, we consider the MIMO extension of FBMC. In each FBMC symbol, let there be P independent streams of data transmitted using $2M$ sub-carriers, N_T transmit and N_R receive antennas. There are P synthesis filter banks, one for each stream of data. Denote the filter for the mth sub-carrier for the pth synthesis filter bank by $p_{p,m}(t)$. A weight $w_{n_t,p}$ is assigned to the pth SFB output for the n_tth antenna. The weights $w_{n_t,p}$ are the coefficients of a linear precoder that couples the P streams of data with the N_T transmit antennas. Denote the data symbol for the pth stream, mth sub-carrier and nth FBMC symbol by $d_{p,m,n}$. As for the scalar case in (3.17), each data symbol $d_{p,m,n}$ is multiplied by a phase term $e^{j\psi_{m,n}}$. We get the transmit signal from the n_tth antenna as

$$x_{n_t}(t) = \sum_{n=-\infty}^{+\infty} \sum_{m=0}^{2M-1} \sum_{p=0}^{P-1} w_{n_t,p} p_{p,m}(t - n\frac{T}{2}) e^{j\psi_{m,n}} d_{p,m,n} \tag{3.24}$$

Denoting the channel between the n_tth transmit and the n_rth receive antenna by $h_{n_r,n_t}(t)$, the received signal on the n_rth antenna is

$$r_{n_r}(t) = \sum_{n_t} h_{n_r,n_t}(t) * x_{n_t}(t) = \sum_{n_t} h_{n_r,n_t}(t) * (\sum_{n=-\infty}^{+\infty} \sum_{m=0}^{2M-1} \sum_{p=0}^{P-1} w_{n_t,p} p_{p,m}(t - n\frac{T}{2}) e^{j\psi_{m,n}} d_{p,m,n}) \tag{3.25}$$

Let $c_{n_r,p,m}(t) = \sum_{n_t} h_{n_r,n_t}(t) * w_{n_t,p} p_{p,m}(t)$. We can then re-write (3.25) as

$$r_{n_r}(t) = \sum_{n=-\infty}^{+\infty} \sum_{m=0}^{2M-1} \sum_{p=0}^{P-1} c_{n_r,p,m}(t - n\frac{T}{2}) e^{j\psi_{m,n}} d_{p,m,n} \tag{3.26}$$

There are P analysis filter banks (AFB) at the receiver, one corresponding to each transmit SFB, that filter the N_R received signals $r_{n_r}(t)$ and produce outputs $y_{p,m}(t)$. Denote the filter for the pth AFB, n_rth receive antenna and mth sub-carrier by $p_{R_{p,m,n_r}}(t)$. The output of the AFB is

$$y_{p,m}(t) = \sum_{n_r} p_{R_{p,m,n_r}}(t) * r_{n_r}(t) = \sum_{n_r} \sum_{n=-\infty}^{+\infty} \sum_{m=0}^{2M-1} \sum_{p=0}^{P-1} p_{R_{p,m,n_r}}(t) * c_{n_r,p,m}(t - n\frac{T}{2}) e^{j\psi_{m,n}} d_{p,m,n} \tag{3.27}$$

If the AFB is designed to be matched to the combined channel and transmit filter banks, then we have $p_{R_{p,m,n_r}}(t) = c^*_{n_r,p,m}(-t)$. Using the Tensor Framework, the data to be transmitted on the nth multi-carrier symbol is $\boldsymbol{\mathcal{D}}[n] \in \mathbb{C}_n^{P\times 2M}$ with components

$$\boldsymbol{\mathcal{D}}_{p,m}[n] = e^{\psi_{(m-1),n}} d_{(p-1),(m-1),n} \quad \text{for } p = 1, \dots, P; m = 1, \dots, 2M \tag{3.28}$$

As discussed in Chapter 2, the data tensor in the tensor framework may be constellation mapped data symbols or data symbols with some form of precoding. In this case, the data tensor is the latter due to the multiplication by the phase term. The transmit system tensor $\boldsymbol{\mathcal{H}}_T(t) \in \mathbb{C}_t^{N_T\times P\times 2M}$ converts $\boldsymbol{\mathcal{D}}[n]$ into a signal tensor $\boldsymbol{\mathcal{X}}(t) \in \mathbb{C}_t^{N_T}$. The components of the transmit system tensor are

$$\mathcal{H}_{T_{n_t,p,m}}(t) = w_{n_t,(p-1)}p_{(p-1),(m-1)}(t) \quad \text{for } p = 1,\dots,P; m = 1,\dots,2M; n_t = 1,\dots,N_T \tag{3.29}$$

We may now re-write (3.24) in tensor notation as $\mathcal{X}(t) = \{\mathcal{H}_T(t - n\frac{T}{2}), \mathcal{D}[n]\}_{(2)}$. The channel system tensor is $\mathcal{H}_C(t) \in \mathbb{C}_t^{N_R \times N_T}$ whose components

$$\mathcal{H}_{C_{n_r,n_t}}(t) = h_{n_r,n_t}(t) \quad \text{for } n_r = 1,\dots,N_R; n_t = 1,\dots,N_T \tag{3.30}$$

are the channel between the n_tth transmit and n_rth receive antenna. The output of the channel tensor $\mathcal{R}(t) \in \mathbb{C}_t^{N_R}$ is $\mathcal{R}(t) = \{\mathcal{H}_C(t) * \mathcal{X}(t)\}_{(1)}$ and the combined channel and transmit system tensor is $\mathcal{C}(t) = \{\mathcal{H}_C(t) * \mathcal{H}_T(t)\}_{(1)} \in \mathbb{C}_t^{N_R \times P \times 2M}$ with components $\mathcal{C}_{n_r,p,m}(t) = \sum_{n_t} \mathcal{H}_{C_{n_r,n_t}}(t) * \mathcal{H}_{T_{n_t,p,m}}(t) = \sum_{n_t} h_{n_r,n_t}(t) * w_{n_t,(p-1)}p_{(p-1),(m-1)}(t) = c_{n_r,(p-1),(m-1)}(t)$. If a system matched to the combined channel and synthesis filter bank is used, then the receive system tensor $\mathcal{H}_R(t) \in \mathbb{C}_t^{P \times 2M \times N_R}$ is $\mathcal{H}_R(t) = \mathcal{C}^H(-t)$ and converts $\mathcal{R}(t)$ into a signal tensor $\mathcal{Y}(t) \in \mathbb{C}_t^{P \times 2M}$. The overall system tensor $\mathcal{H}(t) \in \mathbb{C}_t^{P \times 2M \times P \times 2M}$ is

$$\mathcal{H}(t) = \{\mathcal{H}_R(t) * \{\mathcal{H}_C(t) * \mathcal{H}_T(t)\}_{(1)}\}_{(1)} = \{\mathcal{C}^H(-t) * \mathcal{C}(t)\}_{(1)} \tag{3.31}$$

with components $\mathcal{H}_{p',m',p,m}(t) = \sum_{n_r=1}^{N_R} \mathcal{C}^*_{p',m',n_r}(-t) * \mathcal{C}_{n_r,p,m}(t) = c^*_{(p'-1),(m'-1),n_r}(-t) * c_{n_r,(p-1),(m-1)}(t)$. We may now re-write (3.27), for the case when $p_{R_{p,m,n_r}}(t) = c^*_{p,m,n_r}(-t)$, as

$$\begin{aligned} y_{(p'-1),(m'-1)}(t) &= \sum_{n=-\infty}^{+\infty} \sum_{n_r=1}^{N_R} \sum_{m=1}^{2M} \sum_{p=1}^{P} c^*_{(p'-1),(m'-1),n_r}(-t) * c_{nr,(p-1),(m-1)}(t - n\frac{T}{2}) e^{j\psi_{(m-1),n}} d_{(p-1),(m-1),n} \\ &= \sum_{n=-\infty}^{+\infty} \sum_{m=1}^{2M} \sum_{p=1}^{P} \mathcal{H}_{p',m',p,m}(t - n\frac{T}{2}) \mathcal{D}_{p,m}[n] = \mathcal{Y}_{p',m'}(t) \end{aligned} \tag{3.32}$$

Writing $\mathcal{Y}_{p',m'}(t) = \sum_{n=-\infty}^{+\infty} \sum_{m=1}^{2M} \sum_{p=1}^{P} \mathcal{H}_{p',m',p,m}(t - n\frac{T}{2}) \mathcal{D}_{p,m}[n]$ in tensor notation gives

$$\mathcal{Y}(t) = \sum_{n=-\infty}^{+\infty} \{\mathcal{H}(t - nT_0), \mathcal{D}[n]\}_{(2)} \tag{3.33}$$

where $T_0 = \frac{T}{2}$. Sampling (3.33) at intervals kT_0 we get

$$\mathcal{Y}[k] = \mathcal{Y}(kT_0) = \sum_{n=-\infty}^{+\infty} \{\mathcal{H}((k-n)T_0), \mathcal{D}[n]\}_{(2)} \tag{3.34}$$

An estimate of the data tensor $\hat{\boldsymbol{\mathcal{D}}}[k] \in \mathbb{C}_k^{P \times 2M}$ is found by feeding $\boldsymbol{\mathcal{Y}}[k]$ into a tensor detector $\boldsymbol{\mathcal{G}}[k] \in \mathbb{C}_k^{P \times 2M \times P \times 2M}$, some of which are derived in the next section, such that $\hat{\boldsymbol{\mathcal{D}}}[k] = \{\boldsymbol{\mathcal{G}}[k], \boldsymbol{\mathcal{Y}}[k]\}_{(2)}$.

3.3 Generalized Frequency Division Multiplexing (GFDM)

Generalized frequency division multiplexing (GFDM) [52] is a block-based multicarrier modulation scheme that employs circular filtering. Consider a time-frequency resource block of duration T and bandwidth B. The available bandwidth is divided into K equally-spaced subcarriers with subcarrier spacing $\Delta_f = \frac{B}{K}$ [52], and the time slot is divided into M subsymbols with subsymbol spacing $T_{\text{sub}} = \frac{T}{M}$. The relation between the subcarrier spacing and the subsymbol spacing is given by $\Delta_f T_{\text{sub}} = 1$. The data symbol transmitted on the mth subsymbol, and kth sub-carrier is modulated by a pulse $p_{k,m}(t)$ given by

$$p_{k,m}(t) = w_T(t) p_T(t - mT_{\text{sub}}) e^{j2\pi\Delta_f kt} \quad \text{for } m = 0, \ldots, M-1; k = 0, \ldots, K-1 \tag{3.35}$$

where $p_T(t)$ is a prototype periodic pulse shape of period T, T_{sub} is the duration of one subsymbol, $T = KT_{\text{sub}}$ is the duration of the entire GFDM symbol, and $w_T(t)$ is a rectangular window of duration T such that

$$w_T(t) = \begin{cases} 1 & \text{for } t \in [0,\ T] \\ 0 & \text{elsewhere} \end{cases} \tag{3.36}$$

The rectangular window $w_T(t)$ is used to limit the final modulating pulse $p_{k,m}(t)$ to the interval $t \in [0,\ T]$. A GFDM block hence comprises of pulse shapes generated by time and frequency shifts of a periodic prototype pulse shape followed by multiplication by a finite time window. The transmit signal is given by

$$x(t) = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} p_{k,m}(t - nT) d_{k,m,n} \tag{3.37}$$

where $d_{k,m,n}$ is the complex data transmitted during the mth subsymbol on the kth subcarrier and the nth GFDM symbol. Denoting the channel by $c(t)$ and the additive white Gaussian noise (AWGN) process by $v(t)$, the received signal is

$$r(t) = c(t) * x(t) + v(t)$$

$$= \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \Big(c(t) * p_{k,m}(t-nT) \Big) d_{k,m,n} + v(t) \tag{3.38}$$

Defining second order tensors $\mathfrak{D}[n] \in \mathbb{C}_n^{K\times M}$ and $\mathfrak{H}(t) \in \mathbb{C}_t^{K\times M}$ with components $\mathfrak{D}_{k,m}[n] = d_{k,m,n}$ and $\mathfrak{H}_{k,m}(t) = c(t) * p_{k,m}(t)$, we re-write (3.38) using the tensor representation as

$$r(t) = \sum_{n=-\infty}^{+\infty} \{\mathfrak{H}(t-nT), \mathfrak{D}[n]\}_{(2)} + v(t) \tag{3.39}$$

The signal $r(t)$ is the input to a system tensor $\mathfrak{H}_R(t) \in \mathbb{C}_t^{K\times M\times 1}$ where the singleton dimension is used to indicate that the input to this system is a scalar function $(r(t))$. If there is a bank of filters matched to the transmit filters at the receiver then we have $\mathfrak{H}_{R_{k,m}}(t) = p^*_{k,m}(-t)$ whose output $\mathfrak{Y}(t) \in \mathbb{C}_t^{K\times M}$ has components $\mathfrak{Y}_{k,m}(t) = p^*_{k,m}(-t) * r(t)$. Extending this to the MIMO case, let P independent streams of data be transmitted using K sub-carriers, M sub-symbols and N_T transmit antennas. Let there be N_R receive antennas. Assuming that different banks of filters are used at each transmit and receive antenna, we get the signal transmitted by the n_tth antenna as

$$x_{n_t}(t) = \sum_{n=-\infty}^{+\infty} \sum_{p=1}^{P} \sum_{k=1}^{K} \sum_{m=1}^{M} w_{n_t,p} d_{n,p,k,m} p_{T_{n_t,k,m}}(t-nT) \tag{3.40}$$

where $d_{n,p,k,m}$ is the data transmitted on the nth GFDM symbol, during the mth sub-symbol, on the kth sub-carrier and on the pth stream. Using the Tensor Framework, the complex data to be transmitted on the nth GFDM symbol is $\mathfrak{D}[n] \in \mathbb{C}_n^{P\times K\times M}$ with components $\mathfrak{D}_{p,k,m}[n] = d_{n,p,k,m}$. The transmit tensor $\mathfrak{H}_T(t) \in \mathbb{C}_t^{N_T\times P\times K\times M}$ whose (n_t, p, k, m)th component is $\mathfrak{H}_{T_{n_t,p,k,m}} = w_{n_t,p} p_{T_{n_t,k,m}}(t)$ converts the data tensor into a signal tensor $\mathfrak{X}(t) \in \mathbb{C}_t^{N_T}$. We write (3.40) using tensor notation as $\mathfrak{X}(t) = \sum_{n=-\infty}^{+\infty} \{\mathfrak{H}_T(t-nT), \mathfrak{D}[n]\}_{(3)}$. The channel system tensor is $\mathfrak{H}_C(t) \in \mathbb{C}_t^{N_R\times N_T}$ whose components $\mathfrak{H}_{C_{n_r,n_t}}(t)$ are the channel between the n_tth transmit and n_rth receive antenna. The output of the channel $\mathfrak{R}(t) \in \mathbb{C}_t^{N_R}$ is $\mathfrak{R}(t) = \{\mathfrak{H}_C(t) * \mathfrak{X}(t)\}_{(1)}$ and the combined channel and transmit tensor $\mathfrak{C}(t) \in \mathbb{C}_t^{N_R\times P\times K\times M}$ is $\mathfrak{C}(t) = \{\mathfrak{H}_C(t) * \mathfrak{H}_T(t)\}_{(1)}$. If a matched filter is used at the receiver, the receive system tensor $\mathfrak{H}_R(t) = \mathfrak{C}^H(-t) \in \mathbb{C}_t^{P\times K\times M\times N_R}$ converts $\mathfrak{R}(t)$ into a signal tensor $\mathfrak{Y}(t) \in \mathbb{C}_t^{P\times F\times K}$. The overall system tensor $\mathfrak{H}(t) \in \mathbb{C}_t^{P\times K\times M\times P\times K\times M}$ is

$$\mathfrak{H}(t) = \{\mathfrak{H}_R(t) * \{\mathfrak{H}_C(t) * \mathfrak{H}_T(t)\}_{(1)}\}_{(1)} = \{\mathfrak{C}^H(-t) * \mathfrak{C}(t)\}_{(1)} \tag{3.41}$$

and

$$\boldsymbol{\mathcal{Y}}(t) = \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(t-nT), \boldsymbol{\mathcal{D}}[n]\}_{(3)} \tag{3.42}$$

Sampling (3.42) at intervals of T gives $\boldsymbol{\mathcal{Y}}[k] = \boldsymbol{\mathcal{Y}}(kT) = \{\boldsymbol{\mathcal{H}}(kT-nT), \boldsymbol{\mathcal{D}}[n]\}_{(3)}$. An estimate of the data tensor $\hat{\boldsymbol{\mathcal{D}}}[k] \in \mathbb{C}_k^{P\times K\times M}$ is found by feeding $\boldsymbol{\mathcal{Y}}[k]$ into a tensor detector $\boldsymbol{\mathcal{G}}[k] \in \mathbb{C}_k^{P\times K\times M\times P\times K\times M}$, some of which are derived in the next section, such that $\hat{\boldsymbol{\mathcal{D}}}[k] = \{\boldsymbol{\mathcal{G}}[k], \boldsymbol{\mathcal{Y}}[k]\}_{(3)}$

Chapter 4

Detection Methods

Equalization methods for single input single output (SISO) systems have been extensively studied in the literature [53]. Such equalizers are represented by discrete or continuous scalar functions. For the case of more than one input, there have been several publications on multi-channel extentions in the form of MIMO equalizers. W. Etten presents a maximum likelihood receiver for a MIMO transmission system in [54]. The notion of a matrix matched filter is defined in [55] where a MIMO zero forcing linear equalizer is derived. Duel-Hallen presents an optimal (in the minimum mean squared error sense) linear multichannel equalizer in [56]. A decision feedback equalizer is also described in [56] which is derived using factorization of matrix spectra. J. G. Proakis et. al. present optimal and sub-optimal detectors for multiple antenna systems for both frequency-selective and frequency flat fading in [57].

Matrices are well suited for the design of systems where there are several inputs and outputs belonging to the same domain such as uncoded multiple antenna systems. However, tensors are a more natural tool for systems with multiple domains at both the input and output. For example, systems that employ space-time-frequency coding have an inherent multidimensional structure and are better represented using tensors. In this chapter we derive some tensor based equalization methods for multi-domain communication systems. We present multi-domain tensor linear equalizers that are optimized based on the peak distortion and minimum mean squared error criterion. Further, we present non-linear equalization methods in the form of decision feedback equalizers (DFE). Finally, we present some performance results for the equalizers described above and also show that equalizers for MIMO and scalar systems are a specific case of these tensor based equalizers.

As we have seen from our discussions in Chapter 2 and Chapter 3, several current communication systems are inherently multidimensional in nature and hence warrant the use of mathematical tools that maintain this structure. Many of the developments in the tensor based equalizers presented in this chapter follow a similar line of thought as those of vector-matrix based systems allowing an extension of the existing understanding of such systems into multiple domains. The aim of this chapter is to build a foundation that allows one to exploit the benefits of using a tensor based approach, an example of which is shown in Chapter 5 in the form of Tensor Partial Response Signalling (TPRS), while at the same time dispelling the perceived complexity of tensor mathematics.

4.1 Preliminaries

We begin with some important definitions that will be used throughout this chapter.

Definition 4.1.1. *D-transform of a Discrete Tensor Sequence:* The D-transform of a scalar sequence $x[k]$ is defined as

$$\breve{x}(D) = \sum_k x[k] D^k \tag{4.1}$$

where D is the delay operator. A discrete tensor sequence $\boldsymbol{\mathcal{X}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ is a function tensor with a discrete argument k. The D-transform of $\boldsymbol{\mathcal{X}}[k]$ is a tensor of the D-transform of its components defined as

$$\breve{\boldsymbol{\mathcal{X}}}(D) = \sum_k \boldsymbol{\mathcal{X}}[k] D^k \tag{4.2}$$

with components $\breve{\boldsymbol{\mathcal{X}}}_{i_1,\ldots,i_N}(D) = \sum_k \boldsymbol{\mathcal{X}}_{i_1,\ldots,i_N}[k] D^k$

Definition 4.1.2. *Random Tensor:* A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times \ldots \times I_N}$ is said to be *random* if its components $\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_N}$ are random variables. Similarly, a function tensor $\boldsymbol{\mathcal{A}}(x) \in \mathbb{C}_x^{I_1 \times \ldots \times I_N}$ is a random function tensor if its components are random processes.

Definition 4.1.3. *Mean:* The mean of a random tensor sequence $\mathcal{X}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ is defined as

$$\mu_{\mathcal{X}}[k] = \mathbb{E}\left[\mathcal{X}[k]\right] \tag{4.3}$$

with components $\mu_{\mathcal{X}_{i_1,\ldots,i_N}}[k] = \mathbb{E}\left[\mathcal{X}_{i_1,\ldots,i_N}[k]\right]$

Definition 4.1.4. *Auto-correlation and Cross-correlation of a Random Tensor Sequence:* The *auto-correlation* function of a random tensor sequence $\mathcal{X}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ is a tensor $\mathcal{R}_{\mathcal{X}}[k,i] \in \mathbb{C}_{(k,i)}^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ defined as

$$\mathcal{R}_{\mathcal{X}}[k,i] = \mathbb{E}\left[\mathcal{X}[k] \circ \mathcal{X}^*[k-i]\right] \tag{4.4}$$

The pseudo-diagonal elements of $\mathcal{R}_{\mathcal{X}}[k,i]$, $\mathcal{R}_{\mathcal{X}_{i_1,\ldots,i_N,i_1,\ldots,i_N}}[k,i]$, are the auto-correlation functions of $\mathcal{X}_{i_1,\ldots,i_N}[k]$ and the cross-correlation between two different components $\mathcal{X}_{i_1,\ldots,i_N}[k]$ and $\mathcal{X}_{i'_1,\ldots,i'_N}[k]$ is $\mathcal{R}_{\mathcal{X}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}}[k,i]$. The *cross-correlation* of two random tensor sequences $\mathcal{X}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ and $\mathcal{Y}[k] \in \mathbb{C}_k^{J_1 \times \ldots \times J_M}$ is a tensor $\mathcal{R}_{\mathcal{X},\mathcal{Y}}[k,i] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_M}$ defined as

$$\mathcal{R}_{\mathcal{X},\mathcal{Y}}[k,i] = \mathbb{E}\left[\mathcal{X}[k] \circ \mathcal{Y}^*[k-i]\right] \tag{4.5}$$

where $\mathcal{R}_{\mathcal{X},\mathcal{Y}_{i_1,\ldots,i_N,j_1,\ldots,j_M}}[k,i] = \mathbb{E}\left[\mathcal{X}_{i_1,\ldots,i_N}[k]\mathcal{Y}_{j_1,\ldots,j_M}[k-i]\right]$.

Definition 4.1.5. *Wide Sense Stationary Tensor Sequence:* A random tensor sequence $\mathcal{X}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ is said to be wide sense stationary (WSS) if its mean $\mathbb{E}\left[\mathcal{X}[k]\right]$ is independent of k and its auto-correlation $\mathbb{E}\left[\mathcal{X}[k] \circ \mathcal{X}^*[k-i]\right]$ depends only on i.Two random tensor sequences $\mathcal{X}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ and $\mathcal{Y}[k] \in \mathbb{C}_k^{J_1 \times \ldots \times J_M}$ are jointly WSS if both $\mathcal{X}[k]$ and $\mathcal{Y}[k]$ are WSS and their cross-correlation $\mathbb{E}\left[\mathcal{X}[k] \circ \mathcal{Y}^*[k-i]\right]$ depends only on i In the rest of this thesis, auto-correlation and cross-correlation tensors of WSS and jointly WSS tensor sequences are indexed by one variable ($\mathcal{R}_{\mathcal{X}}[i]$ and $\mathcal{R}_{\mathcal{X},\mathcal{Y}}[i]$ respectively). It can be shown that if the input to a linear time invariant system tensor is a WSS tensor sequence, then the output is also WSS and the input and output tensor sequences are jointly WSS.

Proof. The output of the tensor system $\mathcal{H}[k] \in \mathbb{C}_k^{J_1 \times \ldots \times J_M \times I_1 \times \ldots \times I_N}$ to a WSS input $\mathcal{X}[k] \in$

$\mathbb{C}_k^{I_1 \times \ldots \times I_N}$ is

$$\boldsymbol{\mathcal{Y}}[k] = \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[k-n], \boldsymbol{\mathcal{X}}[n]\}_{(N)} \tag{4.6}$$

The mean of $\boldsymbol{\mathcal{Y}}[k]$ is

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\mathcal{Y}}[k]\right] &= \mathbb{E}\Big[\sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[k-n], \boldsymbol{\mathcal{X}}[n]\}_{(N)}\Big] &(4.7)\\
&= \sum_{n=-\infty}^{+\infty} \mathbb{E}\left[\{\boldsymbol{\mathcal{H}}[k-n], \boldsymbol{\mathcal{X}}[n]\}_{(N)}\right] &(4.8)\\
&= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[k-n], \mathbb{E}\left[\boldsymbol{\mathcal{X}}[n]\right]\}_{(N)} &(4.9)\\
&= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[k-n], \mu_{\boldsymbol{\mathcal{X}}}\}_{(N)} &(4.10)\\
&= \left\{\sum_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{H}}[k-n], \mu_{\boldsymbol{\mathcal{X}}}\right\}_{(N)} &(4.11)
\end{aligned}$$

Since the summation is over all values of n, $\sum\limits_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{H}}[k-n]$ does not depend on the value of k. Hence, $\mathbb{E}\left[\boldsymbol{\mathcal{Y}}[k]\right]$ does not depend on k.

The auto-correlation of $\boldsymbol{\mathcal{Y}}[k]$ is

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\mathcal{Y}}[k] \circ \boldsymbol{\mathcal{Y}}^*[k-i]\right] &= \mathbb{E}\left[\Big(\sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{X}}[k-n]\}_{(N)}\Big) \circ \Big(\sum_{m=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[m], \boldsymbol{\mathcal{X}}[k-i-m]\}_{(N)}\Big)^*\right]\\
&= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \mathbb{E}\left[\left(\{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{X}}[k-n]\}_{(N)}\right) \circ \left(\{\boldsymbol{\mathcal{H}}[m], \boldsymbol{\mathcal{X}}[k-i-m]\}_{(N)}\right)^*\right]
\end{aligned} \tag{4.12}$$

Using (2.35), (4.12) becomes

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\mathcal{Y}}[k] \circ \boldsymbol{\mathcal{Y}}^*[k-i]\right] &= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \mathbb{E}\left[\left(\{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{X}}[k-n]\}_{(N)}\right) \circ \left(\{\boldsymbol{\mathcal{X}}[k-i-m], \boldsymbol{\mathcal{H}}^T[m]\}_{(N)}\right)^*\right]\\
&= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \mathbb{E}\left[\left(\{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{X}}[k-n]\}_{(N)}\right) \circ \left(\{\boldsymbol{\mathcal{X}}^*[k-i-m], \boldsymbol{\mathcal{H}}^H[m]\}_{(N)}\right)\right]
\end{aligned}$$

Using the associativity property of the contracted and outer products from (2.8) and (2.20) we get

$$= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \mathbb{E}\left[\left\{\left\{\boldsymbol{\mathcal{H}}[n], \left(\boldsymbol{\mathcal{X}}[k-n] \circ \boldsymbol{\mathcal{X}}^*[k-i-m]\right)\right\}_{(N)}, \boldsymbol{\mathcal{H}}^H[m]\right\}_{(N)}\right]$$

$$= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \left\{\left\{\boldsymbol{\mathcal{H}}[n], \mathbb{E}\left[\boldsymbol{\mathcal{X}}[k-n] \circ \boldsymbol{\mathcal{X}}^*[k-i-m]\right]\right\}_{(N)}, \boldsymbol{\mathcal{H}}^H[m]\right\}_{(N)}$$

$$= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \left\{\left\{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{X}}}[m+i-n]\right\}_{(N)}, \boldsymbol{\mathcal{H}}^H[m]\right\}_{(N)} \tag{4.13}$$

We can see that the RHS of (4.13) depends only on i and hence the LHS also depends only on i. Since the mean of $\boldsymbol{\mathcal{Y}}[k]$ is constant and its auro-correlation depends only on the difference i, it is WSS. The cross-correlation between $\boldsymbol{\mathcal{X}}[k]$ and $\boldsymbol{\mathcal{Y}}[k]$ is

$$\mathbb{E}\left[\boldsymbol{\mathcal{Y}}[k] \circ \boldsymbol{\mathcal{X}}^*[k-i]\right] = \mathbb{E}\left[\Big(\sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{X}}[k-n]\}_{(N)}\Big) \circ \boldsymbol{\mathcal{X}}^*[k-i]\right]$$

$$= \sum_{n=-\infty}^{+\infty} \mathbb{E}\left[\Big(\{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{X}}[k-n]\}_{(N)}\Big) \circ \boldsymbol{\mathcal{X}}^*[k-i]\right]$$

Using the associativity property of the contracted and outer products from equations (2.8) and (2.20) we get

$$= \sum_{n=-\infty}^{+\infty} \mathbb{E}\left[\left\{\boldsymbol{\mathcal{H}}[n], \left(\boldsymbol{\mathcal{X}}[k-n] \circ \boldsymbol{\mathcal{X}}^*[k-i]\right)\right\}_{(N)}\right]$$

$$= \sum_{n=-\infty}^{+\infty} \left\{\boldsymbol{\mathcal{H}}[n], \mathbb{E}\left[\boldsymbol{\mathcal{X}}[k-n] \circ \boldsymbol{\mathcal{X}}^*[k-i]\right]\right\}_{(N)}$$

$$= \sum_{n=-\infty}^{+\infty} \left\{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{X}}}[i-n]\right\}_{(N)} \tag{4.14}$$

The RHS of (4.14) depends only on i and hence the cross-correlation of $\boldsymbol{\mathcal{X}}[k]$ and $\boldsymbol{\mathcal{Y}}[k]$ also depends only on i. Thus, the output and input are jointly WSS. $\square$

Definition 4.1.6. *Spectrum and Cross-spectrum of a tensor sequence:* The *spectrum* of a WSS tensor sequence $\boldsymbol{\mathfrak{X}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ is a tensor $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D) \in \mathbb{C}_D^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ defined as

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D) = \sum_i \boldsymbol{\mathfrak{R}}_{\boldsymbol{\mathfrak{X}}}[i] D^i \tag{4.15}$$

The pseudo-diagonal elements of $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D)$, $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}}(D)$, are the spectra of $\boldsymbol{\mathfrak{X}}_{i_1,\ldots,i_N}$ and the cross-spectrum between two different components $\boldsymbol{\mathfrak{X}}_{i_1,\ldots,i_N}$ and $\boldsymbol{\mathfrak{X}}_{i'_1,\ldots,i'_N}$ is $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}}(D)$.
The *cross-spectrum* of two tensor sequences $\boldsymbol{\mathfrak{X}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ and $\boldsymbol{\mathfrak{Y}}[k] \in \mathbb{C}_k^{J_1 \times \ldots \times J_M}$ is a tensor $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}},\boldsymbol{\mathfrak{Y}}}(D) \in \mathbb{C}_D^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_M}$ defined as

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}},\boldsymbol{\mathfrak{Y}}}(D) = \sum_i \boldsymbol{\mathfrak{R}}_{\boldsymbol{\mathfrak{X}},\boldsymbol{\mathfrak{Y}}}[i] D^i \tag{4.16}$$

where $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}},\boldsymbol{\mathfrak{Y}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}}(D) = \sum_i \boldsymbol{\mathfrak{R}}_{\boldsymbol{\mathfrak{X}},\boldsymbol{\mathfrak{Y}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}}[i] D^i$.
For example, consider a tensor $\boldsymbol{\mathcal{A}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ such that all the components $\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N}[k]$ are un-correlated and have average power $\sigma^2_{\boldsymbol{\mathcal{A}}}$. The auto-correlation of $\boldsymbol{\mathcal{A}}[k]$ is

$$\boldsymbol{\mathfrak{R}}_{\boldsymbol{\mathcal{A}}}[i] = \sigma^2_{\boldsymbol{\mathcal{A}}} \boldsymbol{\mathfrak{I}}_N \delta(i) \tag{4.17}$$

and the spectrum of $\boldsymbol{\mathcal{A}}[k]$ is

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathcal{A}}}(D) = \sigma^2_{\boldsymbol{\mathcal{A}}} \boldsymbol{\mathfrak{I}}_N \tag{4.18}$$

where $\boldsymbol{\mathfrak{I}}_N$ is the identity tensor of size $I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N$

Definition 4.1.7. *Discrete System tensors:* A discrete system tensor is a tensor $\boldsymbol{\mathfrak{H}}[n] \in \mathbb{C}_n^{I_1 \times \ldots I_N \times J_1 \ldots \times J_M}$ that couples an input tensor sequence $\boldsymbol{\mathfrak{X}}[n] \in \mathbb{C}_n^{J_1 \times \ldots \times J_M}$ with an output tensor sequence $\boldsymbol{\mathfrak{Y}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ through a discrete contracted convolution defined as:

$$\boldsymbol{\mathfrak{Y}}[k] = \sum_n \{\boldsymbol{\mathfrak{H}}[n], \boldsymbol{\mathfrak{X}}[k-n]\}_{(M)} \tag{4.19}$$

Taking the D-transform of (4.19) we get

$$\begin{aligned}
\breve{\boldsymbol{\mathfrak{Y}}}(D) &= \sum_k \boldsymbol{\mathfrak{Y}}[k] D^k \\
&= \sum_k \left(\sum_n \{\boldsymbol{\mathfrak{H}}[n], \boldsymbol{\mathfrak{X}}[k-n]\}_{(M)} \right) D^k \\
&= \sum_k \left(\sum_n \{\boldsymbol{\mathfrak{H}}[n], \boldsymbol{\mathfrak{X}}[k-n]\}_{(M)} \right) D^{k-n} D^n
\end{aligned}$$

$$
\begin{aligned}
&= \sum_n \{\mathcal{H}[n], \Big(\sum_k \mathcal{X}[k-n]D^{k-n}\Big)\}_{(M)} D^n \\
&= \sum_n \{\mathcal{H}[n]D^n, \breve{\mathcal{X}}(D)\}_{(M)} \\
&= \{\Big(\sum_n \mathcal{H}[n]D^n\Big), \breve{\mathcal{X}}(D)\}_{(M)} \\
&= \{\breve{\mathcal{H}}(D), \breve{\mathcal{X}}(D)\}_{(M)} \qquad (4.20)
\end{aligned}
$$

Next, we show that if the input $\mathcal{X}[k]$ to the system $\mathcal{H}[k]$ is wide sense stationary (WSS), then we have

$$
\breve{\mathcal{S}}_{\mathcal{Y}}(D) = \{\breve{\mathcal{H}}(D), \{\breve{\mathcal{S}}_{\mathcal{X}}(D), \breve{\mathcal{H}}^H(D^{-1})\}_{(M)}\}_{(M)} \qquad (4.21)
$$

where $\breve{\mathcal{S}}_{\mathcal{X}}(D)$ is the spectrum of $\mathcal{X}[k]$. Further, $\breve{\mathcal{X}}(D)$ and the output $\breve{\mathcal{Y}}(D)$ are jointly WSS with cross spectrum

$$
\breve{\mathcal{S}}_{\mathcal{Y},\mathcal{X}}(D) = \{\breve{\mathcal{H}}(D), \breve{\mathcal{S}}_{\mathcal{X}}(D)\}_{(M)} \qquad (4.22)
$$

Proof. The auto-correlation tensor of the output $\mathcal{Y}[k]$ has components

$$
\begin{aligned}
\mathcal{R}_{\mathcal{Y}_{m_1,\dots,m_N,n_1,\dots,n_N}}(i) &= \mathbb{E}\big[\mathcal{Y}_{m_1,\dots,m_N}[k] \circ \mathcal{Y}^*_{n_1,\dots,n_N}[k-i]\big] \\
&= \mathbb{E}\big[\big(\sum_m \sum_{j_1} \dots \sum_{j_M} \mathcal{H}_{m_1,\dots,m_N,j_1,\dots,j_M}[m] \mathcal{X}_{j_1,\dots,j_M}[k-m]\big) \\
&\quad \cdot \big(\sum_n \sum_{p_1} \dots \sum_{p_M} \mathcal{H}^*_{n_1,\dots,n_N,p_1,\dots,p_M}[n] \mathcal{X}^*_{p_1,\dots,p_M}[k-i-n]\big)\big] \qquad (4.23) \\
&= \sum_m \sum_{j_1} \dots \sum_{j_M} \mathcal{H}_{m_1,\dots,m_N,j_1,\dots,j_M}[m] \\
&\quad \cdot \sum_n \sum_{p_1} \dots \sum_{p_M} \mathcal{H}^*_{n_1,\dots,n_N,p_1,\dots,p_M}[n] \mathcal{R}_{\mathcal{X}_{j_1,\dots,j_M,p_1,\dots,p_M}}[n+i-m] \qquad (4.24)
\end{aligned}
$$

Taking the D-Transform of $\mathcal{R}_{\mathcal{Y}_{m_1,\dots,m_N,n_1,\dots,n_N}}[i]$ we get

$$
\begin{aligned}
\breve{\mathcal{S}}_{\mathcal{Y}_{m_1,\dots,m_N,n_1,\dots,n_N}}(D) &= \sum_i \mathcal{R}_{\mathcal{Y}_{m_1,\dots,m_N,n_1,\dots,n_N}}[i] D^i \\
&= \sum_i \Bigg(\sum_m \sum_{j_1} \dots \sum_{j_M} \mathcal{H}_{m_1,\dots,m_N,j_1,\dots,j_M}[m] \\
&\quad \cdot \sum_n \sum_{p_1} \dots \sum_{p_M} \mathcal{H}^*_{n_1,\dots,n_N,p_1,\dots,p_M}[n] \mathcal{R}_{\mathcal{X}_{j_1,\dots,j_M,p_1,\dots,p_M}}[n+i-m] \Bigg) D^i
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_i \Bigg(\sum_m \sum_{j_1} \dots \sum_{j_M} \mathcal{H}_{m_1,\dots,m_N,j_1,\dots,j_M}[m] \\
&\cdot \sum_n \sum_{p_1} \dots \sum_{p_M} \mathcal{H}^*_{n_1,\dots,n_N,p_1,\dots,p_M}[n] \mathcal{R}_{\mathcal{X}_{j_1,\dots,j_M,p_1,\dots,p_M}}[n+i-m] \Bigg) D^m D^{-n} D^{n+i-m} \\
&= \sum_m \sum_{j_1} \dots \sum_{j_M} \mathcal{H}_{m_1,\dots,m_N,j_1,\dots,j_M}[m] D^m \\
&\cdot \sum_n \sum_{p_1} \dots \sum_{p_M} \mathcal{H}^*_{n_1,\dots,n_N,p_1,\dots,p_M}[n] D^{-n} \sum_i \mathcal{R}_{\mathcal{X}_{j_1,\dots,j_M,p_1,\dots,p_M}}[n+i-m] D^{n+i-m} \\
&= \sum_{j_1,\dots,j_M} \breve{\mathcal{H}}_{m_1,\dots,m_N,j_1,\dots,j_M}(D) \sum_{p_1,\dots,p_M} \breve{\mathcal{H}}^*_{n_1,\dots,n_N,p_1,\dots,p_M}(D^{-1}) \breve{\mathcal{S}}_{\mathcal{X}_{j_1,\dots,j_M,p_1,\dots,p_M}}(D)
\end{aligned}
\tag{4.25}
$$

which in tensor notation gives

$$
\breve{\mathcal{S}}_{\mathcal{Y}}(D) = \{\breve{\mathcal{H}}(D), \{\breve{\mathcal{S}}_{\mathcal{X}}(D), \breve{\mathcal{H}}^H(D^{-1})\}_{(M)}\}_{(M)} \tag{4.26}
$$

The cross correlation tensor $\mathcal{R}_{\mathcal{Y},\mathcal{X}}[i]$ has components

$$
\begin{aligned}
\mathcal{R}_{\mathcal{Y},\mathcal{X}_{i_1,\dots,i_N,j_1,\dots,j_M}}(i) &= \mathbb{E}\big[\mathcal{Y}_{i_1,\dots,i_N}[k] \circ \mathcal{X}^*_{j_1,\dots,j_M}[k-i]\big] \\
&= \mathbb{E}\big[\big(\sum_m \sum_{p_1} \dots \sum_{p_M} \mathcal{H}_{i_1,\dots,i_N,p_1,\dots,p_M}[m] \mathcal{X}_{p_1,\dots,p_M}[k-m]\big) \mathcal{X}^*_{j_1,\dots,k_M}[k-i]\big)\big]
\end{aligned}
\tag{4.27}
$$

$$
= \sum_n \sum_{p_1} \dots \sum_{p_M} \mathcal{H}_{i_1,\dots,i_N,p_1,\dots,p_M}[m] \mathcal{R}_{\mathcal{X}_{p_1,\dots,p_M,j_1,\dots,j_M}}[i-m] \tag{4.28}
$$

Taking the D transform we get the Cross spectrum

$$
\breve{\mathcal{S}}_{\mathcal{Y},\mathcal{X}}(D) = \{\breve{\mathcal{H}}(D), \breve{\mathcal{S}}_{\mathcal{X}}(D)\}_{(M)} \tag{4.29}
$$

□

Next we show that

$$
\breve{\mathcal{S}}_{\mathcal{Y},\mathcal{X}}(D) = \breve{\mathcal{S}}^H_{\mathcal{X},\mathcal{Y}}(D^{-1}) \tag{4.30}
$$

Proof. The cross correlation tensor $\mathcal{R}_{\mathcal{X},\mathcal{Y}}[i]$ is

$$
\mathcal{R}_{\mathcal{X},\mathcal{Y}}[i] = \mathbb{E}\big[\mathcal{Y}[k] \circ \mathcal{X}^*[k-i]\big] \tag{4.31}
$$

and has components

$$\begin{aligned}
\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}_{j_1,\dots,j_M,i_1,\dots,i_N}}[i] &= \mathbb{E}\big[\boldsymbol{\mathcal{X}}_{j_1,\dots,j_M}[k]\boldsymbol{\mathcal{Y}}^*_{i_1,\dots,i_N}[k-i]\big] \\
&= \mathbb{E}\big[\boldsymbol{\mathcal{X}}_{j_1,\dots,k_M}[k]\Big(\sum_m\sum_{p_1}\dots\sum_{p_M}\boldsymbol{\mathcal{H}}_{i_1,\dots,i_N,p_1,\dots,p_M}[m]\boldsymbol{\mathcal{X}}_{p_1,\dots,p_M}[k-i-m]\Big)^*\big] \\
&= \mathbb{E}\big[\boldsymbol{\mathcal{X}}_{j_1\dots k_M}[k]\Big(\sum_m\sum_{p_1}\dots\sum_{p_M}\boldsymbol{\mathcal{H}}^*_{i_1,\dots,i_N,p_1,\dots,p_M}[m]\boldsymbol{\mathcal{X}}^*_{p_1,\dots,p_M}[k-i-m]\Big)\big] \\
&= \mathbb{E}\big[\sum_m\sum_{p_1}\dots\sum_{p_M}\boldsymbol{\mathcal{H}}^*_{i_1,\dots,i_N,p_1,\dots,p_M}[m]\boldsymbol{\mathcal{X}}^*_{p_1,\dots,p_M}[k-i-m]\boldsymbol{\mathcal{X}}_{j_1,\dots,k_M}[k]\big] \\
&= \sum_m\sum_{p_1}\dots\sum_{p_M}\boldsymbol{\mathcal{H}}^*_{i_1,\dots,i_N,p_1,\dots,p_M}[m]\,\mathbb{E}\big[\boldsymbol{\mathcal{X}}^*_{p_1,\dots,p_M}[k-i-m]\boldsymbol{\mathcal{X}}_{j_1,\dots,j_M}[k]\big] \\
&= \sum_m\sum_{p_1}\dots\sum_{p_M}\boldsymbol{\mathcal{H}}^*_{i_1,\dots,i_N,p_1,\dots,p_M}[m]\Big(\,\mathbb{E}\big[\boldsymbol{\mathcal{X}}_{p_1,\dots,p_M}[k-i-m]\boldsymbol{\mathcal{X}}^*_{j_1,\dots,j_M}[k]\big]\Big)^* \\
&= \sum_m\sum_{p_1}\dots\sum_{p_M}\boldsymbol{\mathcal{H}}^*_{i_1,\dots,i_N,p_1,\dots,p_M}[m]\boldsymbol{\mathcal{R}}^*_{\boldsymbol{\mathcal{X}}_{p_1,\dots,p_M,j_1,\dots,j_M}}[-i-m] \\
&= \sum_m\sum_{p_1}\dots\sum_{p_M}\boldsymbol{\mathcal{R}}^H_{\boldsymbol{\mathcal{X}}_{j_1,\dots,j_M,p_1,\dots,p_M}}[i+m]\boldsymbol{\mathcal{H}}^H_{p_1,\dots,p_M,i_1,\dots,i_N}[m]
\end{aligned} \tag{4.32}$$

In tensor notation, (4.32) can be written as

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}}[i] = \sum_m\{\boldsymbol{\mathcal{R}}^H_{\boldsymbol{\mathcal{X}}}[i+m],\boldsymbol{\mathcal{H}}^H[m]\}_{(M)} \tag{4.33}$$

taking the D-transform of (4.33) we have

$$\begin{aligned}
\breve{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}}(D) &= \sum_i\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}}[i]D^i \\
&= \sum_i\Big(\sum_m\{\boldsymbol{\mathcal{R}}^H_{\boldsymbol{\mathcal{X}}}[-i-m],\boldsymbol{\mathcal{H}}^H[m]\}_{(M)}\Big)D^i \\
&= \sum_i\Big(\sum_m\{\boldsymbol{\mathcal{R}}^H_{\boldsymbol{\mathcal{X}}}[-i-m],\boldsymbol{\mathcal{H}}^H[m]\}_{(M)}\Big)D^{i+m}D^{-m} \\
&= \sum_m\{\Big(\sum_i\boldsymbol{\mathcal{R}}^H_{\boldsymbol{\mathcal{X}}}[-(i+m)]D^{i+m}\Big),\boldsymbol{\mathcal{H}}^H[m]\}_{(M)}D^{-m} \\
&= \sum_m\{\breve{\boldsymbol{\mathcal{S}}}^H_{\boldsymbol{\mathcal{X}}}(D^{-1}),\Big(\boldsymbol{\mathcal{H}}^H[m]D^{-m}\Big)\}_{(M)} \\
&= \{\breve{\boldsymbol{\mathcal{S}}}^H_{\boldsymbol{\mathcal{X}}}(D^{-1}),\breve{\boldsymbol{\mathcal{H}}}^H(D^{-1})\}_{(M)}
\end{aligned}$$

$$= \left(\{\breve{\mathcal{H}}(D^{-1}), \breve{\mathcal{S}}_{\mathcal{X}}(D^{-1})\}_{(M)} \right)^H$$
$$= \breve{\mathcal{S}}^H_{\mathcal{Y},\mathcal{X}}(D^{-1}) \tag{4.34}$$

$\square$

Definition 4.1.8. *Causality of a Discrete System Tensor* Causality for scalar systems is well known. A scalar system $h[n]$ is said to be causal if the output $y[n]$ depends only on the present and past values, $x[n], x[n-1], \ldots$, of $x[n]$. However, the definition of causality for tensor systems offers more flexibility since there are several components within each tensor. We start with the definition of *loose* causality of a system tensor. A system tensor $\mathcal{H}[n] \in \mathbb{C}_n^{J_1 \times \ldots \times J_M \times I_1 \times \ldots \times I_N}$ is said to be loosely causal if

$$\mathcal{H}[n] = \mathbf{0}_T \quad \text{for } n < 0 \tag{4.35}$$

The D-transform of such a system has the form

$$\breve{\mathcal{H}}(D) = \mathcal{H}[0] + \mathcal{H}[1]D + \mathcal{H}[2]D^2 + \ldots \tag{4.36}$$

Next, for system tensors whose input and output have the same dimensions, we define strict causality. For a matrix system $\mathbf{A}[k] \in \mathbb{C}_k^{N \times N}$ with an input $\mathbf{v}[k] \in \mathbb{C}_k^N$, authors in [56] define the system $\mathbf{A}[k]$ to be causal if the output $\mathbf{w}_i[k] = \sum_n \sum_{j=1}^N \mathbf{A}_{i,j}[n]\mathbf{v}_j[k-n]$ does not depend on 'future' inputs $\mathbf{v}_{i+1}[k], \mathbf{v}_{i+2}[k], \ldots, \mathbf{v}_N[k], \mathbf{v}[k+1], \mathbf{v}[k+2] \ldots$, i.e., $\breve{\mathbf{A}}(D) = \mathbf{A}[0] + \mathbf{A}[1]D + \mathbf{A}[2]D^2 + \ldots$ and $\mathbf{A}[0]$ is a lower triangular matrix. This means that within a vector symbol $\mathbf{v}[k]$, there is a sequencing of the components where $\mathbf{v}_j[k]$ appears before $\mathbf{v}_i[k]$ if $j < i$ that indicates what future means. This is just one possible sequencing of the components of $\mathbf{v}[k]$ and a different sequencing would result in a different structure of $\mathbf{A}[0]$. We extend this definition of causality to tensor systems. Such systems are encountered, for example, in the case of decision feedback equalizers that take into consideration past decisions within the same tensor symbol as well as past decisions of previous tensor symbols, the feedback system is a strictly causal system.

Consider a system tensor $\mathcal{G}[n] \in \mathbb{C}_n^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ with input $\mathcal{X}[n] \in \mathbb{C}_n^{I_1 \times \ldots \times I_N}$ and output $\mathcal{Y}[n] \in \mathbb{C}_n^{I_1 \times \ldots \times I_N}$. Next, define a sequencing of the components of these input and output tensors such that for two components $\mathfrak{X}_{i_1,\ldots,i_N}$ and $\mathfrak{X}_{i'_1,\ldots,i'_N}$, $\mathfrak{X}_{i'_1,\ldots,i'_N}$ is a future component if :

$$\big(i'_1 + \sum_{k=2}^{N}(i'_k - 1)\prod_{l=k-1} I_l\big) > \big(i_1 + \sum_{k=2}^{N}(i_k - 1)\prod_{l=k-1} I_l\big) \tag{4.37}$$

For such a sequencing, the system $\mathfrak{G}[n]$ is said to be strictly causal if a particular component $\mathfrak{Y}_{i_1,\ldots,i_N}[n]$ does not depend on 'future' components as defined in (4.37). In such a case, $\mathfrak{G}[0]$ is a pseudo-lower triangular tensor. Thus, we have

$$\mathfrak{G}[n] = \mathbf{0}_T \quad \text{for } n < 0 \tag{4.38}$$

and

$$\mathfrak{G}_{i'_1,\ldots,i'_N,i_1,\ldots,i_N}[0] \begin{cases} 0 & \text{if } \big(i'_1 + \sum_{k=2}^{N}(i'_k - 1)\prod_{l=k-1} I_l\big) < \big(i_1 + \sum_{k=2}^{N}(i_k - 1)\prod_{l=k-1} I_l\big) \\ g_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} & \text{otherwise} \end{cases} \tag{4.39}$$

where $g_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}$ are arbitrary scalars where at least one of them is non-zero.
As an example, consider a second order input $\mathfrak{X}[n] \in \mathbb{C}^{3\times 3}$ and a system $\mathfrak{G}[n] \in \mathbb{C}_n^{3\times 3\times 3\times 3}$. Then (4.39) becomes

$$\mathfrak{G}_{i'_1,i'_2,i_1,i_2} = \begin{cases} 0 & \text{if } i'_1 + 3(i'_2 - 1) < i_1 + 3(i_2 - 1) \\ g_{i'_1,i'_2,i_1,i_2} & \text{otherwise} \end{cases} \tag{4.40}$$

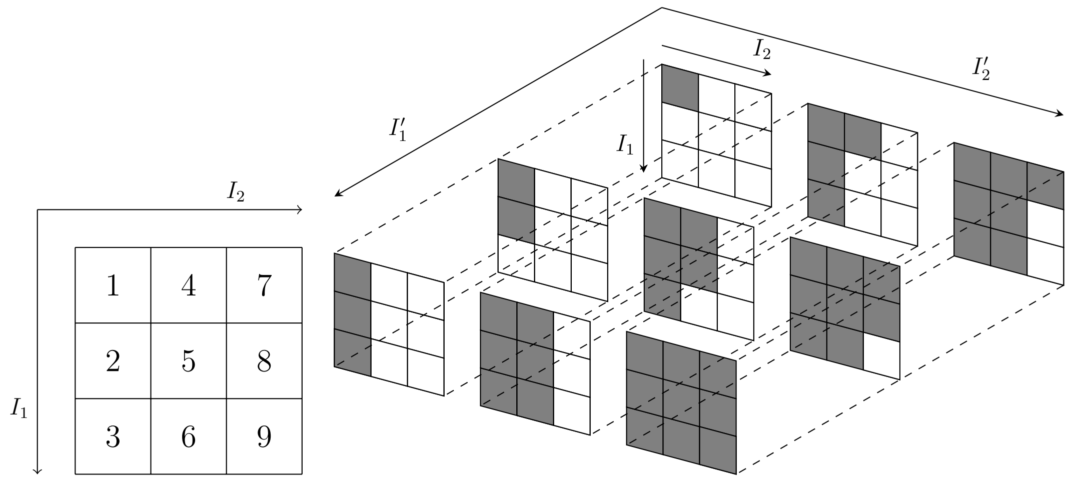
Figure 4.1 shows the ordering of the components of the input and the non-zero components of the system tensor highlighted in gray. It is important to note here that the ordering given here is just *one* of the many possible sequences and the structure of a strictly causal tensor would change accordingly. However, this is the sequencing used in the rest of this thesis.
For any system $\breve{\mathfrak{C}}(D)$, we define its *purely* causal part as

$$\breve{\mathfrak{C}}(D)^+ = \mathfrak{C}[0]^+ + \mathfrak{C}[1]D + \ldots \tag{4.41}$$

and its anti-causal component as

$$\breve{\mathfrak{C}}(D)^- = \breve{\mathfrak{C}}(D) - \breve{\mathfrak{C}}(D)^+ \tag{4.42}$$

where $\mathfrak{C}[0]^-$ is the pseudo-upper triangular part of $\mathfrak{C}[0]$. For the case of matrix systems



(a) a (b) b

Fig. 4.1 (a) Sequencing of input components (b) non zero elements of causal system

(4.38) and (4.39) reduce to the definitions found in [56]. A matrix filter $\breve{\mathbf{G}}(D) \in \mathbb{C}_D^{N \times N}$ that couples a vector input $\breve{\mathbf{x}}(D) \in \mathbb{C}_D^N$ with a vector output $\breve{\mathbf{y}}(D) \in \mathbb{C}_D^N$ through the relation $\breve{\mathbf{y}}(D) = \breve{\mathbf{G}}(D)\breve{\mathbf{x}}(D)$ is causal if it has the form

$$\mathbf{G}(D) = \mathbf{G}[0] + \mathbf{G}[1]D + \mathbf{G}[2]D^2 + \dots \tag{4.43}$$

If, further $\mathbf{G}[0]$ is lower triangular, then the system $\mathbf{G}(D)$ is purely causal.

Both causal and strictly causal system tensors can be represented using a tensor tapped delay line model, the structure of which is shown in Fig. 4.2.

Definition 4.1.9. *Factoring the Spectral Tensor* The factoring of a scalar spectrum $\breve{\mathbf{S}}_x(D)$ such that $\breve{\mathbf{S}}_x(D) = \breve{g}(D)\breve{g}^*(D^{-1})$ with $\breve{g}(D)$ being a stable minimum phase transform is well known in the literature [57]. Authors of [58] and [59] generalize this factorization for the case of the spectrum of a vector. We wish to find a similar factorization for the spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D) \in \mathbb{C}_D^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ of a tensor $\boldsymbol{\mathfrak{X}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$. It is assumed that the spectral tensor $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D)$ is rational and stable. This means that all of its components are rational and stable transfer functions. The problem is to find a factorization

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D) = \{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1})\}_{(N)} \tag{4.44}$$

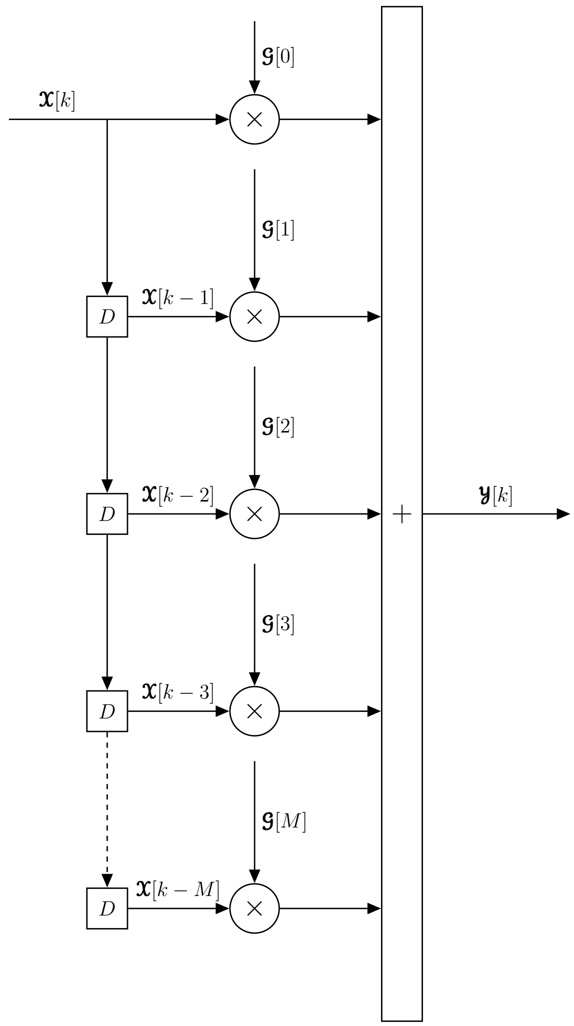


Fig. 4.2 Tensor Tapped Delay Line

where $\breve{\mathfrak{Q}}(D) \in \mathbb{C}_D^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$ is causal and stable and has a stable inverse $\breve{\mathfrak{Q}}^{-1}(D)$. To solve the factorization problem we consider the case when an input tensor $\breve{\mathfrak{P}}(D)$ with spectrum $\breve{\mathfrak{S}}_{\mathfrak{P}}(D) = \mathfrak{I}_N$ excites a system $\breve{\mathfrak{Q}}(D)$. From (4.21) we have that the output, say $\breve{\mathfrak{R}}(D)$, of this system will have a spectrum

$$\breve{\mathfrak{S}}_{\mathfrak{R}}(D) = \{\breve{\mathfrak{Q}}(D), \{\breve{\mathfrak{S}}_{\mathfrak{P}}(D), \breve{\mathfrak{Q}}^H(D^{-1})\}_{(N)}\}_{(N)} = \{\breve{\mathfrak{Q}}(D), \breve{\mathfrak{Q}}^H(D^{-1})\}_{(N)} \tag{4.45}$$

If $\breve{\mathfrak{R}}(D)$ excites a series of cascaded linear systems $\breve{\mathfrak{T}}_1(D), \breve{\mathfrak{T}}_2(D), \ldots \breve{\mathfrak{T}}_L(D)$ such that the output of this overall system, say $\breve{\mathfrak{U}}(D)$, has an identity spectrum $\breve{\mathfrak{S}}_{\mathfrak{U}}(D) = \mathfrak{I}_N$ then this cascade is the inverse of $\breve{\mathfrak{Q}}(D)$. We have

$$\breve{\mathfrak{Q}}^{-1}(D) = \{\{\breve{\mathfrak{T}}_L(D), \breve{\mathfrak{T}}_{L-1}(D)\}_{(N)}, \ldots \breve{\mathfrak{T}}_1(D)\}_{(N)} \tag{4.46}$$

and

$$\breve{\mathfrak{Q}}(D) = \{\{\breve{\mathfrak{T}}_1^{-1}(D), \breve{\mathfrak{T}}_2^{-1}(D)\}_{(N)}, \ldots \breve{\mathfrak{T}}_L^{-1}(D)\}_{(N)} \tag{4.47}$$

The solution of the factorization problem is performed by a series of L tensor transformations on the spectral tensor that converts it into an identity tensor with the constraint that each $\breve{\mathfrak{T}}_i(D)$ and its inverse $\breve{\mathfrak{T}}_i^{-1}(D)$ are stable.

4.2 Linear Equalization

In this section, we investigate some linear tensor equalization architectures for systems with multi-domain interference (both inter-tensor and intra-tensor).

4.2.1 Equivalent Discrete Time Model

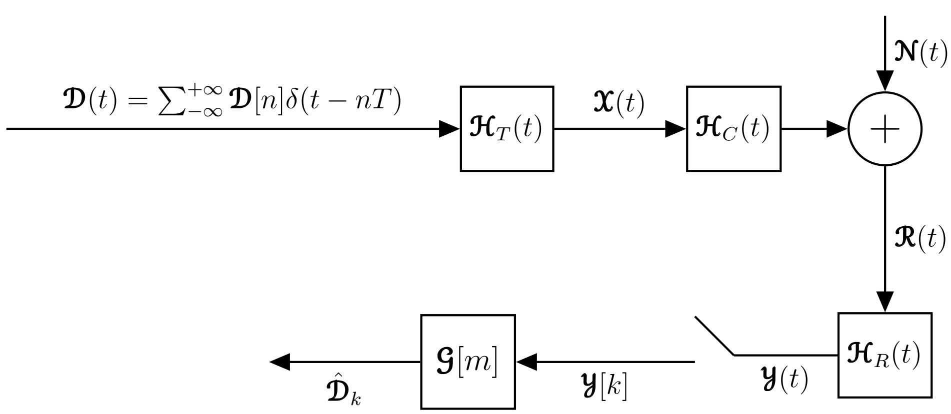


Fig. 4.3 Noisy System Model

Consider the system in Figure 4.3. The data transmitted on the nth tensor symbol is $\mathfrak{D}[n] \in \mathbb{C}_n^{I_1 \times \ldots \times I_N}$. The transmit, channel and receive system tensors are $\mathfrak{H}_T(t) \in \mathbb{C}_t^{J_1 \times \ldots \times J_M \times I_1 \times \ldots \times I_N}$, $\mathfrak{H}_C(t) \in \mathbb{C}_t^{K_1 \times \ldots \times K_O \times J_1 \times \ldots \times J_M}$, $\mathfrak{H}_R(t) \in \mathbb{C}_t^{L_1 \times \ldots \times L_P \times K_1 \times \ldots \times K_O}$ respectively, $\mathfrak{N}(t) \in \mathbb{C}_t^{K_1 \times \ldots \times K_O}$ is an additive noise tensor, and $\mathfrak{X}(t) \in \mathbb{C}_t^{J_1 \times \ldots \times J_M}$ is the output of the transmit system tensor and $\boldsymbol{\mathcal{Y}}(t) \in \mathbb{C}_t^{L_1 \times \ldots \times L_P}$ is the output of the receive system tensor. Furthermore, $\boldsymbol{\mathcal{G}}[n] \in \mathbb{C}_n^{I_1 \times \ldots \times I_N \times L_1 \times \ldots \times L_P}$ is a discrete system tensor whose output $\hat{\boldsymbol{\mathcal{D}}}[n]$ is the estimate of the data $\boldsymbol{\mathcal{D}}[n]$. The input to the receive system tensor is

$$\begin{aligned} \boldsymbol{\mathcal{R}}(t) &= \left\{ \{\boldsymbol{\mathcal{H}}_C(t) * \boldsymbol{\mathcal{H}}_T(t)\}_{(M)} * \sum_n \boldsymbol{\mathcal{D}}[n]\delta(t - nT) \right\}_{(N)} + \boldsymbol{\mathcal{N}}(t) \\ &= \{\boldsymbol{\mathcal{C}}(t - nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)} + \boldsymbol{\mathcal{N}}(t) \end{aligned} \tag{4.48}$$

where the cascade of the transmit tensor and the channel tensor is represented by $\boldsymbol{\mathcal{C}}(t) = \{\boldsymbol{\mathcal{H}}_C(t) * \boldsymbol{\mathcal{H}}_T(t)\}_{(M)} \in \mathbb{C}_t^{K_1 \times \ldots \times K_O \times I_1 \times \ldots I_N}$. The output of the receive system tensor is

$$\begin{aligned} \boldsymbol{\mathcal{Y}}(t) &= \{\boldsymbol{\mathcal{H}}_R(t) * \boldsymbol{\mathcal{R}}(t)\}_{(O)} + \{\boldsymbol{\mathcal{H}}_R(t) * \boldsymbol{\mathcal{N}}(t)\}_{(O)} \\ &= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(t - nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)} + \boldsymbol{\mathcal{V}}(t) \end{aligned} \tag{4.49}$$

where $\boldsymbol{\mathcal{H}}(t) \in \mathbb{C}_t^{L_1 \times \ldots \times L_P \times I_1 \times \ldots \times I_N}$ is the overall system tensor comprising the transmit, channel and receive system tensors, and $\boldsymbol{\mathcal{V}}(t) = \{\boldsymbol{\mathcal{H}}_R(t) * \boldsymbol{\mathcal{N}}(t)\}_{(O)} \in \mathbb{C}_t^{L_1 \times \ldots \times L_P}$. Sampling the output $\boldsymbol{\mathcal{Y}}(t)$ at a rate of $\frac{1}{T}$, we get

$$\begin{aligned} \boldsymbol{\mathcal{Y}}[k] &= \boldsymbol{\mathcal{Y}}(kT) \\ &= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(kT - nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)} + \boldsymbol{\mathcal{V}}[kT] \\ &= \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[k], \boldsymbol{\mathcal{D}}[k - n]\}_{(N)} + \boldsymbol{\mathcal{V}}[k] \end{aligned} \tag{4.50}$$

where $\boldsymbol{\mathcal{H}}[k] = \boldsymbol{\mathcal{H}}(kT)$ is the kth sample of $\boldsymbol{\mathcal{H}}(t)$ and $\boldsymbol{\mathcal{V}}[k] = \{\boldsymbol{\mathcal{H}}_R(t) * \boldsymbol{\mathcal{N}}(t)\}_{(O)}\Big|_{(t=kT)}$ is the sampled noise with autocorrelation

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{V}}}[i] = \mathbb{E}\left[\boldsymbol{\mathcal{V}}[k] \circ \boldsymbol{\mathcal{V}}^*[k - i]\right] \tag{4.51}$$

The D-transform of (4.50) is

$$\breve{\boldsymbol{\mathcal{Y}}}(D) = \{\breve{\boldsymbol{\mathcal{H}}}(D), \breve{\boldsymbol{\mathcal{D}}}(D)\}_{(N)} + \breve{\boldsymbol{\mathcal{V}}}(D) \tag{4.52}$$

The output of $\boldsymbol{\mathcal{G}}[n]$, which is the estimate of the data tensor $\boldsymbol{\mathcal{D}}[k]$, is

$$\hat{\boldsymbol{\mathcal{D}}}[k] = \sum_m \{\boldsymbol{\mathcal{G}}[m], \boldsymbol{\mathcal{Y}}[k - m]\}_{(P)}$$

$$
\begin{aligned}
&= \sum_m \sum_n \{\boldsymbol{\mathcal{G}}[m], \{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{D}}[k-m-n]\}_{(N)}\}_{(P)} + \sum_m \{\boldsymbol{\mathcal{G}}[m], \boldsymbol{\mathcal{V}}[k-m]\}_{(P)} \\
&= \sum_m \sum_n \{\boldsymbol{\mathcal{G}}[m], \{\boldsymbol{\mathcal{H}}[n], \boldsymbol{\mathcal{D}}[k-m-n]\}_{(N)}\}_{(P)} + \tilde{\boldsymbol{\mathcal{V}}}[k] \qquad (4.53)
\end{aligned}
$$

The D-transform of (4.53) is

$$
\begin{aligned}
\hat{\breve{\boldsymbol{\mathcal{D}}}}(D) &= \{\breve{\boldsymbol{\mathcal{G}}}(D), \breve{\boldsymbol{\mathcal{Y}}}(D)\}_{(P)} \\
&= \{\breve{\boldsymbol{\mathcal{G}}}(D), \{\breve{\boldsymbol{\mathcal{H}}}(D), \breve{\boldsymbol{\mathcal{D}}}(D)\}_{(N)}\}_{(P)} + \breve{\tilde{\boldsymbol{\mathcal{V}}}}(D) \qquad (4.54)
\end{aligned}
$$

The system defined by (4.53) is called the equivalent discrete time system model and is shown in Figure 4.4.

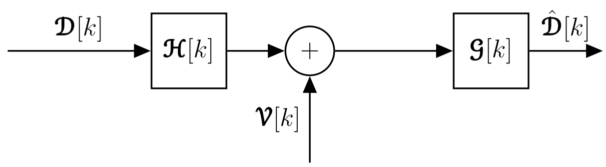


Fig. 4.4 Equivalent Discrete Time System Model

4.2.2 Whitened matched filter

Theorem 2. Consider an input $\boldsymbol{\mathcal{X}}(t) = \sum_{n=-\infty}^{+\infty} \boldsymbol{\mathcal{X}}[n]\delta(t - nT) \in \mathbb{C}_t^{I_1 \times \ldots \times I_N}$ to a system tensor $\boldsymbol{\mathcal{A}}(t) \in \mathbb{C}_t^{J_1 \times \ldots \times J_M \times I_1 \times \ldots \times I_N}$. The output of this filter, corrupted by additive white Gaussian noise, is $\boldsymbol{\mathcal{R}}(t) = \{\boldsymbol{\mathcal{A}}(t) * \boldsymbol{\mathcal{X}}(t)\}_{(N)} + \boldsymbol{\mathcal{N}}(t)$ where $\boldsymbol{\mathcal{N}}(t)$ is a tensor whose components are white Gaussian noise processes. Let $\boldsymbol{\mathcal{R}}(t)$ be the input to a system tensor $\boldsymbol{\mathcal{B}}(t) \in \mathbb{C}_t^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_M}$ with an output $\boldsymbol{\mathcal{Y}}(t) \in \mathbb{C}_t^{I_1 \times \ldots \times I_N}$. The per component SNR of the samples $\boldsymbol{\mathcal{Y}}(kT) = \boldsymbol{\mathcal{Y}}[k]$ is maximized when $\boldsymbol{\mathcal{B}}(t) = \boldsymbol{\mathcal{A}}^H(-t)$.

The proof of this can be found in Appendix A.1. Assuming that the receive filter is a Tensor Matched filter, that is matched to the combined channel and transmit tensors, we have $\boldsymbol{\mathcal{H}}_R(t) = \boldsymbol{\mathcal{C}}^H(-t) \in \mathbb{C}_t^{I_1 \times \ldots \times I_N \times K_1 \times \ldots \times K_O}$. The overall system tensor is $\boldsymbol{\mathcal{H}}(t) = \{\boldsymbol{\mathcal{C}}^H(-t) * \boldsymbol{\mathcal{C}}(t)\}_{(O)} \in \mathbb{C}_t^{I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N}$ which can be written component wise as:

$$
\boldsymbol{\mathcal{H}}_{j_1,\ldots,j_N i_1,\ldots,i_N}(t) = \sum_{k_1} \ldots \sum_{k_O} \boldsymbol{\mathcal{C}}^H_{i_1,\ldots,i_N,k_1,\ldots,k_O}(-t) * \boldsymbol{\mathcal{C}}_{k_1,\ldots,k_O,j_1,\ldots,j_N}(t)
$$

$$= \sum_{k_1} \cdots \sum_{k_O} \int_{-\infty}^{+\infty} \mathbfcal{C}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(\tau - t)\mathbfcal{C}_{k_1,\ldots,k_O,j_1,\ldots,j_N}(\tau)d\tau' \tag{4.55}$$

For the noise tensor $\mathbfcal{V}(t)$, denote its continuous auto-correlation tensor as $\Phi_{\mathbfcal{V}}(p) = \mathbb{E}\big[\mathbfcal{V}(t) \circ \mathbfcal{V}^*(t-p)\big]$ with components

$$\begin{aligned}
\Phi_{\mathbfcal{V}_{i_1,\ldots,i_N,j_1,\ldots,j_N}}(p) = \mathbb{E}\Bigg[&\sum_{k_1} \cdots \sum_{k_O} \int_{-\infty}^{+\infty} \mathbfcal{C}^H_{i_1,\ldots,i_N,k_1,\ldots,k_O}(-\tau)\mathbfcal{N}_{k_1,\ldots,k_O}(t-\tau)d\tau \\
&\cdot \sum_{k'_1} \cdots \sum_{k'_O} \int_{-\infty}^{+\infty} \mathbfcal{C}^H_{j_1,\ldots,j_N,k'_1,\ldots,k'_O}(-\tau')\mathbfcal{N}_{k'_1,\ldots,k'_O}(t-p-\tau')d\tau' \Bigg] \\
= &\sum_{k_1} \cdots \sum_{k_O} \sum_{k'_1} \cdots \sum_{k'_O} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbfcal{C}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(-\tau)\mathbfcal{C}_{k'_1,\ldots,k'_O,j_1,\ldots,j_N}(-\tau') \\
&\cdot \mathbb{E}\big[\mathbfcal{N}_{k_1,\ldots,k_O}(t-\tau)\mathbfcal{N}_{k'_1,\ldots,k'_O}(t-p-\tau')\big] d\tau d\tau' \\
= &\sum_{k_1} \cdots \sum_{k_O} \sum_{k'_1} \cdots \sum_{k'_O} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbfcal{C}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(-\tau)\mathbfcal{C}_{k'_1,\ldots,k'_O,j_1,\ldots,j_N}(-\tau') \\
&\cdot \Phi_{\mathbfcal{N}_{k_1,\ldots,k_O,k'_1,\ldots,k'_O}}(p+\tau'-\tau)d\tau d\tau'
\end{aligned} \tag{4.56}$$

When the components of the noise tensor $\mathbfcal{N}(t)$ are white uncorrelated Gaussian random processes with double sided spectral density N_0 we get

$$\Phi_{\mathbfcal{N}_{k_1,\ldots,k_O,k'_1,\ldots,k'_O}}(p+\tau'-\tau) = \begin{cases} N_0\delta(p+\tau'-\tau) & \text{if } k_1 = k'_1, k_2 = k'_2, \ldots, k_O = k'_O \\ 0 & \text{otherwise} \end{cases} \tag{4.57}$$

Using (4.57) in (4.56) gives

$$\begin{aligned}
\Phi_{\mathbfcal{V}_{i_1,\ldots,i_N,j_1,\ldots,j_N}}(p) &= \sum_{k_1,\ldots,k_O} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbfcal{C}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(-\tau)\mathbfcal{C}_{k_1,\ldots,k_O,j_1,\ldots,j_N}(-\tau')N_0\delta(p+\tau'-\tau)d\tau d\tau' \\
&= N_0 \sum_{k_1,\ldots,k_O} \int_{-\infty}^{+\infty} \mathbfcal{C}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(-\tau) \int_{-\infty}^{+\infty} \mathbfcal{C}_{k_1,\ldots,k_O,j_1,\ldots,j_N}(-\tau')\delta(p+\tau'-\tau)d\tau d\tau' \\
&= N_0 \sum_{k_1,\ldots,k_O} \int_{-\infty}^{+\infty} \mathbfcal{C}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(-\tau)\mathbfcal{C}_{k_1,\ldots k_O,j_1,\ldots,j_N}(-(\tau-p))d\tau \\
&= N_0 \sum_{k_1,\ldots,k_O} \int_{-\infty}^{+\infty} \mathbfcal{C}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(-\tau)\mathbfcal{C}_{k_1,\ldots,k_O,j_1,\ldots,j_N}(p-\tau)d\tau
\end{aligned}$$

Making the substitution $p - \tau = k$ gives

$$\begin{aligned}\Phi_{\boldsymbol{\mathcal{V}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}}(p) &= N_0 \sum_{k_1,\ldots,k_O} - \int_{+\infty}^{-\infty} \boldsymbol{\mathfrak{C}}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(k-p)\boldsymbol{\mathfrak{C}}_{k_1,\ldots,k_O,j_1,\ldots,j_N}(k)dk \\ &= N_0 \sum_{k_1,\ldots,k_O} \int_{-\infty}^{+\infty} \boldsymbol{\mathfrak{C}}^*_{k_1,\ldots,k_O,i_1,\ldots,i_N}(k-p)\boldsymbol{\mathfrak{C}}_{k_1,\ldots,k_O,j_1,\ldots,j_N}(k)dk \qquad (4.58)\end{aligned}$$

Comparing (4.58) with (4.55), we can see that $\Phi_{\boldsymbol{\mathcal{V}}}(p) = N_0 \boldsymbol{\mathfrak{H}}(p)$. Sampling $\Phi_{\boldsymbol{\mathcal{V}}}(p)$ and taking its D-transform gives us the noise spectral tensor

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathcal{V}}}(D) = N_0 \breve{\boldsymbol{\mathfrak{H}}}(D) \qquad (4.59)$$

The spectrum of the noise may be factorized into factor tensors as

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathcal{V}}}(D) = N_0 \{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{Q}}}^H(D)^{-1}\}_{(N)} \qquad (4.60)$$

and using (4.60) in (4.59) gives

$$\breve{\boldsymbol{\mathfrak{H}}}(D) = \{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1})\}_{(N)} \qquad (4.61)$$

An example of such a factorization is provided in Appendix 3.1. (4.52) may now be re-written as

$$\breve{\boldsymbol{\mathfrak{Y}}}(D) = \{\{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1})\}_{(N)}, \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)} + \{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{N}}}(D)\}_{(N)} \qquad (4.62)$$

The output noise in (4.62) is coloured and is whitened by passing $\breve{\boldsymbol{\mathfrak{Y}}}(D)$ through a system $\breve{\boldsymbol{\mathfrak{Q}}}^{-1}(D)$ whose output is

$$\begin{aligned}\breve{\boldsymbol{\mathfrak{Z}}}(D) &= \{\breve{\boldsymbol{\mathfrak{Q}}}^{-1}(D), \breve{\boldsymbol{\mathfrak{Y}}}(D)\}_{(N)} \qquad (4.63)\\ &= \{\breve{\boldsymbol{\mathfrak{Q}}}^{-1}(D), \{\{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1})\}_{(N)}, \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)} + \{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{N}}}(D)\}_{(N)}\}_{(N)} \qquad (4.64)\\ &= \{\breve{\boldsymbol{\mathfrak{Q}}}^{-1}(D), \{\{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1})\}_{(N)}, \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)}\}_{(N)} + \{\breve{\boldsymbol{\mathfrak{Q}}}^{-1}(D), \{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{N}}}(D)\}_{(N)}\}_{(N)}\end{aligned}$$

Using the associativity property (2.8) we get

$$\begin{aligned}&= \{\{\{\breve{\boldsymbol{\mathfrak{Q}}}^{-1}(D), \breve{\boldsymbol{\mathfrak{Q}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1})\}_{(N)}, \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)} + \{\{\breve{\boldsymbol{\mathfrak{Q}}}^{-1}(D), \breve{\boldsymbol{\mathfrak{Q}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathfrak{N}}}(D)\}_{(N)}\\ &= \{\breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1}), \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)} + \breve{\boldsymbol{\mathfrak{N}}}(D) \qquad (4.65)\end{aligned}$$

From Theorem 2, we know that the SNR at the output of the matched filter is maximized. However, the matched filter correlates the noise. This correlation is removed by the noise whitening filter. Define $\boldsymbol{\mathfrak{F}}[k]$ as a sequence of tap co-efficients whose D-transform is $\breve{\boldsymbol{\mathfrak{F}}}(D) =$

$\breve{\mathfrak{Q}}^H(D^{-1})$. The overall input-output relation of a system comprising of the transmit system tensor, the multi-domain channel, the tensor matched filter and the noise whitening tensor tapped delay line is then

$$\mathfrak{Z}[k] = \sum_n \{\mathfrak{F}[n], \mathfrak{D}[k-n]\}_{(N)} + \mathfrak{N}[k] \tag{4.66}$$

4.2.3 Peak Distortion Criterion

Let the cascade of the the overall system $\mathfrak{H}[k]$ and the equalizer $\mathfrak{G}[k]$ from (4.53) be denoted by $\mathfrak{P}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$. We have

$$\breve{\mathfrak{P}}(D) = \{\breve{\mathfrak{G}}(D), \breve{\mathfrak{H}}(D)\}_{(P)} \tag{4.67}$$

with tap co-efficients

$$\mathfrak{P}[k] = \sum_m \{\mathfrak{G}[m], \mathfrak{H}[k-m]\}_{(P)} \tag{4.68}$$

Rewriting (4.53) using (4.68) we get

$$\hat{\mathfrak{D}}[k] = \sum_m \sum_n \{\mathfrak{G}[m], \{\mathfrak{H}[k-m-n], \mathfrak{D}[n]\}_{(N)}\}_{(P)} + \tilde{\mathfrak{V}}[k]$$

Using the associativity property gives

$$\begin{aligned}
\hat{\mathfrak{D}}[k] &= \sum_n \left\{ \left\{ \sum_m \mathfrak{G}[m], \mathfrak{H}[k-m-n] \right\}_{(P)}, \mathfrak{D}[n] \right\}_{(N)} + \tilde{\mathfrak{V}}[k] \\
&= \sum_n \{\mathfrak{P}[k-n], \mathfrak{D}[n]\}_{(N)} + \sum_m \{\mathfrak{G}[m], \mathfrak{V}[k-m]\}_{(P)} \\
&= \sum_n \{\mathfrak{P}[n], \mathfrak{D}[k-n]\}_{(N)} + \sum_m \{\mathfrak{G}[m], \mathfrak{V}[k-m]\}_{(P)}
\end{aligned} \tag{4.69}$$

which on expansion gives

$$\hat{\mathfrak{D}}[k] = \{\mathfrak{P}[0], \mathfrak{D}[k]\}_{(N)} + \sum_{n, n\neq 0} \{\mathfrak{P}[n], \mathfrak{D}[k-n]\}_{(N)} + \sum_m \{\mathfrak{G}[m], \mathfrak{V}[k-m]\}_{(P)} \tag{4.70}$$

whose components are

$$\hat{\mathfrak{D}}_{i_1,\ldots,i_N}[k] = \mathfrak{P}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[0]\mathfrak{D}_{i_1,\ldots,i_N}[k] + \sum_{\substack{j_1,\ldots,j_N \\ (j_1\ldots j_N)\neq(i_1\ldots i_N)}} \mathfrak{P}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[0]\mathfrak{D}_{j_1,\ldots,j_N}[k]$$

$$+\sum_{n,n\neq 0}\sum_{j_1,\ldots,j_N}\mathcal{P}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[n]\mathcal{D}_{j_1,\ldots,j_N}[k-n]+\sum_{m}\sum_{l_1,\ldots,l_P}\mathcal{G}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[k-m]\mathcal{V}_{l_1,\ldots,l_P}[m] \tag{4.71}$$

The first term on the RHS of (4.71) is a scaled version of the required data symbol, the second term is interference from other components within the same tensor (intra-tensor interference), the third term is the interference from other tensor symbols (inter-tensor interference) and the fourth term is the filtered noise. The inter-tensor interference and intra-tensor interference combined are dubbed multi-domain interference (MDI). For an output $\hat{\mathcal{D}}_{i_1\ldots i_N}[k]$, we define the worst case or maximal possible value of the amplitude of the combined multi-domain interference relative to the magnitude of the desired signal sample as the *peak distortion* at $\hat{\mathcal{D}}_{i_1,\ldots,i_N}$. The peak distortion at $i_1,\ldots,i_N$ is then denoted by

$$I_{i_1,\ldots,i_N}=\max\left(\left|\sum_{\substack{j_1,\ldots,j_N\\(j_1,\ldots,j_N)\neq\\(i_1,\ldots,i_N)}}\mathcal{P}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[0]\mathcal{D}_{j_1,\ldots,j_N}[k]+\sum_{\substack{n,n\neq 0\\j_1,\ldots,j_N}}\mathcal{P}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[n]\mathcal{D}_{j_1,\ldots,j_N}[k-n]\right|\right) \tag{4.72}$$

The overall worst-case distortion is the maximum value of $I_{i_1,\ldots,i_N}$ over all the outputs $i_1,\ldots,i_N$ and is denoted by

$$I_0=\max_{i_1,\ldots,i_N} I_{i_1,\ldots,i_N} \tag{4.73}$$

For $z_1,\ldots,z_P\in\mathbb{C}$ we have $|\sum_i z_i|\leq\sum_i|z_i|$ with equality only when $z_1,\ldots,z_N$ have the same argument. When the arguments are different, the difference between the two sides of the inequality reduces as the largest difference in argument reduces. In general, the peak value of the distortion is achieved when for a given $\mathcal{P}[k]$, the components of $\mathcal{D}[k]$ are such that the difference in argument between any two terms in (4.72) is minimum. If the argument of the components are the same, then we have

$$I_{i_1,\ldots,i_N}=\max\left(\sum_{\substack{j_1,\ldots,j_N\\(j_1,\ldots,j_N)\neq\\(i_1,\ldots,i_N)}}\left|\mathcal{P}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[0]\mathcal{D}_{j_1,\ldots,j_N}[k]\right|+\sum_{\substack{n,n\neq 0\\j_1,\ldots,j_N}}\left|\mathcal{P}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[n]\mathcal{D}_{j_1,\ldots,j_N}[k-n]\right|\right)$$

Let D_{max} denote the maximum value that the modulus of a component of $\mathcal{D}[k]$ can take.

Then the peak value of the distortion occurs when $|\boldsymbol{\mathfrak{D}}_{j_1,\ldots,j_N}[k]| = D_{max}$. For such a case we have

$$I_{i_1,\ldots,i_N} = D_{max}\Bigg(\sum_{\substack{j_1,\ldots,j_N \\ (j_1,\ldots,j_N)\neq \\ (i_1,\ldots,i_N)}} \Big|\boldsymbol{\mathfrak{P}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[0]\Big| + \sum_{\substack{n,n\neq 0 \\ j_1,\ldots,j_N}} \Big|\boldsymbol{\mathfrak{P}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[n]\Big|\Bigg) \tag{4.74}$$

For the scalar case where the channel and equalizer are represented by $f[n]$ and $c[n]$ respectively, (4.73) reduces to

$$I_0 = \max(|\sum_{n,n\neq 0} q[n]d[k-n]|) \tag{4.75}$$

where $q[n] = \sum_{m=-\infty}^{+\infty} c[n]f[k-n]$ denotes the convolution of the channel and the equalizer. If $q[n]$ and $d[n]$ are real, and the maximum value of $|d[k]|$ is denoted by d_{max}, the maximum value of (4.75) occurs when all the $|d[k-n]| = d_{max}$ and the algebraic sign of $d[k-n]$ is the same as $q[n]$. In this case, (4.75) becomes

$$I_0 = d_{max} \sum_{n,n\neq 0} |q[n]| \tag{4.76}$$

which is consistent with the definition of peak distortion of [60]. The aim of the first linear equalizer is to minimize I_o and this is called the Peak Distortion or Zero Forcing (ZF) Criterion. From our previous discussions of the generalized Nyquist criterion if the overall system follows the strict Generalized Nyquist criterion then $\boldsymbol{\mathfrak{P}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}[k] = 0$ for $k \neq 0$ and $(i_1,\ldots,i_N) \neq (j_1,\ldots,j_N)$ and hence $I_0 = 0$. Denote the equalizer with the optimal tap coefficient tensors such that the worst case distortion is minimized by $\boldsymbol{\mathfrak{G}}^{ZF}[k]$. From (2.68) we have that

$$\breve{\boldsymbol{\mathfrak{P}}}(D) = \{\breve{\boldsymbol{\mathfrak{G}}}^{ZF}(D), \breve{\boldsymbol{\mathfrak{H}}}(D)\}_{(P)} = \boldsymbol{\mathfrak{I}}_N \tag{4.77}$$

or

$$\breve{\boldsymbol{\mathfrak{G}}}^{ZF}(D) = \breve{\boldsymbol{\mathfrak{H}}}^{+}(D) \tag{4.78}$$

For the specific case of zero inter-tensor interference, we have $\boldsymbol{\mathfrak{H}}[k] = \mathbf{0}_T$ for $k \neq 0$ and $\breve{\boldsymbol{\mathfrak{H}}}(D) = \boldsymbol{\mathfrak{H}}[0]$. Hence, (4.78) becomes

$$\breve{\boldsymbol{\mathfrak{G}}}^{ZF}(D) = \breve{\boldsymbol{\mathfrak{H}}}^{+}(D) = \boldsymbol{\mathfrak{H}}^{+}[0] \tag{4.79}$$

If the matched filter system is used then, using (4.66), we see that the input to the equalizer is $\boldsymbol{\mathcal{Z}}[k]$ such that

$$\hat{\boldsymbol{\mathcal{D}}}[k] = \sum_m \sum_n \{\boldsymbol{\mathcal{G}}[m], \{\boldsymbol{\mathcal{F}}[k-m-n], \boldsymbol{\mathcal{D}}[n]\}_{(N)}\}_{(N)} + \sum_n \{\boldsymbol{\mathcal{G}}[n], \boldsymbol{\mathcal{N}}[k-n]\}_{(N)} \tag{4.80}$$

Denoting the cascade of the equalizer $\boldsymbol{\mathcal{G}}[k]$ and the system $\boldsymbol{\mathcal{F}}[k]$ by $\boldsymbol{\mathcal{R}}[k] = \sum_n \{\boldsymbol{\mathcal{G}}[k], \boldsymbol{\mathcal{F}}[k-n]\}_{(N)}$, and using the associativity property, (4.80) becomes

$$\begin{aligned}\hat{\boldsymbol{\mathcal{D}}}[k] &= \sum_n \left\{ \left\{ \sum_m \boldsymbol{\mathcal{G}}[m], \boldsymbol{\mathcal{F}}[k-m-n] \right\}_{(N)}, \boldsymbol{\mathcal{D}}[n] \right\}_{(N)} + \sum_m \{\boldsymbol{\mathcal{G}}[m], \boldsymbol{\mathcal{N}}[k-m]\}_{(N)} \\ &= \sum_n \{\boldsymbol{\mathcal{R}}[k-n], \boldsymbol{\mathcal{D}}[n]\}_{(N)} + \sum_m \{\boldsymbol{\mathcal{G}}[m], \boldsymbol{\mathcal{N}}[k-m]\}_{(N)} \end{aligned} \tag{4.81}$$

The peak distortion from (4.72) becomes

$$I_{i_1,\dots,i_N} = \max \left(\left| \sum_{\substack{j_1,\dots,j_N \\ (j_1,\dots,j_N)\neq \\ (i_1,\dots,i_N)}} \boldsymbol{\mathcal{R}}_{i_1,\dots,i_N,j_1,\dots,j_N}[0] \boldsymbol{\mathcal{D}}_{j_1,\dots,j_N}[k] + \sum_{\substack{n, n\neq 0 \\ j_1,\dots,j_N}} \boldsymbol{\mathcal{R}}_{i_1,\dots,i_N,j_1,\dots,j_N}[n] \boldsymbol{\mathcal{D}}_{j_1,\dots,j_N}[k-n] \right| \right) \tag{4.82}$$

If the overall system follows the strict Generalized Nyquist criterion then $\boldsymbol{\mathcal{R}}_{i_1,\dots,i_N,j_1,\dots,j_N}[k] = 0$ for $k \neq 0$ and $(i_1,\dots,i_N) \neq (j_1,\dots,j_N)$ and hence $I_0 = 0$. This means that to minimize the peak distortion we require that the overall system $\breve{\boldsymbol{\mathcal{R}}}(D) = \boldsymbol{\mathcal{I}}_N$. Hence we have

$$\breve{\boldsymbol{\mathcal{R}}}(D) = \{\breve{\boldsymbol{\mathcal{G}}}(D), \breve{\boldsymbol{\mathcal{F}}}(D)\}_{(N)} = \boldsymbol{\mathcal{I}}_N \tag{4.83}$$

and the optimal zero forcing equalizer is

$$\breve{\boldsymbol{\mathcal{G}}}^{ZF}(D) = \breve{\boldsymbol{\mathcal{F}}}^{+}(D) \tag{4.84}$$

4.2.4 Tensor Minimum Mean Square Error Equalization

The next equalization scheme that we look at is the equalizer that minimizes the mean squared error between the data tensor $\boldsymbol{\mathcal{D}}[k]$ and its estimate $\hat{\boldsymbol{\mathcal{D}}}[k]$. Define an error tensor $\boldsymbol{\mathcal{E}}[k] = \boldsymbol{\mathcal{D}}[k] - \hat{\boldsymbol{\mathcal{D}}}[k]$. The mean squared error is defined as

$$\mathrm{MSE} = \sum_{i_1,\dots,i_N} \mathbb{E}\left[\left| \boldsymbol{\mathcal{E}}_{i_1,\dots,i_N}[k] \right|^2 \right]$$

$$= \sum_{i_1,\ldots,i_N} \mathbb{E}\left[\left|\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] - \hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\right|^2\right] \tag{4.85}$$

$$= \mathbb{E}\Big[\Big\|\boldsymbol{\mathcal{D}}[k] - \hat{\boldsymbol{\mathcal{D}}}[k]\Big\|_F^2\Big] \tag{4.86}$$

The auto-correlation tensor of the error at zero delay is $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}}[0] = \mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{E}}^*[k]\right]$. For simplicity, we remove the index and denote this by $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}}$. The components of $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}}$ are

$$\begin{aligned}
\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}} &= \mathbb{E}\left[\boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N}[k]\boldsymbol{\mathcal{E}}^*_{j_1,\ldots,j_N}[k]\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] - \hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\right)\left(\boldsymbol{\mathcal{D}}_{j_1,\ldots,j_N}[k] - \hat{\boldsymbol{\mathcal{D}}}_{j_1,\ldots,j_N}[k]\right)^*\right]
\end{aligned} \tag{4.87}$$

When $i_1 = j_1, \ldots, i_N = j_N$, (4.87) becomes

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}} = \mathbb{E}\left[\left|\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] - \hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\right|^2\right] \tag{4.88}$$

Using (4.88) in (4.85) gives

$$\text{MSE} = \sum_{i_1,\ldots,i_N} \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}} = \text{trace}(\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}}) \tag{4.89}$$

We first look at the case of an equalizer with an infinite number of tensor taps, and then move on to restricting the number of taps to M. To find the optimal tap co-efficients we begin with a generalization of the well-known principle of linear estimation for scalar systems [57] that the error must be uncorrelated with all the observed random variables for the MSE to be minimized. Authors in [61] mention the orthogonality principle for MIMO systems using matrices.

Theorem 3. The mean squared error between a tensor $\boldsymbol{\mathcal{D}}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N}$ and its estimate $\hat{\boldsymbol{\mathcal{D}}} \in \mathbb{C}_k^{I_1\times\ldots\times I_N}$ is minimized if and only if the error is uncorrelated with all the observed tensors $\boldsymbol{\mathcal{Y}}[k] \in \mathbb{C}_k^{L_1\times\ldots\times L_M}$

$$\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{Y}}^*[k-i]\right] = \mathbf{0}_T \quad \forall i \tag{4.90}$$

where $\boldsymbol{\mathcal{E}}[k] = \boldsymbol{\mathcal{D}}[k] - \hat{\boldsymbol{\mathcal{D}}}[k]$.

The proof of this theorem can be found in appendix A.2. Using Theorem 3, we have that the optimal multi-linear equalizer must satisfy

$$\begin{aligned}
\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}},\boldsymbol{\mathcal{Y}}}[i] &= \mathbb{E}\big[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{Y}}^*[k-i]\big] \\
&= \mathbb{E}\big[(\boldsymbol{\mathcal{D}}[k] - \hat{\boldsymbol{\mathcal{D}}}[k]) \circ \boldsymbol{\mathcal{Y}}^*[k-i]\big] = \mathbf{0}_T
\end{aligned} \tag{4.91}$$

Component wise, (4.91) becomes

$$\mathcal{R}_{\mathcal{E},\mathcal{Y}_{i_1,\ldots,i_N,l_1,\ldots,l_P}}[i] = \mathbb{E}\big[(\mathcal{D}_{i_1,\ldots,i_N}[k] - \hat{\mathcal{D}}_{i_1,\ldots,i_N}[k])\mathcal{Y}^*_{l_1,\ldots,l_P}[k-i]\big] = 0 \tag{4.92}$$

which implies

$$\begin{aligned}\mathbb{E}\big[\mathcal{D}_{i_1,\ldots,i_N}[k]\mathcal{Y}^*_{l_1,\ldots,l_P}[k-i]\big] &= \mathbb{E}\big[\hat{\mathcal{D}}_{i_1,\ldots,i_N}[k]\mathcal{Y}^*_{l_1,\ldots,l_P}[k-i]\big]\\ \mathcal{R}_{\mathcal{D},\mathcal{Y}_{i_1,\ldots,i_N,l_1,\ldots,l_P}}[i] &= \mathcal{R}_{\hat{\mathcal{D}},\mathcal{Y}_{i_1,\ldots,i_N,l_1,\ldots,l_P}}[i]\\ \breve{\mathcal{S}}_{\mathcal{D},\mathcal{Y}_{i_1,\ldots,i_N,l_1,\ldots,l_P}}(D) &= \breve{\mathcal{S}}_{\hat{\mathcal{D}},\mathcal{Y}_{i_1,\ldots,i_N,l_1,\ldots,l_P}}(D)\end{aligned} \tag{4.93}$$

which in tensor notation gives

$$\breve{\mathcal{S}}_{\mathcal{D},\mathcal{Y}}(D) = \breve{\mathcal{S}}_{\hat{\mathcal{D}},\mathcal{Y}}(D) \tag{4.94}$$

From (4.21), (4.22) and (4.30) we get the following relations:

$$\breve{\mathcal{S}}_{\hat{\mathcal{D}},\mathcal{Y}}(D) = \{\breve{\mathcal{G}}(D), \breve{\mathcal{S}}_{\mathcal{Y}}(D)\}_{(P)} \tag{4.95}$$

$$\begin{aligned}\breve{\mathcal{S}}_{\mathcal{D},\mathcal{Y}}(D) &= \breve{\mathcal{S}}^{\mathrm{H}}_{\mathcal{Y},\mathcal{D}}(D^{-1})\\ &= \{\breve{\mathcal{S}}_{\mathcal{D}}(D^{-1}), \breve{\mathcal{H}}^{H}(D^{-1})\}_{(N)}\end{aligned} \tag{4.96}$$

$$\breve{\mathcal{S}}_{\mathcal{Y}}(D) = \{\breve{\mathcal{H}}(D), \{\breve{\mathcal{S}}_{\mathcal{D}}(D), \breve{\mathcal{H}}^{H}(D^{-1})\}_{(N)}\}_{(N)} + \breve{\mathcal{S}}_{\mathcal{V}}(D) \tag{4.97}$$

where $\breve{\mathcal{S}}_{\mathcal{V}}(D)$ is the spectrum of the noise. Using (4.97),(4.96) and (4.95) in (4.94) we get

$$\begin{aligned}\{\breve{\mathcal{S}}_{\mathcal{D}}(D^{-1}), \breve{\mathcal{H}}^{H}(D^{-1})\}_{(N)} &= \{\{\breve{\mathcal{G}}(D), \{\breve{\mathcal{S}}_{\mathcal{Y}}(D)\}_{(P)}\\ &= \left\{\breve{\mathcal{G}}(D), \left(\big\{\breve{\mathcal{H}}(D), \{\breve{\mathcal{S}}_{\mathcal{D}}(D), \breve{\mathcal{H}}^{H}(D^{-1})\}_{(N)}\big\}_{(N)} + \breve{\mathcal{S}}_{\mathcal{V}}(D)\right)\right\}_{(P)}\end{aligned} \tag{4.98}$$

which gives the optimal tap co-efficients for the infinite tap tensor MMSE equalizer $\breve{\mathcal{G}}^{\mathrm{MMSE}}(D)$ as

$$\breve{\mathcal{G}}^{\mathrm{MMSE}}(D) = \left\{\breve{\mathcal{S}}_{\mathcal{D}}(D^{-1}), \left\{\breve{\mathcal{H}}^{H}(D^{-1}), \left(\big\{\breve{\mathcal{H}}(D), \{\breve{\mathcal{S}}_{\mathcal{D}}(D), \breve{\mathcal{H}}^{H}(D^{-1})\}_{(N)}\big\}_{(N)} + \breve{\mathcal{S}}_{\mathcal{V}}(D)\right)^{-1}\right\}_{(P)}\right\}_{(N)} \tag{4.99}$$

If we use a tensor matched filter at the receiver along with noise whitening, then the input-output relation is described by (4.66) and the optimal tap coefficients for this case reduce to

$$\breve{\mathfrak{G}}^{\mathrm{MMSE}}(D) = \left\{ \breve{\mathfrak{S}}_{\mathfrak{D}}(D^{-1}), \left\{ \breve{\mathfrak{F}}^{H}(D^{-1}), \left(\{ \breve{\mathfrak{F}}(D), \{ \breve{\mathfrak{S}}_{\mathfrak{D}}(D), \breve{\mathfrak{F}}^{H}(D^{-1}) \}_{(N)} \}_{(N)} + N_0 \mathfrak{J}_N \right)^{-1} \right\}_{(N)} \right\}_{(N)} \tag{4.100}$$

For a scalar system where the input is represented by $d[k]$, the overall channel is represented by $h[k]$ and the filtered noise is represented by $v[k]$, (4.99) becomes

$$\breve{g}^{\mathrm{MMSE}}(D) = \frac{\breve{S}_d(D^{-1})\breve{h}^*(D^{-1})}{\breve{h}(D)\breve{S}_d(D)\breve{h}^*(D^{-1}) + \breve{S}_v(D)} \tag{4.101}$$

The definition of the linear MMSE equalizer of (4.101), is consistent with the definition found in [60]. Let $\breve{h}(D)$ consist of a channel and a matched filter. Then equation (4.61) becomes

$$\breve{h}(D) = \breve{q}(D)\breve{q}^*(D^{-1}) \tag{4.102}$$

and the spectrum of $\breve{v}(D)$ is

$$\breve{S}_v(D) = N_0\breve{h}(D) \tag{4.103}$$

Assuming that $d[k]$ is uncorrelated such that $R_d[0] = \mathbb{E}[d[k]d^*[k]]$ and $R_d[i] = 0$ for $i \neq 0$, where $R_d[i]$ denotes the auto-correlation of $d[k]$. Denoting $R_d[0]$ by R_d, (4.101) becomes

$$\breve{g}^{\mathrm{MMSE}}(D) = \frac{R_d\breve{h}^*(D^{-1})}{\breve{h}(D)R_d\breve{h}^*(D^{-1}) + N_0\breve{h}(D)} \tag{4.104}$$

From (4.102) we can see that $\breve{h}(D) = \breve{h}^*(D^{-1})$. This gives

$$\breve{g}^{\mathrm{MMSE}}(D) = \frac{R_d\breve{h}^*(D^{-1})}{\breve{h}^*(D^{-1})(\breve{h}(D)R_d + N_0)} = \frac{R_d}{(\breve{h}(D)R_d + N_0)} = \frac{1}{(\breve{h}(D) + \frac{N_0}{R_d})} \tag{4.105}$$

We can see that (4.105) is consistent with [60]. Further, by using the substitution $D = z^{-1}$ (4.105) becomes

$$\breve{g}^{\mathrm{MMSE}}(z^{-1}) = \frac{1}{(\breve{h}(z^{-1}) + \frac{N_0}{R_d})} = \frac{1}{(\breve{h}(z) + \frac{N_0}{R_d})} \tag{4.106}$$

which is consistent with [57].

If we further assume that components of the input tensor $\mathfrak{D}[k]$ are i.i.d and uncorrelated with unit energy such that $\breve{\mathcal{S}}_{\mathcal{D}}(D) = \mathcal{I}_N$ from (4.100) we have

$$\breve{\mathcal{G}}(D) = \left\{ \breve{\mathcal{F}}^H(D^{-1}), \left(\{\breve{\mathcal{F}}(D), \breve{\mathcal{F}}^H(D^{-1})\}_{(N)} + N_0 \mathcal{I}_N \right)^{-1} \right\}_{(N)} \tag{4.107}$$

As for the zero forcing equalizer, we look at the specific case of zero inter-tensor interference. Even for this case, the MMSE equalizer should perform better due to the presence of the additive noise tensor. For zero inter-tensor interference, (4.107) becomes

$$\breve{\mathcal{G}}(D) = \left\{ \mathcal{F}^H[0], \left(\{\mathcal{F}]0], \mathcal{F}^H[0]\}_{(N)} + N_0 \mathcal{I}_N \right)^{-1} \right\}_{(N)} \tag{4.108}$$

Next, consider the case where the equalizer has a finite number (M) of tensor taps $\{\mathcal{G}[i] \in \mathbb{C}_i^{I_1 \times \ldots \times I_N \times L_1 \times \ldots \times L_P}\}$, $i = 0, 1, \ldots, M-1$. Further, assume that the overall channel contains $v+1$ tensor taps $\{\mathcal{H}[i] \in \mathbb{C}_i^{L_1 \times \ldots \times L_P \times I_1 \times \ldots \times I_N}\}$, $i = 0, 1, \ldots, v$ such that the estimate of the data tensor is given by

$$\hat{\mathcal{D}}[k] = \sum_{i=0}^{M-1} \{\mathcal{G}[i], \mathcal{Y}[k-i]\}_{(P)} \tag{4.109}$$

There is a decision delay Δ, such that $0 \leq \Delta \leq M + v - 1$, to ensure causality. This delay is important when designing finite-length equalizers as non-causal filters cannot be implemented in practice. For the case of infinite-length equalizers, the delay is not considered as infinite-length systems are not realizable. Hence, the tensor $\hat{\mathcal{D}}[k]$ is an estimate of $\mathcal{D}[k-\Delta]$. The error tensor is

$$\mathcal{E}[k] = \hat{\mathcal{D}}[k] - \mathcal{D}[k-\Delta] = \sum_{i=0}^{M-1} \{\mathcal{G}[i], \mathcal{Y}[k-i]\}_{(P)} - \mathcal{D}[k-\Delta] \tag{4.110}$$

We define a tensor $\bar{\mathcal{Y}}[k] \in \mathbb{C}_k^{M \times L_1 \times \ldots \times L_P}$ with $P+1$ domains. Collecting the received tensor for different delays $\mathcal{Y}[k], \mathcal{Y}[k-1], \ldots, \mathcal{Y}[k-(M-1)]$ into an extended tensor $\bar{\mathcal{Y}}[k] \in \mathbb{C}_k^{M \times L_1 \times \ldots \times L_P}$, where the additional domain is the delay domain, we have

$$\bar{\mathcal{Y}}_{m,l_1,\ldots,l_P}[k] = \mathcal{Y}_{l_1,\ldots,l_P}[k-(m-1)] \quad \text{for } m = 1, \ldots, M \tag{4.111}$$

Similarly, define extended noise tensor $\bar{\mathcal{V}}[k] \in \mathbb{C}_k^{M \times L_1 \times \ldots \times L_P}$ and extended data tensor $\bar{\mathcal{D}}[k] \in \mathbb{C}_k^{(M+v) \times I_1 \times \ldots \times I_N}$ such that

$$\begin{aligned} \bar{\mathcal{D}}_{q,i_1,\ldots,i_N}[k] &= \mathcal{D}_{i_1,\ldots,i_N}[k-(q-1)] \quad \text{for } q = 1, \ldots, M+v \\ \bar{\mathcal{V}}_{m,l_1,\ldots,l_P}[k] &= \mathcal{V}_{l_1,\ldots,l_P}[k-(m-1)] \quad \text{for } m = 1, \ldots, M \end{aligned} \tag{4.112}$$

The *slice* of a tensor is defined as a two-dimensional section of a tensor obtained by fixing all but two indices [5]. For example, a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ has three slices denoted by $\boldsymbol{\mathcal{A}}_{i_1,:,:}$,$\boldsymbol{\mathcal{A}}_{:,i_2,:}$ and $\boldsymbol{\mathcal{A}}_{:,:,i_3}$. Define $\bar{\boldsymbol{\mathcal{H}}} \in \mathbb{C}_k^{M \times L_1 \times \ldots \times L_P \times (M+v) \times I_1 \times \ldots \times I_N}$ with two additional domains corresponding to the delays at the receiver and the transmitter such that the slice

$$\bar{\boldsymbol{\mathcal{H}}}_{:,l_1,\ldots,l_P,:,i_1,\ldots,i_N} =$$
$$\begin{bmatrix} \boldsymbol{\mathcal{H}}_{l_1 \ldots l_P i_1 \ldots i_N}[0] & \ldots & \boldsymbol{\mathcal{H}}_{l_1 \ldots l_P i_1 \ldots i_N}[v] & 0 & \ldots & \ldots & 0 \\ 0 & \boldsymbol{\mathcal{H}}_{l_1 \ldots l_P i_1 \ldots i_N}[0] & \ldots & \boldsymbol{\mathcal{H}}_{l_1 \ldots l_P i_1 \ldots i_N}[v] & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 0 & \boldsymbol{\mathcal{H}}_{l_1 \ldots l_P i_1 \ldots i_N}[0] & \ldots & \boldsymbol{\mathcal{H}}_{l_1 \ldots l_P i_1 \ldots i_N}[v] \end{bmatrix} \tag{4.113}$$

For a channel $\boldsymbol{\mathcal{H}}[k]$ with $v+1$ non-zero taps, (4.50) becomes $\boldsymbol{\mathcal{Y}}[k] = \sum_{n=0}^{v} \{\boldsymbol{\mathcal{H}}[k], \boldsymbol{\mathcal{D}}[k-n]\}_{(N)} + \boldsymbol{\mathcal{V}}[k]$ with components

$$\boldsymbol{\mathcal{Y}}_{l_1,\ldots,l_P}[k] = \sum_{n=0}^{v} \sum_{i_1} \ldots \sum_{i_N} \boldsymbol{\mathcal{H}}_{l_1,\ldots,l_P,i_1,\ldots,I_N}[n] \boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k-n] + \boldsymbol{\mathcal{V}}_{l_1,\ldots,l_P}[k] \tag{4.114}$$

Using (4.111) and (4.114) we have

$$\begin{aligned} \bar{\boldsymbol{\mathcal{Y}}}_{m,l_1,\ldots,l_P}[k] &= \boldsymbol{\mathcal{Y}}_{l_1,\ldots,l_P}[k-(m-1)] \\ &= \sum_{n=0}^{v} \sum_{i_1} \ldots \sum_{i_N} \boldsymbol{\mathcal{H}}_{l_1,\ldots,l_P,i_1,\ldots,I_N}[n] \boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k-(m-1)-n] + \boldsymbol{\mathcal{V}}_{l_1,\ldots,l_P}[k-(m-1)] \end{aligned}$$

and with (4.113) this becomes

$$\bar{\boldsymbol{\mathcal{Y}}}_{m,l_1,\ldots,l_P}[k] = \sum_{q=1}^{M+v} \sum_{i_1} \ldots \sum_{i_N} \bar{\boldsymbol{\mathcal{H}}}_{m,l_1,\ldots,l_P,q,i_1,\ldots,I_N} \bar{\boldsymbol{\mathcal{D}}}_{q,i_1,\ldots,i_N}[k] + \bar{\boldsymbol{\mathcal{V}}}_{m,l_1,\ldots,l_P}[k] \tag{4.115}$$

which in tensor notation gives

$$\bar{\boldsymbol{\mathcal{Y}}}[k] = \{\bar{\boldsymbol{\mathcal{H}}}, \bar{\boldsymbol{\mathcal{D}}}[k]\}_{(N+1)} + \bar{\boldsymbol{\mathcal{V}}}[k] \tag{4.116}$$

Further, define an augmented tensor $\bar{\boldsymbol{\mathcal{G}}} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times M \times L_1 \times \ldots \times L_P}$, which is a collection of the tensor equalizer taps $\boldsymbol{\mathcal{G}}[0], \ldots, \boldsymbol{\mathcal{G}}[M-1]$, whose components are

$$\bar{\boldsymbol{\mathcal{G}}}_{i_1,\ldots,i_N,m,l_1,\ldots,l_P} = \boldsymbol{\mathcal{G}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[(m-1)] \quad \text{for } m = 1, \ldots, M, \tag{4.117}$$

The components of $\hat{\boldsymbol{\mathfrak{D}}}[k]$ of (4.109) can be written as

$$\begin{aligned}
\hat{\mathfrak{D}}_{i_1,\ldots,i_N}[k] &= \sum_{m=0}^{M-1} \Big(\sum_{l_1} \ldots \sum_{l_P} \boldsymbol{\mathfrak{G}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[m] \boldsymbol{\mathfrak{Y}}_{l_1,\ldots,l_P}[k-m] \Big) \\
&= \sum_{l_1} \ldots \sum_{l_P} \boldsymbol{\mathfrak{G}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[0] \boldsymbol{\mathfrak{Y}}_{l_1,\ldots,l_P}[k] + \ldots \\
&\quad + \sum_{l_1} \ldots \sum_{l_P} \boldsymbol{\mathfrak{G}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[M-1] \boldsymbol{\mathfrak{Y}}_{l_1,\ldots,l_P}[k-M+1] \\
&= \sum_{l_1} \ldots \sum_{l_P} \bar{\boldsymbol{\mathfrak{G}}}_{i_1,\ldots,i_N,1,l_1,\ldots,l_p} \bar{\boldsymbol{\mathfrak{Y}}}_{1,l_1,\ldots,l_P}[k] + \ldots + \sum_{l_1} \ldots \sum_{l_P} \bar{\boldsymbol{\mathfrak{G}}}_{i_1,\ldots,i_N,M,l_1,\ldots,l_p} \bar{\boldsymbol{\mathfrak{Y}}}_{M,l_1,\ldots,l_P}[k] \\
&= \Big(\{\bar{\boldsymbol{\mathfrak{G}}}, \bar{\boldsymbol{\mathfrak{Y}}}[k]\}_{(P+1)} \Big)_{i_1,\ldots,i_N} \qquad (4.118)
\end{aligned}$$

hence (4.109) becomes

$$\hat{\boldsymbol{\mathfrak{D}}}[k] = \{\bar{\boldsymbol{\mathfrak{G}}}, \bar{\boldsymbol{\mathfrak{Y}}}[k]\}_{(P+1)} \qquad (4.119)$$

and (4.110) becomes

$$\boldsymbol{\mathfrak{E}}[k] = \{\bar{\boldsymbol{\mathfrak{G}}}, \bar{\boldsymbol{\mathfrak{Y}}}[k]\}_{(P+1)} - \boldsymbol{\mathfrak{D}}[k-\Delta] \qquad (4.120)$$

We wish to find the optimal co-efficients such that the mean squared error is minimized. Using Theorem 3 we have that the optimal multi-linear equalizer must satisfy

$$\boldsymbol{\mathfrak{R}}_{\boldsymbol{\mathfrak{E}},\boldsymbol{\mathfrak{Y}}}[i] = \mathbb{E}\big[(\hat{\boldsymbol{\mathfrak{D}}}[k] - \boldsymbol{\mathfrak{D}}[k-\Delta]) \circ \boldsymbol{\mathfrak{Y}}^*[k-i]\big] = \mathbf{0}_T \quad \text{for } |i| \leq M \qquad (4.121)$$

where $\mathbf{0}_T \in \mathbb{C}^{I_1 \times \ldots \times I_N \times L_1 \times \ldots \times L_P}$, which is equivalent to

$$\mathbb{E}\big[(\hat{\boldsymbol{\mathfrak{D}}}[k] - \boldsymbol{\mathfrak{D}}[k-\Delta]) \circ \bar{\boldsymbol{\mathfrak{Y}}}^*[k]\big] = \mathbf{0}_T \qquad (4.122)$$

$$\implies \mathbb{E}\big[\hat{\boldsymbol{\mathfrak{D}}}[k] \circ \bar{\boldsymbol{\mathfrak{Y}}}^*[k]\big] = \mathbb{E}\big[\boldsymbol{\mathfrak{D}}[k-\Delta]) \circ \bar{\boldsymbol{\mathfrak{Y}}}^*[k]\big] \qquad (4.123)$$

where $\mathbf{0}_T \in \mathbb{C}^{I_1 \times \ldots \times I_N \times M \times L_1 \times \ldots \times L_P}$. Substituting the value of $\hat{\boldsymbol{\mathfrak{D}}}[k]$ from (4.119), we get

$$\mathbb{E}\left[\{\bar{\boldsymbol{\mathfrak{G}}}, \bar{\boldsymbol{\mathfrak{Y}}}[k]\}_{(P+1)} \circ \bar{\boldsymbol{\mathfrak{Y}}}^*[k]\right] = \mathbb{E}\left[\boldsymbol{\mathfrak{D}}[k-\Delta] \circ \bar{\boldsymbol{\mathfrak{Y}}}^*[k]\right] \qquad (4.124)$$

Denoting $\boldsymbol{\mathfrak{R}}_{\bar{\boldsymbol{\mathfrak{y}}}}[0] = \mathbb{E}\left[\bar{\boldsymbol{\mathfrak{Y}}}[k] \circ \bar{\boldsymbol{\mathfrak{Y}}}^*[k]\right] = \boldsymbol{\mathfrak{R}}_{\bar{\boldsymbol{\mathfrak{y}}}}$, we have

$$\begin{aligned}
\boldsymbol{\mathfrak{R}}_{\bar{\boldsymbol{\mathfrak{y}}}} &= \mathbb{E}\left[\bar{\boldsymbol{\mathfrak{Y}}}[k] \circ \bar{\boldsymbol{\mathfrak{Y}}}^*[k]\right] \\
&= \mathbb{E}\left[\Big(\{\bar{\boldsymbol{\mathfrak{H}}}, \bar{\boldsymbol{\mathfrak{D}}}[k]\}_{(N+1)} + \bar{\boldsymbol{\mathfrak{V}}}[k] \Big) \circ \Big(\{\bar{\boldsymbol{\mathfrak{H}}}, \bar{\boldsymbol{\mathfrak{D}}}[k]\}_{(N+1)} + \bar{\boldsymbol{\mathfrak{V}}}[k] \Big)^* \right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\{\bar{\mathcal{H}}, \bar{\mathcal{D}}[k]\}_{(N+1)} \circ \{\bar{\mathcal{H}}^*, \bar{\mathcal{D}}^*[k]\}_{(N+1)}\right] + \mathbb{E}\left[\{\bar{\mathcal{H}}, \bar{\mathcal{D}}[k]\}_{(N+1)} \circ \bar{\mathcal{V}}^*[k]\right] \\
&\quad + \mathbb{E}\left[\bar{\mathcal{V}}[k] \circ \{\bar{\mathcal{H}}^*, \bar{\mathcal{D}}^*[k]\}_{(N+1)}\right] + \mathbb{E}\left[\bar{\mathcal{V}}[k] \circ \bar{\mathcal{V}}^*[k]\right] \\
&= \{\{\bar{\mathcal{H}}, \mathcal{R}_{\bar{\mathcal{D}}}\}_{(N+1)}, \bar{\mathcal{H}}^H\}_{(N+1)} + \mathcal{R}_{\bar{\mathcal{V}}} \qquad (4.125)
\end{aligned}
$$

where $\mathcal{R}_{\bar{\mathcal{V}}} = \mathbb{E}\left[\bar{\mathcal{V}}[k] \circ \bar{\mathcal{V}}^*[k]\right]$ and $\mathcal{R}_{\bar{\mathcal{D}}} = \mathbb{E}\left[\bar{\mathcal{D}}[k] \circ \bar{\mathcal{D}}^*[k]\right]$. Let $\mathbb{E}\left[\mathcal{D}[k-\Delta] \circ \bar{\mathcal{Y}}^*[k]\right] = \mathcal{R}_{\mathcal{D},\bar{\mathcal{Y}}}$. Using (4.116) gives

$$
\begin{aligned}
\mathcal{R}_{\mathcal{D},\bar{\mathcal{Y}}} &= \mathbb{E}\left[\mathcal{D}[k-\Delta] \circ \bar{\mathcal{Y}}^*[k]\right] \\
&= \mathbb{E}\left[\mathcal{D}[k-\Delta] \circ \left(\{\bar{\mathcal{H}}, \bar{\mathcal{D}}[k]\}_{(N+1)} + \bar{\mathcal{V}}[k]\right)^*\right] \\
&= \{\mathbb{E}\left[\mathcal{D}[k-\Delta] \circ \bar{\mathcal{D}}[k]\right], \bar{\mathcal{H}}^H\}_{(N+1)} \\
&= \{\mathcal{R}_{\mathcal{D},\bar{\mathcal{D}}}, \bar{\mathcal{H}}^H\}_{(N+1)} \qquad (4.126)
\end{aligned}
$$

where $\mathcal{R}_{\mathcal{D},\bar{\mathcal{D}}} = \mathbb{E}\left[\mathcal{D}[k-\Delta] \circ \bar{\mathcal{D}}[k]\right]$ and has components

$\mathcal{R}_{\mathcal{D},\bar{\mathcal{D}}_{i_1,\ldots,i_N,m,i'_1,\ldots,i'_N}} = \mathbb{E}\left[\mathcal{D}_{i_1,\ldots,i_N}[k-\Delta]\mathcal{D}^*_{i'_1,\ldots,i'_N}[k-m]\right]$

Assuming the data tensors are uncorrelated, we get

$$
\mathcal{R}_{\mathcal{D},\bar{\mathcal{D}}_{i_1,\ldots,i_N,m,i'_1,\ldots,i'_N}} = \begin{cases} \mathbb{E}\left[\mathcal{D}_{i_1,\ldots,i_N}[k-m] \circ \mathcal{D}^*_{i_1,\ldots,i_N}[k-m]\right] & \text{if } m = \Delta \text{ and } i_1 = i'_1, \ldots, i_N = i'_N \\ 0 & \text{otherwise} \end{cases} \qquad (4.127)
$$

Further, we have

$$
\begin{aligned}
\mathcal{R}_{\bar{\mathcal{Y}},\mathcal{D}} &= \mathbb{E}\left[\bar{\mathcal{Y}}[k] \circ \mathcal{D}^*[k-\Delta]\right] \\
&= \mathbb{E}\left[\left(\{\bar{\mathcal{H}}, \bar{\mathcal{D}}[k]\}_{(N+1)} + \bar{\mathcal{V}}[k]\right) \circ \mathcal{D}^*[k-\Delta]\right] \\
&= \{\bar{\mathcal{H}}, \mathbb{E}\left[\bar{\mathcal{D}}[k] \circ \mathcal{D}^*[k-\Delta]\right]\}_{(N+1)} \\
&= \{\bar{\mathcal{H}}, \left(\mathbb{E}\left[\mathcal{D}[k-\Delta] \circ \bar{\mathcal{D}}^*[k]\right]\right)^H\}_{(N+1)} \\
&= \left(\{\mathbb{E}\left[\mathcal{D}[k-\Delta] \circ \bar{\mathcal{D}}[k]\right], \bar{\mathcal{H}}^H\}_{(N+1)}\right)^H
\end{aligned}
$$

$$= \boldsymbol{\mathcal{R}}^H_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}} \tag{4.128}$$

The LHS of (4.124) becomes

$$\begin{aligned}\mathbb{E}\left[\{\bar{\boldsymbol{\mathcal{G}}},\bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)}\circ\bar{\boldsymbol{\mathcal{Y}}}^*[k]\right] &= \mathbb{E}\left[\{\bar{\boldsymbol{\mathcal{G}}},(\bar{\boldsymbol{\mathcal{Y}}}[k]\circ\bar{\boldsymbol{\mathcal{Y}}}^*[k])\}_{(P+1)}\right]\\ &= \left\{\bar{\boldsymbol{\mathcal{G}}},\left(\mathbb{E}\left[\bar{\boldsymbol{\mathcal{Y}}}[k]\circ\bar{\boldsymbol{\mathcal{Y}}}^*[k]\right]\right)\right\}_{(P+1)}\\ &= \{\bar{\boldsymbol{\mathcal{G}}},\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)}\end{aligned} \tag{4.129}$$

and the RHS of (4.124) becomes

$$\mathbb{E}\left[\boldsymbol{\mathcal{D}}[k-\Delta]\circ\bar{\boldsymbol{\mathcal{Y}}}^*[k]\right] = \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}} \tag{4.130}$$

Using (4.129) and (4.130), (4.124) becomes

$$\{\bar{\boldsymbol{\mathcal{G}}},\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)} = \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}} \tag{4.131}$$

To find the optimal tap co-efficients, we solve (4.131) by contracting both sides of (4.131) by $\boldsymbol{\mathcal{R}}^{-1}_{\bar{\boldsymbol{\mathcal{Y}}}}$. This gives

$$\{\{\bar{\boldsymbol{\mathcal{G}}},\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)},\boldsymbol{\mathcal{R}}^{-1}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)} = \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}},\boldsymbol{\mathcal{R}}^{-1}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)} \tag{4.132}$$

Using the associativity property we get

$$\{\bar{\boldsymbol{\mathcal{G}}},\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}},\boldsymbol{\mathcal{R}}^{-1}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)}\}_{(P+1)} = \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}},\boldsymbol{\mathcal{R}}^{-1}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)} \tag{4.133}$$

Since $\boldsymbol{\mathcal{R}}^{-1}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)} = \boldsymbol{\mathcal{I}}_{P+1}$, we have the optimal tap co-efficients

$$\bar{\boldsymbol{\mathcal{G}}}_{\text{opt}} = \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}},\boldsymbol{\mathcal{R}}^{-1}_{\bar{\boldsymbol{\mathcal{Y}}}}\}_{(P+1)} \tag{4.134}$$

4.2.5 Performance Analysis

In this section we present some performance measures for the detection methods described previously. For the Minimum Mean Square Error equalizer we define the $i_1,\ldots,i_N$th Mean squared error as

$$\begin{aligned}\epsilon_{i_1,\ldots,i_N} &= \mathbb{E}\Big[\big|\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]-\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\big|^2\Big]\\ &= \mathbb{E}\Big[(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]-\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k])(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]-\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k])^*\Big]\\ &= \mathbb{E}\Big[\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]\boldsymbol{\mathcal{D}}^*_{i_1,\ldots,i_N}[k]\Big] - \mathbb{E}\Big[\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\boldsymbol{\mathcal{D}}^*_{i_1,\ldots,i_N}[k]\Big]\\ &\quad - \mathbb{E}\Big[(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]-\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k])\hat{\boldsymbol{\mathcal{D}}}^*_{i_1,\ldots,i_N}[k]\Big]\end{aligned} \tag{4.135}$$

The third term in (4.135) can be expanded as

$$\begin{aligned}
&\mathbb{E}\big[(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]-\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k])\hat{\boldsymbol{\mathcal{D}}}^*_{i_1,\ldots,i_N}[k]\big] \\
&\qquad=\mathbb{E}\Big[\Big(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]-\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\Big)\Big(\sum_m\sum_{i_1}\cdots\sum_{i_N}\boldsymbol{\mathcal{G}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[m]\boldsymbol{\mathcal{Y}}_{i_1,\ldots,i_N}[k-m]\Big)^*\Big] \\
&\qquad=\sum_m\sum_{i_1}\cdots\sum_{i_N}\mathbb{E}\Big[\boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N}\Big(\boldsymbol{\mathcal{G}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[m]\boldsymbol{\mathcal{Y}}_{i_1,\ldots,i_N}[k-m]\Big)^*\Big] \\
&\qquad=0
\end{aligned} \tag{4.136}$$

where the last step is due to the fact that the error is uncorrelated with the observation. Using (4.136) in (4.135) we get

$$\begin{aligned}
\epsilon_{i_1,\ldots,i_N}&=\mathbb{E}\big[\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k]\boldsymbol{\mathcal{D}}^*_{i_1,\ldots,i_N}[k]\big]-\mathbb{E}\big[\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\boldsymbol{\mathcal{D}}^*_{i_1,\ldots,i_N}[k]\big] \\
&=\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}}[0]-\boldsymbol{\mathcal{R}}_{\hat{\boldsymbol{\mathcal{D}}},\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}}[0]
\end{aligned} \tag{4.137}$$

To find $\boldsymbol{\mathcal{R}}_{\hat{\boldsymbol{\mathcal{D}}},\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}}[0]$ we write the cross-correlation

$$\begin{aligned}
\boldsymbol{\mathcal{R}}_{\hat{\boldsymbol{\mathcal{D}}},\boldsymbol{\mathcal{D}}}[i]&=\mathbb{E}\big[\hat{\boldsymbol{\mathcal{D}}}[k]\circ\boldsymbol{\mathcal{D}}^*[k-i]\big] \\
&=\mathbb{E}\Big[\Big(\sum_m\{\boldsymbol{\mathcal{G}}[m],\boldsymbol{\mathcal{Y}}[k-m]\}_{(P)}\Big)\circ\boldsymbol{\mathcal{D}}^*[k-i]\Big]
\end{aligned}$$

Since the noise is uncorrelated with the input, we get

$$\begin{aligned}
&=\mathbb{E}\Big[\Big(\sum_m\{\boldsymbol{\mathcal{G}}[m],\sum_n\{\boldsymbol{\mathcal{H}}[n],\boldsymbol{\mathcal{D}}[k-m-n]\}_{(N)}\}_{(P)}\Big)\circ\boldsymbol{\mathcal{D}}^*[k-i]\Big] \\
&=\sum_m\sum_n\{\boldsymbol{\mathcal{G}}[m],\{\boldsymbol{\mathcal{H}}[n],\mathbb{E}\big[\boldsymbol{\mathcal{D}}[k-m-n]\circ\boldsymbol{\mathcal{D}}^*[k-i]\big]\}_{(N)}\}_{(P)} \\
&=\sum_m\sum_n\{\boldsymbol{\mathcal{G}}[m],\{\boldsymbol{\mathcal{H}}[n],\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}}[i-m-n]\}_{(N)}\}_{(P)}
\end{aligned} \tag{4.138}$$

The D-transform of (4.138) is

$$\breve{\boldsymbol{\mathcal{S}}}_{\hat{\boldsymbol{\mathcal{D}}},\boldsymbol{\mathcal{D}}}=\{\breve{\boldsymbol{\mathcal{G}}}(D),\{\breve{\boldsymbol{\mathcal{H}}}(D)\breve{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\mathcal{D}}}(D)\}_{(N)}\}_{(P)} \tag{4.139}$$

For a tensor $\boldsymbol{\mathcal{A}}[k]\in\mathbb{C}_k^{I_1\times\ldots\times I_N}$ with D-transform $\breve{\boldsymbol{\mathcal{A}}}(D)$, we have $\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N}[0]=\int_0^1\breve{\boldsymbol{\mathcal{A}}}_{i_1,\ldots,i_N}(e^{j2\pi f})df$, where $\breve{\boldsymbol{\mathcal{A}}}_{i_1,\ldots,i_N}(e^{j2\pi f})$ is found by setting $D=e^{j2\pi f}$. Hence, the $i_1,\ldots,i_N$th MSE can be found by integrating the $i_1,\ldots,i_N$th pseudo-diagonal element of $\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{D}}}(D)-\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{D}},\hat{\boldsymbol{\mathcal{D}}}}(D)$ along the unit circle by setting $D = e^{j2\pi f}$. i.e.,

$$\begin{aligned}\epsilon_{i_1,\ldots,i_N} &= \int_0^1 \left(\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{D}}}(e^{j2\pi f}) - \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{D}},\hat{\boldsymbol{\mathcal{D}}}}(e^{j2\pi f}) \right)_{i_1,\ldots,i_N,i_1,\ldots,i_N} df \\ &= \int_0^1 \left(\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{D}}}(e^{j2\pi f}) - \{\breve{\boldsymbol{\mathcal{G}}}(e^{j2\pi f}), \{\breve{\boldsymbol{\mathcal{H}}}(e^{j2\pi f}), \breve{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\mathcal{D}}}(e^{j2\pi f})\}_{(N)}\}_{(P)} \right)_{i_1,\ldots,i_N,i_1,\ldots,i_N} df \end{aligned} \tag{4.140}$$

The overall MSE is given by

$$\text{MSE} = \sum_{i_1 \ldots i_N} \epsilon_{i_1,\ldots,i_N} = trace\left(\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}}[0] - \boldsymbol{\mathcal{R}}_{\hat{\boldsymbol{\mathcal{D}}},\boldsymbol{\mathcal{D}}}[0] \right) \tag{4.141}$$

For the finite tap MMSE equalizer we have

$$\begin{aligned}\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}} &= \mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{E}}[k]^* \right] \\ &= \mathbb{E}\left[\left(\{\bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} - \boldsymbol{\mathcal{D}}[k] \right) \circ \left(\{\bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} - \boldsymbol{\mathcal{D}}[k] \right)^* \right] \\ &= \mathbb{E}\left[\{\bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} \circ \{\bar{\boldsymbol{\mathcal{G}}}^*, \bar{\boldsymbol{\mathcal{Y}}}^*[k]\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} \circ \boldsymbol{\mathcal{D}}^*[k] \right. \\ &\quad \left. - \boldsymbol{\mathcal{D}}[k] \circ \{\bar{\boldsymbol{\mathcal{G}}}^*, \bar{\boldsymbol{\mathcal{Y}}}^*[k]\}_{(P+1)} + \boldsymbol{\mathcal{D}}[k] \circ \boldsymbol{\mathcal{D}}[k]^* \right] \end{aligned} \tag{4.142}$$

Using (2.35) in (4.142) we get

$$\begin{aligned}\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}} &= \mathbb{E}\left[\{\bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} \circ \{\bar{\boldsymbol{\mathcal{Y}}}^*, \bar{\boldsymbol{\mathcal{G}}}^H\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{G}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} \circ \boldsymbol{\mathcal{D}}^*[k] \right. \\ &\quad \left. - \boldsymbol{\mathcal{D}}[k] \circ \{\bar{\boldsymbol{\mathcal{Y}}}^*[k], \bar{\boldsymbol{\mathcal{G}}}^H\}_{(P+1)} + \boldsymbol{\mathcal{D}}[k] \circ \boldsymbol{\mathcal{D}}[k]^* \right] \end{aligned} \tag{4.143}$$

Using the associativitiy property, (4.143) becomes

$$\begin{aligned}\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}} &= \{\bar{\boldsymbol{\mathcal{G}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}, \bar{\boldsymbol{\mathcal{G}}}^H\}_{(P+1)}\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{G}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}},\boldsymbol{\mathcal{D}}}\}_{(P+1)} \\ &\quad - \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \bar{\boldsymbol{\mathcal{G}}}^H\}_{(P+1)} + \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}} \end{aligned} \tag{4.144}$$

Substituting the optimal system tensor from (4.134) we get

$$\begin{aligned}\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}},\min} &= \{\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}^{-1}\}_{(P+1)}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}, (\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}^{-1}\}_{(P+1)})^H\}_{(P+1)}\}_{(P+1)} \\ &\quad - \{\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}^{-1}\}_{(P+1)}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}},\boldsymbol{\mathcal{D}}}\}_{(P+1)} - \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, (\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}^{-1}\}_{(P+1)})^H\}_{(P+1)} + \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}} \end{aligned} \tag{4.145}$$

$$= \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}},\boldsymbol{\mathcal{D}}}\}_{(P+1)}\}_{(P+1)} - 2\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}},\boldsymbol{\mathcal{D}}}\}_{(P+1)}\}_{(P+1)} + \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}} \tag{4.146}$$

$$= \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}} - \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\mathbf{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}},\boldsymbol{\mathcal{D}}}\}_{(P+1)}\}_{(P+1)} \tag{4.147}$$

$$= \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}} - \{\bar{\boldsymbol{\mathcal{G}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathbf{y}},\boldsymbol{\mathcal{D}}}\}_{(P+1)} \tag{4.148}$$

and the minimum mean squared error becomes

$$\mathrm{MSE}_{\min} = trace(\mathcal{R}_{\mathcal{E},\min}) = trace(\mathcal{R}_{\mathcal{D}} - \{\bar{\mathcal{G}}, \mathcal{R}_{\bar{\mathcal{Y}},\mathcal{D}}\}_{(P+1)}) \tag{4.149}$$

It can be shown that for a finite length equalizer with N taps, (4.145) tends to (4.137) in the limit as N tends to infinity. i.e.,

$$\lim_{N\to\infty} \mathcal{R}^{\text{finite}}_{\mathcal{E},\min} = \mathcal{R}^{\text{inf}}_{\mathcal{E},\min} \tag{4.150}$$

where $\mathcal{R}^{\text{finite}}_{\mathcal{E},\min}$ is the error correlation tensor for the finite tap equalizer and $\mathcal{R}^{\text{inf}}_{\mathcal{E},\min}$ is the error correlation tensor for the infinite tap equalizer. The proof of this can be found in Appendix A.4.

As a method of verification, we reproduce some performance results from the literature using the tensor framework. In particular, we use the results from [62] for GFDM as a reference with which we compare our results. We begin with a brief description of the system of [62] and then show the representation of the system using the tensor framework. The system consists of K subcarriers and M subsymbols. denote the data on the kth subcarrier and mth subsymbol for the nth GFDM symbol by $d_{k,m,n}$. Assuming an ideal channel and setting $c(t) = \delta(t)$ in (3.38) we get

$$r(t) = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} p_{k,m}(t-nT) d_{k,m,n} + v(t) \tag{4.151}$$

The analog processing at the receiver consists of a bank of filters $p_{R_{k,m}}(t)$ such that

$$\begin{aligned} y_{k',m'}(t) &= \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} p_{R_{k',m'}}(t) * p_{k,m}(t-nT) d_{k,m,n} + n_{k',m'}(t) \\ &= \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} q_{k',m',k,m}(t-nT) d_{k,m,n} + n_{k',m'}(t) \end{aligned} \tag{4.152}$$

where $q_{k',m',k,m}(t) = p_{R_{k',m'}}(t) * p_{k,m}(t)$ and $n_{k',m'}(t) = p_{R_{k',m'}}(t) * v(t)$. Sampling (4.152) at intervals of T we get

$$y_{k',m'}[s] = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} q_{k',m',k,m}(sT-nT) d_{k,m,n} + n_{k',m'}(sT) \tag{4.153}$$

In GFDM, there is no inter-symbol interference between successive GFDM symbols [52]. Hence, $y_{k',m'}[s]$ does not depend on $d_{k,m,n}$ for $n \neq s$. Without loss of generality, we consider

$y_{k',m'}[0]$ (i.e., $s = 0$) and, denoting it by $y_{k',m'}$, (4.153) becomes

$$y_{k',m'} = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} q_{k',m',k,m} d_{k,m} + n_{k',m'} \tag{4.154}$$

Assuming that the receive filter is band limited to B, where B is the bandwidth of one GFDM symbol, the noise $n(t)$ is also band limited and $n_{k',m'}$ are zero-mean complex Gaussian random variables with variance N_0. The desired symbol from (4.154) is $q_{k',m',k',m'} d_{k',m'}$ and the interference from other sub-carriers and sub-symbols is $\sum\limits_{k \neq k'} \sum\limits_{m \neq m'} q_{k',m',k,m} d_{k,m}$. Assuming that the data symbols $d_{k,m}$ have unit energy, the average energy per symbol is defined as $E_s = \frac{1}{KM} \sum\limits_{k',m'} |q_{k',m',k',m'}|^2$ and the average energy per bit is $\frac{E_s}{N_b}$ where N_b is the number of bits per symbol $d_{k,m}$. The SNR per bit is then defined as $\frac{E_b}{N_0}$. Define a matrix $\mathbf{A}$ with components $\mathbf{A}_{(k'+K(m'-1),(k+K(m-1)} = q_{k',m',k,m}$ and vectors $\mathbf{y}$ with components $\mathbf{y}_{k'+K(m'-1)} = y_{k',m'}$ and $\mathbf{d}$ with components $\mathbf{d}_{k'+K(m'-1)} = d_{k',m'}$. Now (4.154) can be re-written in matrix notation as

$$\mathbf{y} = \mathbf{A}\mathbf{d} + \mathbf{n} \tag{4.155}$$

where $\mathbf{n}$ has components $\mathbf{n}_{k+K(m-1)} = n_{k,m}$. The estimate of the data is

$$\begin{aligned} \hat{\mathbf{d}} &= \mathbf{B}\mathbf{y} \\ &= \mathbf{B}\mathbf{A}\mathbf{d} + \mathbf{B}\mathbf{n} \end{aligned} \tag{4.156}$$

The demodulation matrix $\mathbf{B}$ depends on the type of receiver used. For our purpose of reproduction of results, we look at two examples of matched filtering and zero forcing receivers for which we have $\mathbf{B} = \mathbf{A}^{\mathrm{H}}$ and $\mathbf{B} = \mathbf{A}^{+}$ respectively [62] where $()^{+}$ denotes Moore-Penrose pseudo-inversion.

The tensor framework allows a straightforward representation of (4.154). Define a tensor $\mathfrak{Q} \in \mathbb{C}^{K \times M \times K \times M}$ with components $\mathfrak{Q}_{k',m',k,m} = q_{k',m',k,m}$. We can now write (4.154) in tensor notation as

$$\mathfrak{Y} = \{\mathfrak{Q}, \mathfrak{D}\}_{(2)} + \mathfrak{N} \tag{4.157}$$

where $\boldsymbol{\mathcal{D}} \in \mathbb{C}_k^{K \times M}$ with components $\boldsymbol{\mathcal{D}}_{k,m} = d_{k,m}$ and $\boldsymbol{\mathcal{N}} \in \mathbb{C}_k^{K \times M}$ with components $\boldsymbol{\mathcal{N}}_{k,m} = n_{k,m}$. The data is estimated by

$$\hat{\boldsymbol{\mathcal{D}}} = \{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{Y}}\}_{(2)} = \{\boldsymbol{\mathcal{B}}, \{\boldsymbol{\mathcal{Q}}, \boldsymbol{\mathcal{D}}\}_{(2)}\}_{(2)} + \{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{N}}\}_{(2)} \tag{4.158}$$

where $\bar{\boldsymbol{\mathcal{B}}} \in \mathbb{C}^{K \times M \times K \times M}$ and its components depend on the type of detection used. For the matched filter system we have

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{P}}^{\mathrm{H}} \tag{4.159}$$

with components $\boldsymbol{\mathcal{B}}_{k',m',k'm} = \boldsymbol{\mathcal{P}}^*_{k,m,k',m'}$ and for the zero forcing case we have

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{P}}^{+} \tag{4.160}$$

To show the equivalence of the tensor framework representation to the representation used in [62] we show the performance of the two receivers discussed above and compare that with results from Fig. 3(a) and 3(b) from [62]. The system parameters used are as defined in Table 4.1. Figures 4.5 and 4.6 show the bit error rate (BER) results for the matched filter and zero forcing receivers for two different roll-off factors α and $\frac{E_b}{N_0} = 1, \ldots, 8[dB]$ averaged over 1000 GFDM symbols. We can see that the results from the tensor framework are consistent with those from [62] for both the receivers.

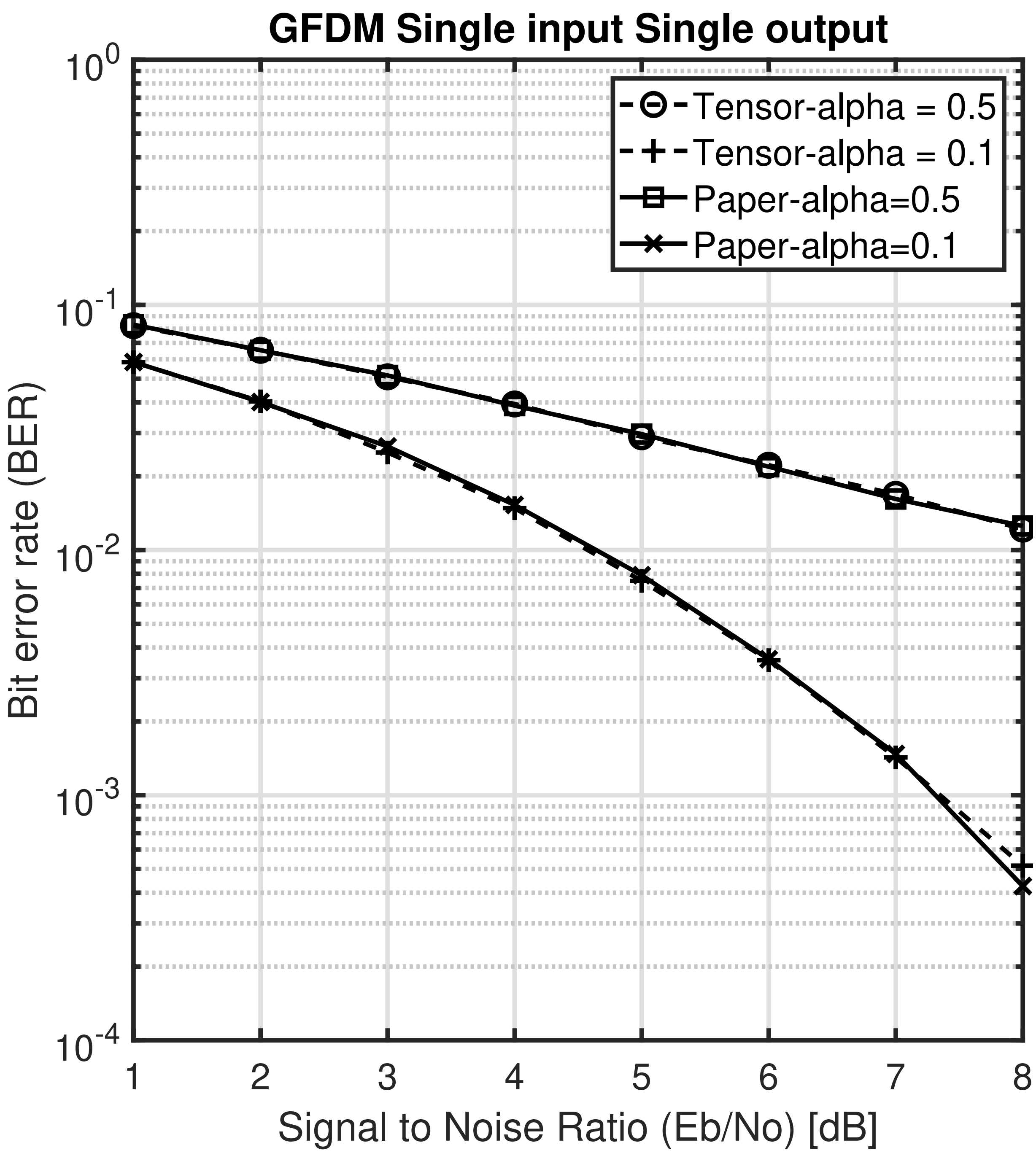


Fig. 4.5 GFDM Matched Filter Receiver

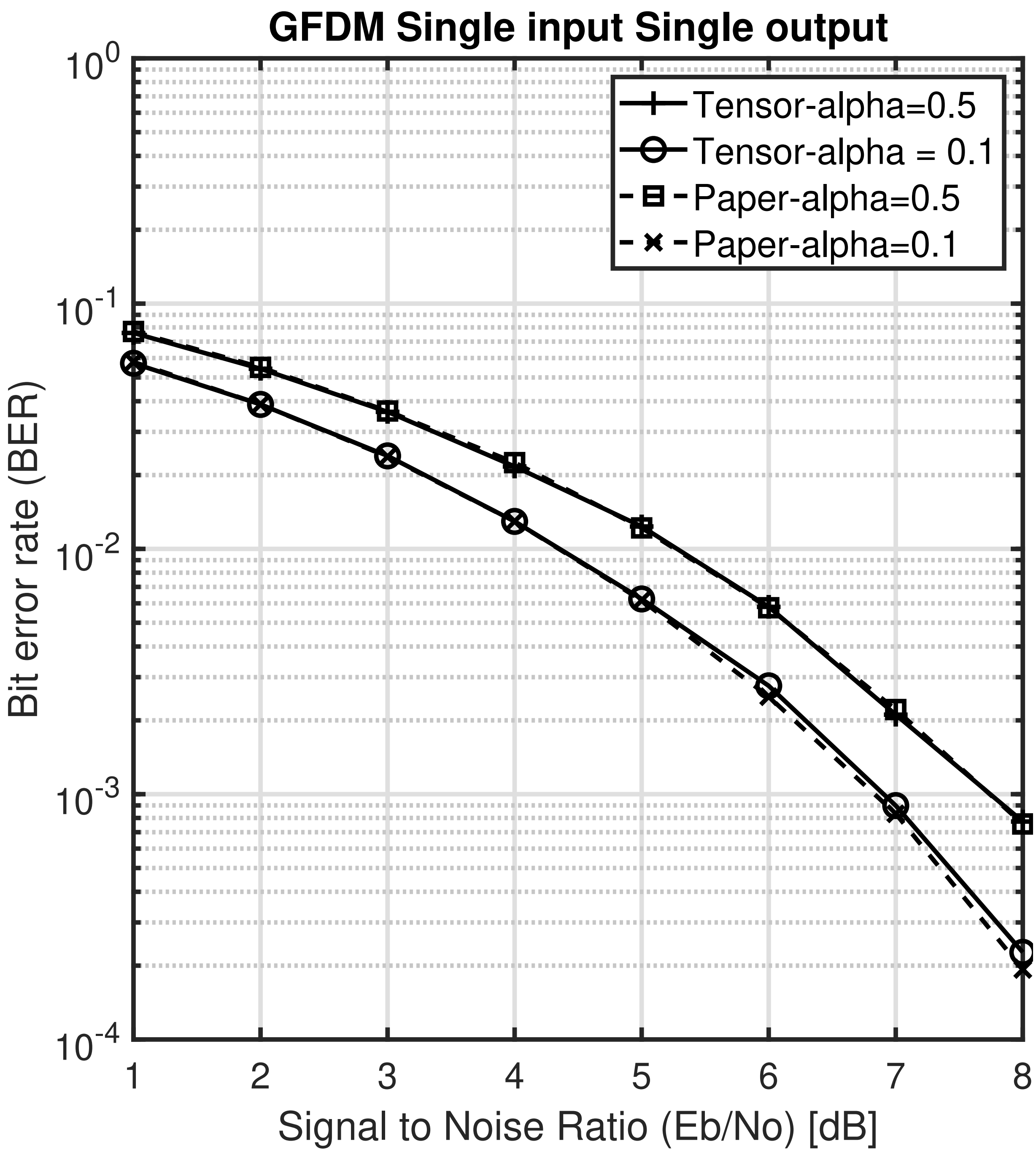


Fig. 4.6 GFDM Zero Forcing Receiver

Table 4.1 Table 1 of [62]

Description	parameter	value
Number of subcarriers	K	128
Number of time slots	M	5
Pulse shaping filter	g	Root Raised Cosine (RRC)
Roll-off factor	α	0.1 and 0.5
Modulation order	μ	2 (QPSK)

In Fig. 4.7 we present the minimum mean squared error vs P, the number of domains at the receiver, at SNR = 30dB. The results are averaged over 1000 channel realizations. The input $\boldsymbol{\mathcal{D}}[k] \in \mathbb{C}_k^{2\times 2\times 2\times 2}$ is a fourth order tensor with components drawn from an i.i.d source with $E_s = 1$ and 16-QAM is used for modulation. Hence we have $N = 4$, implying $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}} = \boldsymbol{\mathcal{I}}_4$ and $\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}} = \boldsymbol{\mathcal{I}}_5$. The noise tensor $\boldsymbol{\mathcal{V}}[k] \in \mathbb{C}_k^{\overbrace{2\times\ldots\times 2}^{P}}$, whose components are complex Gaussian with zero mean and variance N_0, has the same size as the received tensor $\boldsymbol{\mathcal{Y}}[k] \in \mathbb{C}_k^{\overbrace{2\times\ldots\times 2}^{P}}$. The channels used for the simulation contain three tensor taps ($v = 2$) $\boldsymbol{\mathcal{H}}[0]$, $\boldsymbol{\mathcal{H}}[1]$ and $\boldsymbol{\mathcal{H}}[2]$, whose components are randomly generated complex zero-mean uncorrelated Gaussian random variables with unit variance per complex sample. The equalizer used for the simulations contains $M = 7$ taps. Hence, we have $\bar{\boldsymbol{\mathcal{V}}}[k] \in \mathbb{C}_k^{7\times 2\times\ldots\times 2}$, $\bar{\boldsymbol{\mathcal{Y}}}[k] \in \mathbb{C}_k^{7\times 2\times\ldots\times 2}$ and $\bar{\boldsymbol{\mathcal{H}}} \in \mathbb{C}_k^{7\times 2\times\ldots\times 2\times 9\times 2\times\ldots\times 2}$ as defined in (4.112) and (4.113) such that (4.125) and (4.126) become

$$\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}} = \{\bar{\boldsymbol{\mathcal{H}}}, \bar{\boldsymbol{\mathcal{H}}}^H\}_{(5)} + N_0 \boldsymbol{\mathcal{I}}_{P+1} \tag{4.161}$$

and

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}} = \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{H}}}^H\}_{(5)} \tag{4.162}$$

Substituting the values of $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}}$ and $\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}$ from (4.161) and (4.162) in (4.145), and using $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}} = \boldsymbol{\mathcal{R}}^H_{\bar{\boldsymbol{\mathcal{Y}}},\boldsymbol{\mathcal{D}}}$, gives the analytical minimum mean squared error as

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}},\min} = \boldsymbol{\mathcal{I}}_4 - \left\{ \left(\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{H}}}^H\}_{(5)} \right), \left\{ \left(\{\bar{\boldsymbol{\mathcal{H}}}, \bar{\boldsymbol{\mathcal{H}}}^H\}_{(5)} + N_0 \boldsymbol{\mathcal{I}}_{(P+1)} \right)^{-1}, \left(\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{H}}}^H\}_{(5)} \right)^H \right\}_{(P+1)} \right\}_{(P+1)} \tag{4.163}$$

The simulations carried out consider four different number of domains at the receiver which are summarized in Table 4.2. The simulation results are presented in Fig. 4.7. The channels used for the simulation contain three tensor taps $\boldsymbol{\mathcal{H}}[0]$, $\boldsymbol{\mathcal{H}}[1]$ and $\boldsymbol{\mathcal{H}}[2]$, whose components are randomly generated complex zero-mean uncorrelated Gaussian random variables with unit variance per complex sample. The simulated mean squared error is consistent with

Table 4.2 Dimension sizes of the receiver

No. of Domains	Size of Receive tensor $\boldsymbol{\mathcal{Y}}[k]$	Size of channel tensor $\boldsymbol{\mathcal{H}}[k]$
1	$\mathbb{C}^{2}$	$\mathbb{C}^{2\times2\times2\times2\times2}$
2	$\mathbb{C}^{2\times2}$	$\mathbb{C}^{2\times2\times2\times2\times2\times2}$
3	$\mathbb{C}^{2\times2\times2}$	$\mathbb{C}^{2\times2\times2\times2\times2\times2\times2}$
4	$\mathbb{C}^{2\times2\times2\times2}$	$\mathbb{C}^{2\times2\times2\times2\times2\times2\times2\times2}$

the minimum mean squared error obtained from using (4.163). Further, for a fixed SNR ($30dB$), the minimum mean squared error decreases as the number of domains at the receiver are increased. This is because with each additional domain at the receiver, the number of samples in the receive tensor $\boldsymbol{\mathcal{Y}}[k]$ is doubled and hence there is better averaging of the noise.

Another interesting conclusion that can be drawn is that performance improvements can be made to a system by the addition of domains rather than having to increase the size of the individual domains themselves. To illustrate this, we compute the mean squared error for a system with a channel of size $\mathbb{C}^{2\times2\times2\times2\times8}$. This system contains only one domain at the receiver of size 8. The mean squared error for this case, at SNR $= 30dB$, was found to be $MSE = 8.16$. From Fig. 4.7, we can see that this value of mean squared error is reached when there are 3 domains of size 2 each at the receiver. Similarly, the mean squared error for a system of size $\mathbb{C}^{2\times2\times2\times2\times16}$ at SNR $= 30dB$ was found to be $MSE = 1.35$. From Fig. 4.7, we can see that this value of mean squared error is reached when there are 4 domains of size 2 each at the receiver. In summary, the MSE for a system with 3 equal domains of size 2 reaches the same performance as a system with one large receive domain of size 8 and a system with 4 equal domains of size 2 reaches the performance of a system with one large receive domain of size 16. This is useful when the size of a domain is constrained.

For example, in certain cases performance improvements through the addition of frequency or time domains might be more desirable as compared to the addition of more antennas. Table 4.3 shows the equivalent system employing a single large domain for each of the four systems that have been presented in Fig. 4.7.

In Fig. 4.8 we present the error rate for different equalizer tap lengths $M = 2K + 1$

Table 4.3 Comparison of one large receive domain and multiple smaller receive domains.

Size of channel tensor $\boldsymbol{\mathcal{H}}[k]$	Size of equivalent system with one large receive domain
$\mathbb{C}^{2\times2\times2\times2\times2}$	$\mathbb{C}^{2\times2\times2\times2\times2}$
$\mathbb{C}^{2\times2\times2\times2\times2\times2}$	$\mathbb{C}^{2\times2\times2\times2\times4}$
$\mathbb{C}^{2\times2\times2\times2\times2\times2\times2}$	$\mathbb{C}^{2\times2\times2\times2\times8}$
$\mathbb{C}^{2\times2\times2\times2\times2\times2\times2\times2}$	$\mathbb{C}^{2\times2\times2\times2\times16}$

($K = 3, 5, 7$) plotted against the averaged receive SNR for the finite tap MMSE equalizer. Also shown for comparison is the performance for of zero forcing equalizers with the same number of taps. The energy per symbol E_s is the total energy of one tensor symbol E_T divided by the number of symbols per tensor. i.e., $E_s = \frac{E_T}{8}$, and the SNR is defined as $\frac{E_s}{N_0}$. The input data tensor $\boldsymbol{\mathcal{D}}[k]$ is of size $2 \times 2 \times 2$ with components drawn from an i.i.d source, $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}}[i] = \boldsymbol{\mathcal{I}}_3$, and 4-QAM is used for modulation. The channel used consists of two taps ($v = 1$). i.e., the received tensor $\boldsymbol{\mathcal{Y}}[k] \in \mathbb{C}_k^{2\times2\times2}$ only contains inter-tensor interference from $\boldsymbol{\mathcal{D}}[k-1]$. The channel is assumed to be time-invariant and known at the receiver. For each realization of a test channel, the components of $\boldsymbol{\mathcal{H}}[k]$ are drawn from a complex Gaussian distribution such that each complex sample has zero mean and unit variance. In this case (4.125) and (4.126) become

$$\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}} = \{\bar{\boldsymbol{\mathcal{H}}}, \bar{\boldsymbol{\mathcal{H}}}^H\}_{(4)} + N_0\boldsymbol{\mathcal{I}}_4 \tag{4.164}$$

and

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}} = \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{H}}}^H\}_{(4)} \tag{4.165}$$

The coefficients of the equalizer are calculated using (4.134) and the error rate is found by averaging MATLAB simulation results over 100 channel realizations, accumulating 250

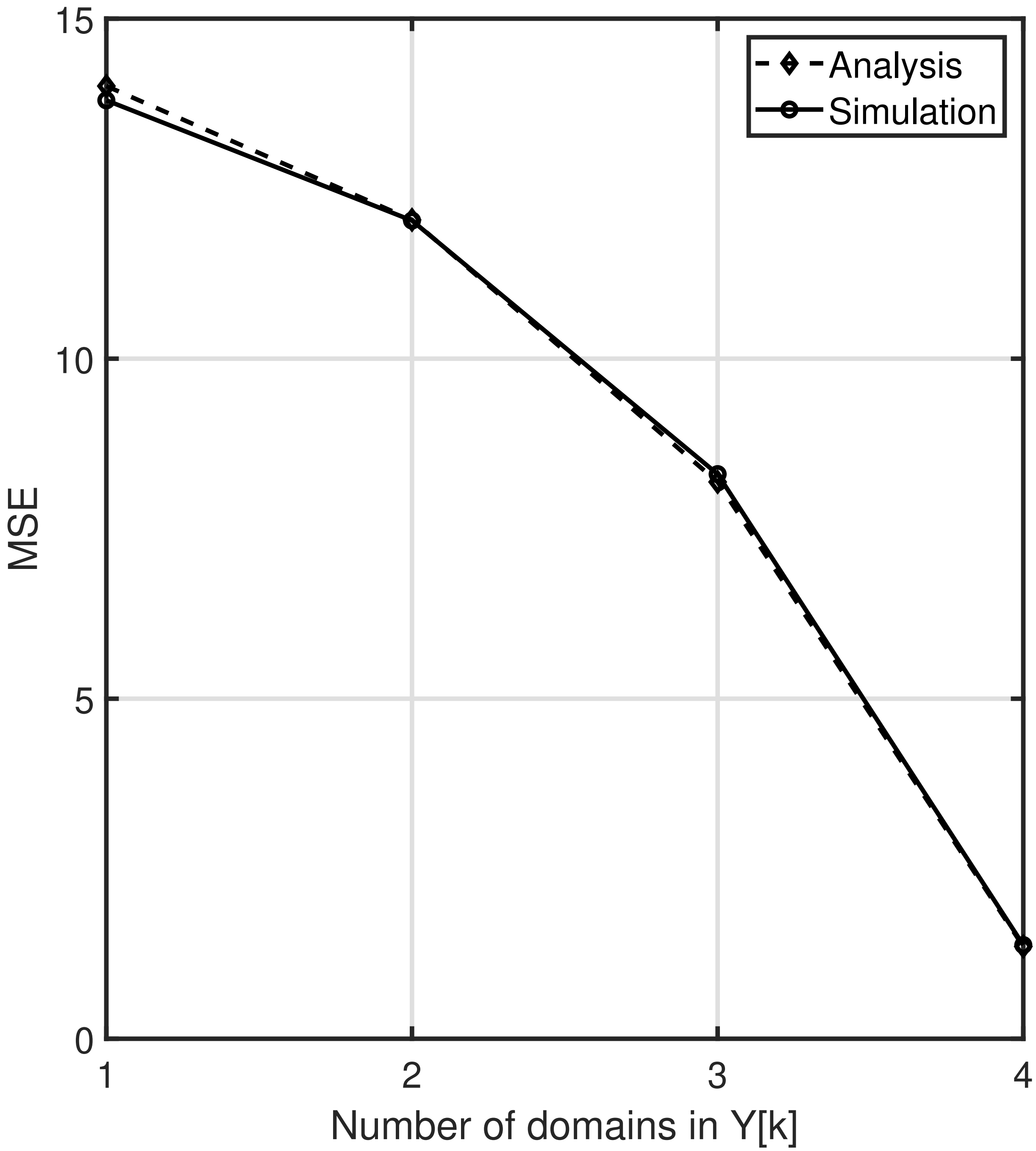


Fig. 4.7 Minimum Mean Squared Error versus number of receive domains for a system with 4 transmit domains of size 2 each

errors at each SNR. Table 4.4 summarizes the simulation parameters for figures 4.8 and 4.9.

We can see that there is an improvement in error performance as the number of tensor taps in the equalizer increases. Further, the performance of the zero forcing equalizer for the same number of taps is worse than the MMSE equalizer but the difference between their performance decreases with increased SNR. In Fig. 4.9 we present results for a channel with three taps ($v = 2$). The SER results of figures 4.9 and 4.8 exhibit a saturation for higher values of SNR. This is because the finite tap equalizers are not able to completely eliminate the interference even as noise vanishes. We can also see that the saturation floors decrease as the number of equalizer taps are increased and that the flooring occurs at a larger value of SNR. This is because the residual interference decreases as we increase the number of equalizer taps. J. G. Proakis et al. present similar error rate saturation results for $M \times N$ MIMO systems in [57]. M. K. Varanasi et al. show that the BER saturation floor for overloaded CDMA systems of size $K \times N$ is a decreasing function of $\beta = \frac{K}{N}$. From the results of figures 4.8 and 4.9, we can see that the SER floors depend on the ratio of the product of the domains of $\bar{\boldsymbol{\mathcal{D}}}[k] \in \mathbb{C}_k^{(M+v)\times 2\times 2\times 2}$ and $\bar{\boldsymbol{\mathcal{Y}}}[k] \in \mathbb{C}_k^{M\times 2\times 2\times 2}$. Factoring out the common domains, we get that the error floors are a decreasing function of

$$\beta = \frac{M+v}{M} \tag{4.166}$$

where M is the number of equalizer taps and $v+1$ is the number of taps in the channel. For the results of Fig. 4.8 ($v = 1$) this becomes $\beta = \frac{M+1}{M}$ and for Fig. 4.9 ($v = 2$) we have $\beta = \frac{M+2}{M}$. For a fixed v, β decreases with an increase in the number of equalizer taps M and hence leads to a lower error floor.

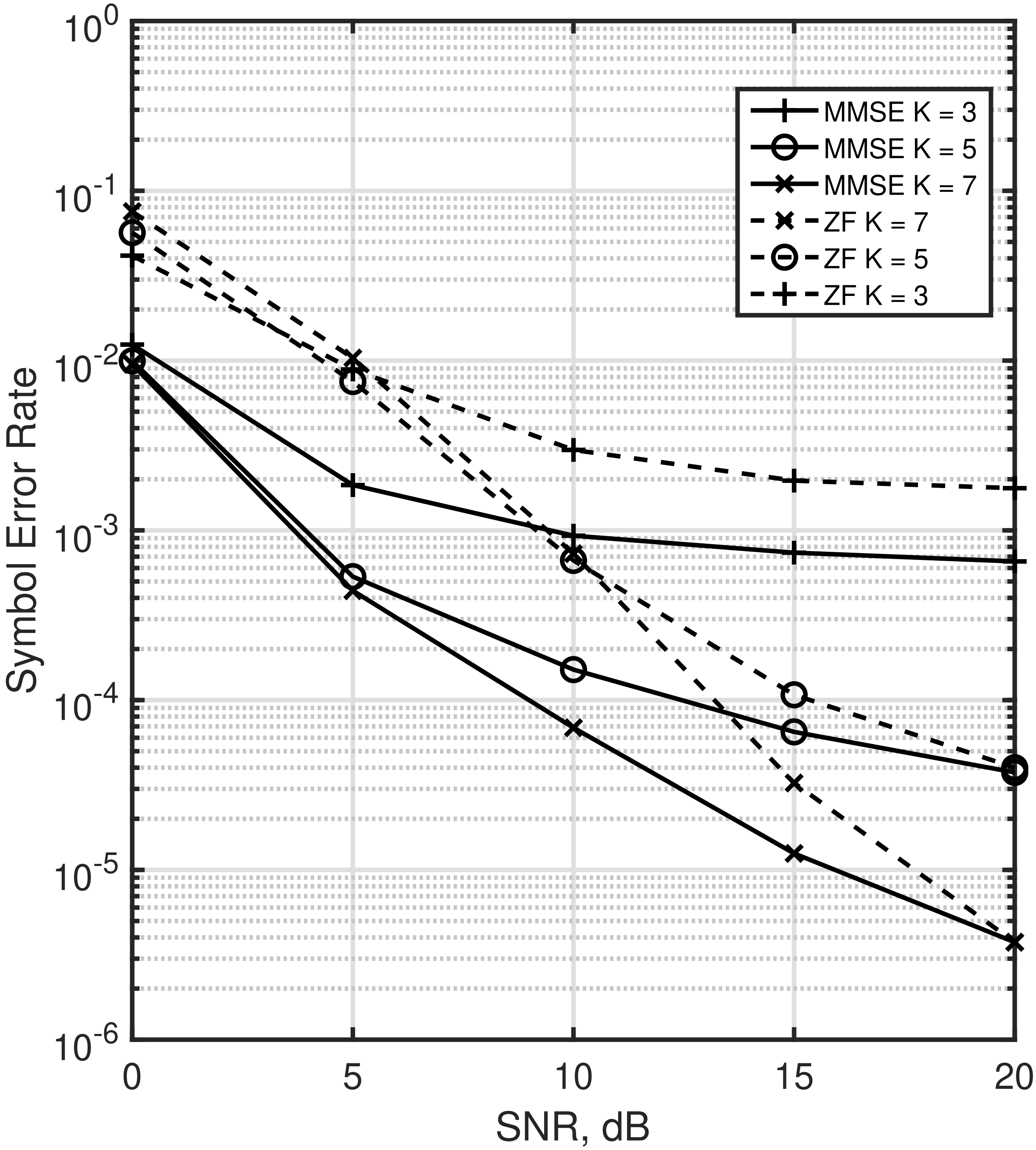


Fig. 4.8 Finite Tap MMSE Equaliser for a channel with L=2

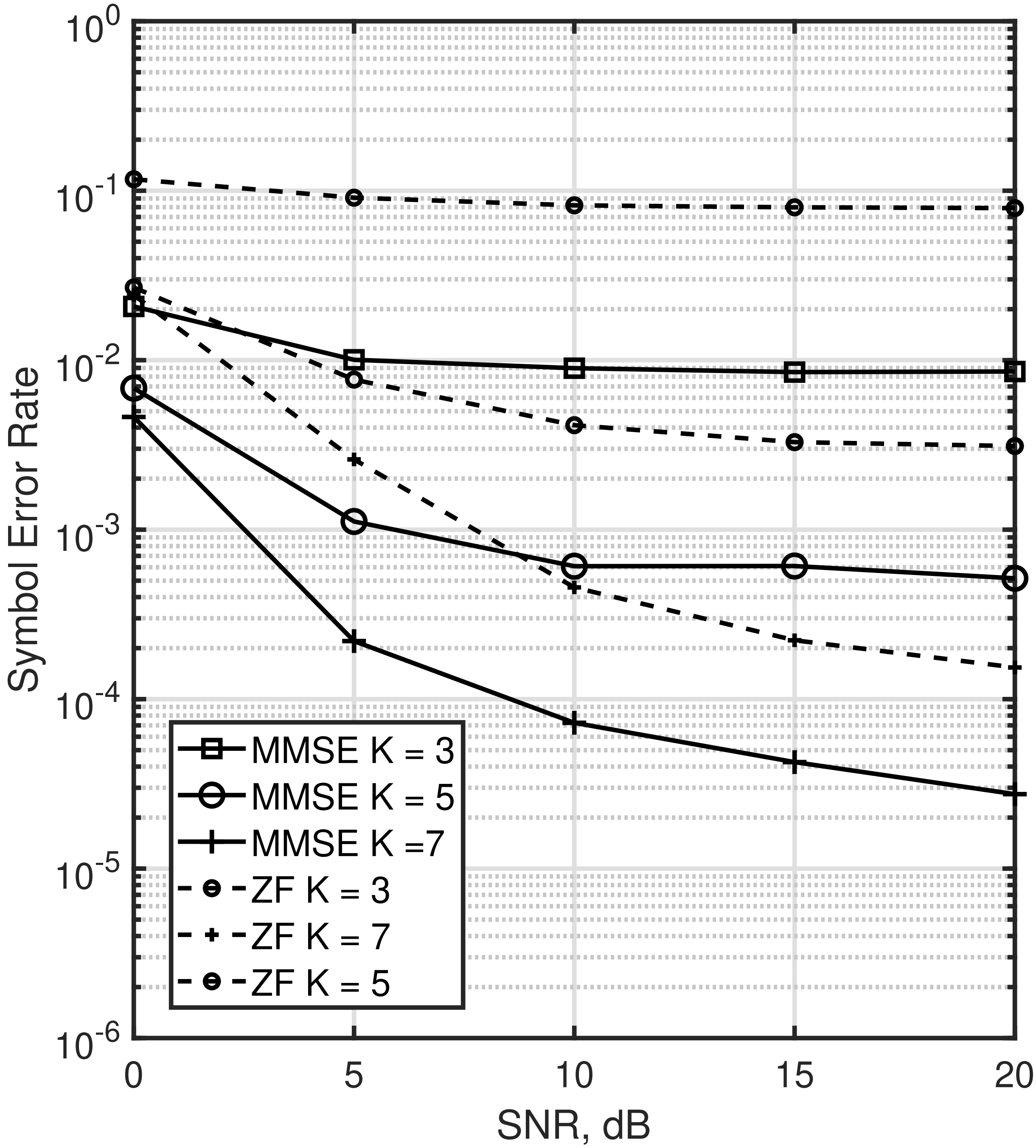


Fig. 4.9 Finite Tap MMSE Equaliser for a channel with L=3

Table 4.4 Simulation Parameters

Description	Fig. 4.8	Fig. 4.9
Number of channel taps	2	3
Size of data	$2 \times 2 \times 2$	$2 \times 2 \times 2$
Size of received tensor	$2 \times 2 \times 2$	$2 \times 2 \times 2$
Modulation	4-QAM	4-QAM

The error performance when there is no inter-tensor interference is illustrated in Fig. 4.10. We consider the case when there is only intra-tensor interference, and hence each component of $\mathcal{Y}[k]$ is a linear combination of the components of $\mathcal{D}[k]$. It is assumed that $\mathcal{R}_{\mathcal{D}}[0] = \mathcal{I}_{(N)}$ and $\mathcal{R}_{\mathcal{V}}[0] = N_0\mathcal{I}_{(N)}$. The channel contains only one non-zero tap ($v = 0$) $\mathcal{H} = \mathcal{H}[0]$ and the equalizer also contains one tap $\mathcal{G}$. The simulations are carried out for channels of three different sizes: $(2 \times 2 \times 2 \times 2 \times 2 \times 2)$, $(2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2)$ and $(2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2)$ and the corresponding sizes of the transmit tensors are: $(2 \times 2 \times 2)$, $(2 \times 2 \times 2 \times 2)$ and $(2 \times 2 \times 2 \times 2 \times 2)$. We can see that the performance of the MMSE equalizer is better than the zero forcing equalizer. This is attributed to the fact that there is significant noise enhancement due to the contraction of the noise tensor $\mathcal{V}[k]$ with $\mathcal{H}^{-1}$. Further, as the size of the channels, and hence the number of domains in the receiver, increases, the MMSE equalizer performs better. This is because increasing the number of domains, and hence the number of samples, results in better averaging of the noise. Unlike the MMSE equalizer that optimizes the mean squared error, the zero forcing equalizer simply eliminates the interference from the other components of the data tensor at the expense of noise enhancement. In the three systems of Fig. 4.10, the number of transmit domains increase at the same rate as the number of receiver domains and any gain in performance from the additional receive domains is nullified by the additional data being transmitted on the added transmit domains that increase the noise enhancement that occurs when channel inversion is performed.

Comparing Fig. 4.10 with Fig. 4.8 we can see that the multi-tap equalizers perform better even though there is more interference in the system of Fig. 4.8. It was found that in this case as the number of equalier taps are increased, the performance of the equalizers in the latter improve. This is illustrated in Fig. 4.11. This is not true for the case where there is only intra-tensor interference. To this end, Fig. 4.11 shows the variation of the mean squared error with the number of equalizer taps at a fixed SNR of $5dB$. As we can see, the equalizer for the case where there is no inter-tensor interference ($L = 1$) does not benefit from increasing the number of taps while the equalizer for $L = 2$ performs worse for $K = 1$ but improves as the number of taps are increased.

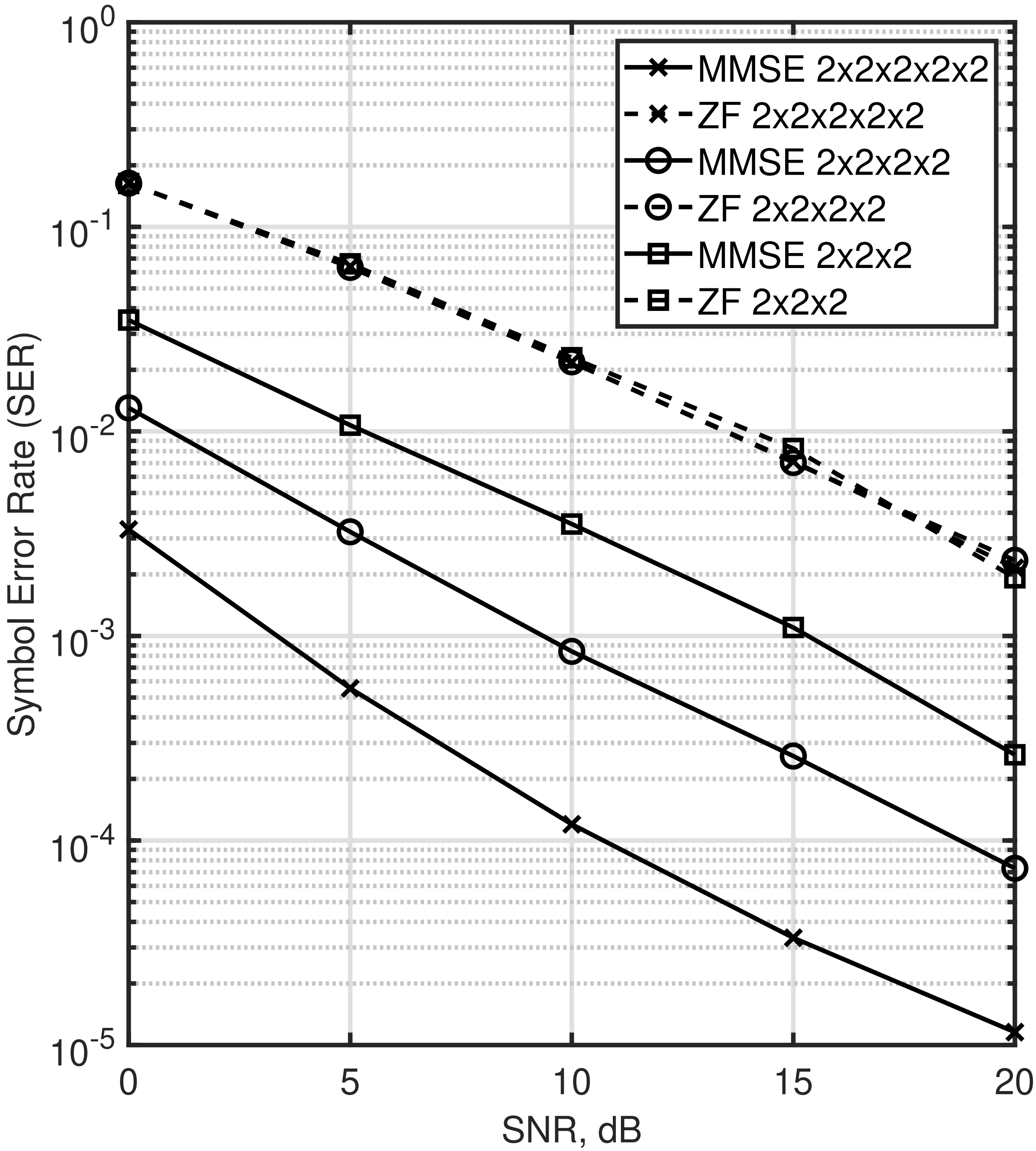


Fig. 4.10 ZF and MMSE Equalizers for zero inter-tensor interference. The size in the legend corresponds to the size of the transmit tensor

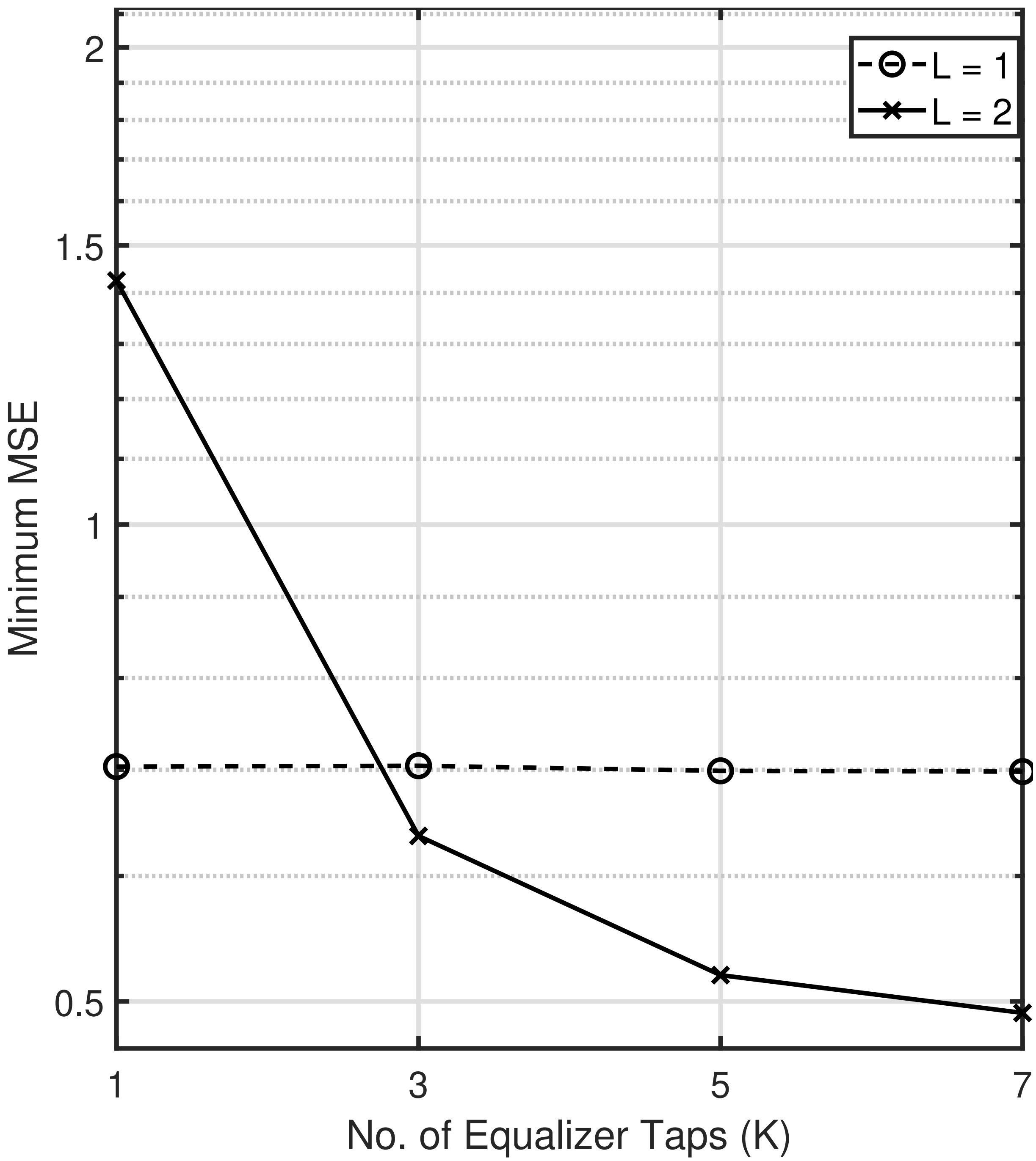


Fig. 4.11 MSE for different equalizer tap lengths

4.3 Decision Feedback Equalization

Decision Feedback Equalization (DFE) is a non-linear equalization strategy where previously detected symbols assist in the equalization and detection of subsequent symbols. The Tensor Decision Feedback Equalizer consists of two parts, a feed-forward system tensor $\boldsymbol{\mathcal{W}}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N\times L_1\times\ldots\times L_P}$ and a feedback system tensor $\boldsymbol{\mathcal{B}}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$. The structure of the decision feedback equalizer is shown in Fig.4.12.

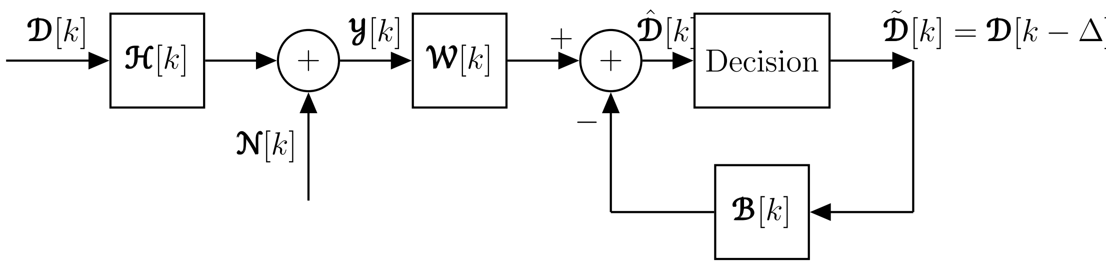


Fig. 4.12 System Model of the DFE

Finite tap DFE

Consider the case where the feedforward system has N_f tensor taps $\{\boldsymbol{\mathcal{W}}[i] \in \mathbb{C}^{I_1\times\ldots\times I_N\times L_1\times\ldots\times L_P},\ i = 0,1,\ldots,N_f-1$ and the feedback filter has N_b+1 tensor taps $\{\boldsymbol{\mathcal{B}}[i] \in \mathbb{C}^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}\},\ i = 0,1,\ldots,N_b$. Further, assume that the overall channel contains $v+1$ tensor taps $\{\boldsymbol{\mathcal{H}}[i] \in \mathbb{C}^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}\},\ i = 0,1,\ldots,v$ such that the estimate of the data tensor is given by

$$\hat{\boldsymbol{\mathcal{D}}}[k] = \sum_{i=0}^{N_f-1} \{\boldsymbol{\mathcal{W}}[i], \boldsymbol{\mathcal{Y}}[k-i]\}_{(P)} - \sum_{j=0}^{N_b} \{\boldsymbol{\mathcal{B}}[j], \tilde{\boldsymbol{\mathcal{D}}}[k-j]\}_{(N)} \tag{4.167}$$

where $\tilde{\boldsymbol{\mathcal{D}}}[k]$ is the tensor containing decisions at time k. There is a constant delay Δ such that the decision $\tilde{\boldsymbol{\mathcal{D}}}[k]$ corresponds to an input $\boldsymbol{\mathcal{D}}[k-\Delta]$. If the kth decision is correct we have $\tilde{\boldsymbol{\mathcal{D}}}[k] = \boldsymbol{\mathcal{D}}[k-\Delta]$. Define an error tensor $\boldsymbol{\mathcal{E}}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N}$ such that

$$\boldsymbol{\mathcal{E}}[k] = \hat{\boldsymbol{\mathcal{D}}}[k] - \boldsymbol{\mathcal{D}}[k-\Delta] = \sum_{i=0}^{N_f-1} \{\boldsymbol{\mathcal{W}}[i], \boldsymbol{\mathcal{Y}}[k-i]\}_{(P)} - \sum_{j=0}^{N_b} \{\boldsymbol{\mathcal{B}}[i], \tilde{\boldsymbol{\mathcal{D}}}[k-i]\}_{(N)} - \boldsymbol{\mathcal{D}}[k-\Delta] \tag{4.168}$$

Under the assumption of correct past decisions, i.e., if $\tilde{\mathcal{D}}[k] = \mathcal{D}[k-\Delta]$, (4.167) becomes

$$\hat{\mathcal{D}}[k] = \sum_{i=0}^{N_f-1} \{\mathcal{W}[i], \mathcal{Y}[k-i]\}_{(P)} - \sum_{j=0}^{N_b} \{\mathcal{B}[j], \mathcal{D}[k-\Delta-j]\}_{(N)} \tag{4.169}$$

and (4.168) becomes

$$\mathcal{E}[k] = \sum_{i=0}^{N_f-1} \{\mathcal{W}[i], \mathcal{Y}[k-i]\}_{(P)} - \sum_{j=1}^{N_b} \{\mathcal{B}[j], \mathcal{D}[k-\Delta-j]\}_{(N)} - \Big(\mathcal{D}[k-\Delta] + \{\mathcal{B}[0], \mathcal{D}[k-\Delta]\}_{(N)}\Big) \tag{4.170}$$

Defining, as for the linear equalizer case, augmented tensors $\bar{\mathcal{Y}}[k] \in \mathbb{C}_k^{N_f \times L_1 \times \ldots \times L_P}$,$\bar{\mathcal{D}}[k] \in \mathbb{C}_k^{(N_f+N_b) \times I_1 \times \ldots \times I_N}$, $\bar{\mathcal{N}}[k] \in \mathbb{C}_k^{N_f \times L_1 \times \ldots \times L_P}$ and $\bar{\mathcal{H}} \in \mathbb{C}^{N_f \times L_1 \times \ldots \times L_P \times (N_f+N_b) \times I_1 \times \ldots \times I_N}$ that are a collection of the receive tensor, data tensor, noise tensor and channel for different delays we have

$$\bar{\mathcal{Y}}_{m,l_1,\ldots,l_P}[k] = \mathcal{Y}_{l_1,\ldots,l_P}[k-(m-1)] \quad \text{for } m = 1, \ldots, N_f \tag{4.171}$$

$$\bar{\mathcal{D}}_{q,l_1,\ldots,l_P}[k] = \mathcal{D}_{i_1,\ldots,i_N}[k-(q-1)] \quad \text{for } q = 1, \ldots, N_f + v \tag{4.172}$$

$$\bar{\mathcal{N}}_{m,l_1,\ldots,l_P}[k] = \mathcal{V}_{l_1,\ldots,l_P}[k-(m-1)] \quad \text{for } m = 1, \ldots, N_f \tag{4.173}$$

and

$$\bar{\mathcal{H}}_{:,l_1,\ldots,l_P,:,i_1,\ldots,i_N} =$$

$$\begin{bmatrix} \mathcal{H}_{l_1\ldots l_P i_1 \ldots i_N}[0] & \ldots & \mathcal{H}_{l_1\ldots l_P i_1 \ldots i_N}[v] & 0 & \ldots & \ldots & 0 \\ 0 & \mathcal{H}_{l_1\ldots l_P i_1 \ldots i_N}[0] & \ldots & \mathcal{H}_{l_1\ldots l_P i_1 \ldots i_N}[v] & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 0 & \mathcal{H}_{l_1\ldots l_P i_1 \ldots i_N}[0] \ldots & \mathcal{H}_{l_1\ldots l_P i_1 \ldots i_N}[v] \end{bmatrix} \tag{4.174}$$

The relation between $\mathcal{H}[k]$, $\mathcal{D}[k]$, $\mathcal{N}[k]$ and $\mathcal{Y}[k]$ is $\mathcal{Y}[k] = \sum_{n=0}^{v} \{\mathcal{H}[k], \mathcal{D}[k-n]\}_{(N)} + \mathcal{N}[k]$ with components

$$\mathcal{Y}_{l_1,\ldots,l_P}[k] = \sum_{n=0}^{v} \sum_{i_1} \ldots \sum_{i_N} \mathcal{H}_{l_1,\ldots,l_P,i_1,\ldots,I_N}[n] \mathcal{D}_{i_1,\ldots,i_N}[k-n] + \mathcal{N}_{l_1,\ldots,l_P}[k] \tag{4.175}$$

Using the components of $\bar{\mathcal{Y}}_{m,l_1,\ldots,l_P}$ from (4.171) and the relation (4.175) we have

$$\bar{\mathcal{Y}}_{m,l_1,\ldots,l_P}[k] = \mathcal{Y}_{l_1,\ldots,l_P}[k-(m-1)] \tag{4.176}$$

$$= \sum_{n=0}^{v} \sum_{i_1} \cdots \sum_{i_N} \mathcal{H}_{l_1,\ldots,l_P,i_1,\ldots,I_N}[n] \mathcal{D}_{i_1,\ldots,i_N}[k-(m-1)-n] + \mathcal{N}_{l_1,\ldots,l_P}[k-(m-1)]$$

Writing (4.176) in terms of (4.113),(4.172) and (4.173) gives

$$\bar{\mathcal{Y}}_{m,l_1,\ldots,l_P}[k] = \sum_{q=1}^{N_f+v} \sum_{i_1} \cdots \sum_{i_N} \bar{\mathcal{H}}_{m,l_1,\ldots,l_P,q,i_1,\ldots,I_N} \bar{\mathcal{D}}_{q,i_1,\ldots,i_N}[k] + \bar{\mathcal{N}}_{m,l_1,\ldots,l_P}[k] \tag{4.177}$$

which in tensor notation gives

$$\bar{\mathcal{Y}}[k] = \{\bar{\mathcal{H}}, \bar{\mathcal{D}}[k]\}_{(N+1)} + \bar{\mathcal{N}}[k] \tag{4.178}$$

Define tensors $\bar{\mathcal{W}}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N\times N_f\times L_1\times\ldots\times L_P}$ and $\bar{\mathcal{B}}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N\times(\Delta+N_b+1)\times I_1\times\ldots\times I_N}$ such that

$$\bar{\mathcal{W}}_{i_1,\ldots,i_N,m,l_1,\ldots,l_P}[k] = \mathcal{W}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[k-(m-1)] \quad \text{for } m = 1,\ldots,N_f \tag{4.179}$$

$$\bar{\mathcal{B}}_{i'_1,\ldots,i'_N,m,i_1,\ldots,i_N} = \begin{cases} 0 \quad \text{for } 1 \le m \le \Delta \\ \mathcal{I}_{N_{i'_1,\ldots,i'_N,i_1,\ldots,i_N}} + \mathcal{B}_{i'_1,\ldots,i'_N,i_1,\ldots,i_N}[0] \quad \text{for } m = \Delta+1 \\ \mathcal{B}_{i'_1,\ldots,i'_N,i_1,\ldots,i_N}[m-\Delta] \quad \text{for } \Delta+1 < m \le \Delta+Nb \end{cases} \tag{4.180}$$

We can then rewrite (4.170) using these tensors as

$$\mathcal{E}[k] = \{\bar{\mathcal{W}}, \bar{\mathcal{Y}}[k]\}_{(P+1)} - \{\bar{\mathcal{B}}, \bar{\mathcal{D}}[k]\}_{(N+1)} \tag{4.181}$$

The mean squared error tensor is

$$\mathcal{R}_{\mathcal{E}} = \mathcal{R}_{\mathcal{E}}[0] = \mathbb{E}\left[\mathcal{E}[k] \circ \mathcal{E}[k]^*\right] \tag{4.182}$$

which can be expanded using (4.181) to give

$$\mathcal{R}_{\mathcal{E}} = \mathbb{E}\left[\mathcal{E}[k] \circ \mathcal{E}[k]^*\right] = \left\{\bar{\mathcal{B}}, \left\{\left(\mathcal{R}_{\bar{\mathcal{D}}} - \{\mathcal{R}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}, \{\mathcal{R}_{\bar{\mathcal{Y}}}^{-1}, \mathcal{R}_{\bar{\mathcal{Y}},\bar{\mathcal{D}}}\}_{(P+1)}\}_{(P+1)}\right), \bar{\mathcal{B}}^H\right\}_{(N+1)}\right\}_{(N+1)}$$
$$= \{\bar{\mathcal{B}}, \{\mathcal{A}, \bar{\mathcal{B}}^H\}_{(N+1)}\}_{(N+1)} \tag{4.183}$$

where we have defined $\mathcal{A} = \left(\mathcal{R}_{\bar{\mathcal{D}}} - \{\mathcal{R}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}, \{\mathcal{R}_{\bar{\mathcal{Y}}}^{-1}, \mathcal{R}_{\bar{\mathcal{Y}},\bar{\mathcal{D}}}\}_{(P+1)}\}_{(P+1)}\right)$. The detailed derivation of this is provided in Appendix A.5. The optimal feedback filter minimizes the mean squared error (4.183). i.e., we need to find $\min_{\bar{\mathcal{B}}} trace(\mathcal{R}_{\mathcal{E}}) = \min_{\bar{\mathcal{B}}} trace(\{\bar{\mathcal{B}}, \{\mathcal{A}, \bar{\mathcal{B}}^H\}_{(N+1)}\}_{(N+1)})$. The decision at time k is aided by past decisions from previous tensors at time $k-1, k-2, \ldots, k-Nb$. From (4.180), we require that $\bar{\mathcal{B}}_{m,i'_1,\ldots,i'_N,i_1,\ldots,i_N} = 0$ for $1 \le m \le \Delta$ and

$\bar{\mathcal{B}}_{m,i'_1,\ldots,i'_N,i_1,\ldots,i_N} = \mathcal{I}_{N_{i'_1,\ldots,i'_N,i_1,\ldots,i_N}}$ for $m = \Delta+1$. Define a tensor $\mathcal{S} \in \mathbb{C}^{(\Delta+N_b+1)\times I_1\times\ldots\times I_N\times(\Delta+1)\times I_1\times\ldots\times I_N}$ with components

$$\mathcal{S}_{m,i'_1,\ldots,i'_N,n,i_1,\ldots,i_N} = \begin{cases} 1 & \text{if } m = n, i_1 = i'_1, \ldots, i'_1 = i'_N \\ 0 & \text{otherwise} \end{cases} \tag{4.184}$$

and a tensor $\mathcal{T} \in \mathbb{C}^{I_1\times\ldots\times I_N\times(\Delta+1)\times I_1\times\ldots\times I_N}$ with components

$$\mathcal{T}_{i_1,\ldots,i_N,n,i'_1,\ldots,i'_N} = \begin{cases} 0 & \text{if } 1 \le n \le \Delta \\ \mathcal{I}_{N_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} & \text{for } n = \Delta + 1 \end{cases} \tag{4.185}$$

The requirement that $\bar{\mathcal{B}}_{m,i'_1,\ldots,i'_N,i_1,\ldots,i_N} = 0$ for $1 \le m \le \Delta$ and $\bar{\mathcal{B}}_{m,i'_1,\ldots,i'_N,i_1,\ldots,i_N} = \mathcal{I}_{N_{i'_1,\ldots,i'_N,i_1,\ldots,i_N}}$ for $m = \Delta + 1$, in terms of tensors $\mathcal{S}$ and $\mathcal{T}$ is $\{\bar{\mathcal{B}}, \mathcal{S}\}_{(N+1)} = \mathcal{T}$. Hence, to find the optimal feedback filter, we need to solve the following constrained optimization problem:

$$\min_{\bar{\mathcal{B}}} trace(\{\bar{\mathcal{B}}, \{\mathcal{A}, \bar{\mathcal{B}}^H\}_{(N+1)}\}_{(N+1)}) \quad \text{s.t. } \{\bar{\mathcal{B}}, \mathcal{S}\}_{(N+1)} = \mathcal{T} \tag{4.186}$$

To solve this, define a tensor of Lagrange coefficients $\boldsymbol{\lambda} \in \mathbb{C}^{I_1\times\ldots\times I_N\times(\Delta+1)\times I_1\times\ldots\times I_N}$ and the Lagrangian function

$$J = trace(\{\bar{\mathcal{B}}, \{\mathcal{A}, \bar{\mathcal{B}}^H\}_{(N+1)}\}_{(N+1)}) + \langle \boldsymbol{\lambda}, (\{\bar{\mathcal{B}}, \mathcal{S}\}_{(N+1)} - \mathcal{T})\rangle \tag{4.187}$$

where $\langle(,)\rangle$ is the tensor inner product. For a tensor $\mathcal{X} = \mathcal{X}^R + j\mathcal{X}^I$, extending the gradient vector in [63], we define a corresponding tensor gradient operator ∇, with components

$$\nabla_{i_1,\ldots,i_N,l_1,\ldots,l_M} = \frac{\partial}{\partial \mathcal{X}^R_{i_1,\ldots,i_N,l_1,\ldots,l_M}} - j\frac{\partial}{\partial \mathcal{X}^I_{i_1,\ldots,i_N,l_1,\ldots,l_M}} \tag{4.188}$$

The gradient of J with respect to $\bar{\mathcal{B}}$ is

$$\nabla_{\bar{\mathcal{B}}} J = \frac{\partial trace(\{\bar{\mathcal{B}}, \{\mathcal{A}, \bar{\mathcal{B}}^H\}_{(N+1)}\}_{(N+1)})}{\partial \bar{\mathcal{B}}} + \frac{\partial \langle \boldsymbol{\lambda}, (\{\bar{\mathcal{B}}, \mathcal{S}\}_{(N+1)} - \mathcal{T})\rangle}{\partial \bar{\mathcal{B}}} \tag{4.189}$$

Using (4.188) and $\bar{\mathcal{B}} = \bar{\mathcal{B}}_R + j\bar{\mathcal{B}}_I$ and solving, it can be shown that

$$\frac{\partial trace(\{\bar{\mathcal{B}}, \{\mathcal{A}, \bar{\mathcal{B}}^H\}_{(N+1)}\}_{(N+1)})}{\partial \bar{\mathcal{B}}} = \{\bar{\mathcal{B}}^*, \mathcal{A}^T\}_{(N+1)} \tag{4.190}$$

and

$$\frac{\partial \langle \boldsymbol{\lambda}, (\{\bar{\mathcal{B}}, \mathcal{S}\}_{(N+1)} - \mathcal{T})\rangle}{\partial \bar{\mathcal{B}}} = \{\boldsymbol{\lambda}, \mathcal{S}^T\}_{(N+1)} \tag{4.191}$$

We omit the proof of the above for the sake of brevity. Using (4.190) and (4.191) we get

$$\nabla_{\bar{\mathcal{B}}} J = \{\bar{\mathcal{B}}^*, \mathcal{A}^T\}_{(N+1)} + \{\boldsymbol{\lambda}, \mathcal{S}^T\}_{(N+1)} \tag{4.192}$$

$$\implies \bar{\boldsymbol{\mathcal{B}}} = -\{\{\boldsymbol{\lambda}^*, \boldsymbol{\mathcal{S}}^H\}_{(N+1)}, \boldsymbol{\mathcal{A}}^{-1}\}_{(N+1)} \tag{4.193}$$

where in the last step we have used the fact that $\boldsymbol{\mathcal{A}}^H = \boldsymbol{\mathcal{A}}$. If the inverse of $\boldsymbol{\mathcal{A}}$ does not exist, then the minimum-norm least square solution of (4.192) is

$$\boldsymbol{\mathcal{B}} = -\{\{\boldsymbol{\lambda}^*, \boldsymbol{\mathcal{S}}^H\}_{(N+1)}, \boldsymbol{\mathcal{A}}^{+}\}_{(N+1)} \tag{4.194}$$

where $\boldsymbol{\mathcal{A}}^+$ is the Moore-Pensore pseudoinverse of $\boldsymbol{\mathcal{A}}$ [22]. Substituting for $\boldsymbol{\mathcal{B}}$ from (4.193) in the constraint equation $\{\bar{\boldsymbol{\mathcal{B}}}, \boldsymbol{\mathcal{S}}\}_{(N+1)} = \boldsymbol{\mathcal{T}}$ gives

$$-\{\{\{\boldsymbol{\lambda}^*, \boldsymbol{\mathcal{S}}^H\}_{(N+1)}, \boldsymbol{\mathcal{A}}^{-1}\}_{(N+1)}, \boldsymbol{\mathcal{S}}\}_{(N+1)} = -\{\boldsymbol{\lambda}^*, (\{\{\boldsymbol{\mathcal{S}}^H, \boldsymbol{\mathcal{A}}^{-1}\}_{(N+1)}, \boldsymbol{\mathcal{S}}\}_{(N+1)})\}_{(N+1)} = \boldsymbol{\mathcal{T}} \tag{4.195}$$

which gives

$$-\boldsymbol{\lambda}^* = \{\boldsymbol{\mathcal{T}}, (\{\{\boldsymbol{\mathcal{S}}^H, \boldsymbol{\mathcal{A}}^{-1}\}_{(N+1)}, \boldsymbol{\mathcal{S}}\}_{(N+1)})^{-1}\}_{(N+1)} \tag{4.196}$$

Substituting for $\boldsymbol{\lambda}$ in (4.193) gives the optimal feedback tensor as

$$\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}} = \{\boldsymbol{\mathcal{T}}, \{(\{\{\boldsymbol{\mathcal{S}}^H, \boldsymbol{\mathcal{A}}^{+}\}_{(N+1)}, \boldsymbol{\mathcal{S}}\}_{(N+1)})^{-1}, \boldsymbol{\mathcal{S}}^H\}_{(N+1)}\}_{(N+1)}, \boldsymbol{\mathcal{A}}^{-1}\}_{(N+1)} \tag{4.197}$$

For the scalar case, $\boldsymbol{\mathcal{T}}$ becomes a $(\Delta+1)$-length row vector $\mathbf{t} = [0 \dots 0, 1]$, $\boldsymbol{\mathcal{S}}$ becomes a matrix

$$\mathbf{S} = \begin{bmatrix} \mathbf{I}_{(\Delta+1)} \\ \mathbf{0}_{N_b \times (\Delta+1)} \end{bmatrix} \tag{4.198}$$

and the optimal feedback filter (4.197) degenerates to

$$\mathbf{b}_{\text{opt}} = \mathbf{t}(\mathbf{S}^H \mathbf{A} \mathbf{S})^{-1} \mathbf{S}^H \mathbf{A}^{-1} \tag{4.199}$$

where $\mathbf{A} = \mathbf{R_D} - \mathbf{R_{D,Y}} \mathbf{R_Y}^{-1} \mathbf{R_{YD}}$ and $\mathbf{b}_{\text{opt}} = [0, \dots, 0, 1, b_1, \dots, b_{N_b}]$. This is consistent with the optimal feedback filter for the SISO case in [61].

The optimal feed-forward tensor is found by using (4.197) in (A.48)

$$\bar{\boldsymbol{\mathcal{W}}}_{\text{opt}} = \{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{\mathcal{Y}}}}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}^{-1}\}_{(P+1)}\}_{(N+1)} \tag{4.200}$$

The minimum mean squared error MMSE_{DFE} is

$$\text{MMSE}_{\text{DFE}} = trace(\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}},\min}) = trace(\{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \{\boldsymbol{\mathcal{A}}, \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^H\}_{(N+1)}\}_{(N+1)}) \tag{4.201}$$

Next, we derive the relation between the MMSE of the DFE and the MMSE of the linear equalizer. Substituting the value of $\boldsymbol{\mathcal{A}}$ in the error auto-correlation gives

$$\begin{aligned}\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}} &= \left\{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \left\{\left(\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}} - \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{\mathcal{Y}}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}\right), \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^{H}\right\}_{(N+1)}\right\}_{(N+1)} \\ &= \{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^{H}\}_{(N+1)}\}_{(N+1)} - \{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{\mathcal{Y}}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}, \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^{H}\}_{(N+1)}\}_{(N+1)}\end{aligned} \tag{4.202}$$

When the number of feedback taps $N_b = 0$, we get

$$\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}_{i'_1,\ldots,i'_N,m,i_1,\ldots,i_N}} = \begin{cases} 0 & \text{for } 1 \le m \le \Delta \\ \boldsymbol{\mathcal{I}}_{N_{i'_1,\ldots,i'_N,i_1,\ldots,i_N}} & \text{for } m = \Delta + 1 \end{cases} \tag{4.203}$$

Using (4.203) we may write $\{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^{H}\}_{(N+1)}\}_{(N+1)}$ component wise as

$$\begin{aligned} &\left(\{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \{(\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^{H}\}_{(N+1)}\}_{(N+1)}\right)_{i'_1,\ldots,i'_N,j'_1,\ldots,j'_N} \\ &= \sum_{m,i_1,\ldots,i_N} \sum_{m',j_1,\ldots,j_N} \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}_{i'_1,\ldots,i'_N,m,i_1,\ldots,i_N}} \mathbb{E}[\bar{\boldsymbol{\mathcal{D}}}_{m,i_1,\ldots,i_N} \bar{\boldsymbol{\mathcal{D}}}^{*}_{m',j_1,\ldots,j_N}] \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}_{i'_1,\ldots,i'_N,m,i_1,\ldots,i_N}} \\ &= \mathbb{E}[\bar{\boldsymbol{\mathcal{D}}}_{(\Delta+1),i'_1,\ldots,i'_N} \bar{\boldsymbol{\mathcal{D}}}^{*}_{(\Delta+1),j'_1,\ldots,j'_N}] \\ &= \mathbb{E}[\boldsymbol{\mathcal{D}}_{i'_1,\ldots,i'_N}[k-\Delta] \boldsymbol{\mathcal{D}}^{*}_{,j'_1,\ldots,j'_N}[k-\Delta]] = \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}_{i'_1,\ldots,i'_N,j'_1,\ldots,j'_N}} \end{aligned} \tag{4.204}$$

which can be written in tensor notation as $\{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \{(\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^{H}\}_{(N+1)}\}_{(N+1)} = \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}}$. Following similar lines, we can show that $\{\bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{\mathcal{Y}}}}\}_{(N+1)} = \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}}$ and $\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}_{\text{opt}}^{H}\}_{(N+1)} = \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}},\boldsymbol{\mathcal{D}}}$. Substituting these in (4.202) gives

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}} = \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}} - \{\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}},\bar{\boldsymbol{\mathcal{Y}}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{Y}}},\boldsymbol{\mathcal{D}}}\}_{(P+1)}\}_{(P+1)} \tag{4.205}$$

Comparing this to (4.145), we can see that when there are no feedback taps, the error auto-correlation for the DFE-MMSE equalizer is the same as the MMSE linear equalizer. Fig. 4.13 shows the minimum mean squared error vs the number of domains at the receiver for different values of SNR $(0, 10, 20, 30)$dB. The channels used for the simulation contain three taps $(v = 2)$ and the results are averaged over 1000 channel realizations. The components of the three taps are randomly generated circular complex zero-mean uncorrelated Gaussian random variables with unit variance per complex sample. The input $\boldsymbol{\mathcal{D}}[k] \in \mathbb{C}_k^{2\times2\times2\times2}$ is a fourth order tensor and $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{D}}}[0] = \boldsymbol{\mathcal{I}}_4$. As the number of domains are increased, the mean squared error decreases. Moreover, the improvement in the mean squared error performance is much greater with each additional domain. This can be seen from the fact that going from 1 domain to 2 results in a 30% reduction in the MSE while going from 2 to 3 domains results in a drop of over 50%. We can hence leverage domain diversity to improve performance in situations where there are restrictions on the size of the domains themselves. These results are consistent with the MSE results of the linear MMSE equalizer.

As an example we consider MIMO GFDM with tensor DFE, and present some performance results. The data being transmitted in the nth MIMO GFDM symbol (which contains $P = 2$ streams of K subcarriers and M subsymbols) is a third order tensor $\mathcal{D}[n] \in \mathbb{C}_n^{K\times M\times P}$ with $\mathcal{R}_{\mathcal{D}} = \mathcal{I}_3$. There are 2 transmit and 2 receive antennas and the *scalar* channel between any transmit-receive antenna pair contains 16 taps spaced $\frac{T}{KM}$ apart whose components are independent complex Gaussian random variables with zero mean and unit variance. The overall channel is a sixth order tensor $\mathcal{H}[n] \in \mathbb{C}_n^{K\times M\times P\times K\times M\times P}$ that couples the input with an output, which is another third order tensor $\mathcal{Y}[n] \in \mathbb{C}_n^{K\times M\times P}$. The channel is assumed to be known at the receiver and consists of 1 *tensor* tap. Further, due to the use of a cyclic prefix, the channel does not cause interference between successive GFDM symbols (i.e., no inter-tensor interference) but interference is caused within each GFDM symbol. The DFE used has $N_f = 1, N_b = 1$ and the decision delay $\Delta = 0$. The tensors $\bar{\mathcal{D}}[n] \in \mathbb{C}_n^{1\times K\times M\times P}$, $\bar{\mathcal{Y}}[n] \in \mathbb{C}_n^{1\times K\times M\times P}$ and $\bar{\mathcal{N}}[n] \in \mathbb{C}_n^{1\times K\times M\times P}$ have auto-correlation $\mathcal{R}_{\bar{\mathcal{D}}} = \mathcal{I}_4$, $\mathcal{R}_{\bar{\mathcal{Y}}} = \{\bar{\mathcal{H}}, \bar{\mathcal{H}}^H\}_{(4)} + N_0\mathcal{I}_4$ and $\mathcal{R}_{\bar{\mathcal{N}}} = N_0\mathcal{I}_4$ respectively. The cross correlation between $\bar{\mathcal{D}}[n]$ and $\bar{\mathcal{Y}}[n]$ is given by

$$\mathcal{R}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}} = \{\mathcal{R}_{\bar{\mathcal{D}}}, \bar{\mathcal{H}}^H\}_{(4)} = \bar{\mathcal{H}}^H \tag{4.206}$$

Using (4.206) and the value of $\mathcal{R}_{\bar{\mathcal{Y}}}$ we get the tensor

$$\mathcal{A} = \mathcal{I}_4 - \{\bar{\mathcal{H}}^H, \left(\{\bar{\mathcal{H}}, \bar{\mathcal{H}}^H\}_{(4)} + N_0\mathcal{I}_4\right)^{-1}, \bar{\mathcal{H}}\}_{(4)}\}_{(4)} \tag{4.207}$$

The feedback and feedforward filters are found by substituting (4.207) in (4.197) and then using the optimal feedback filter in (4.200). The mean squared error is numerically evaluated by using (4.207) in (4.201). The averaged receive signal to noise ratio is defined

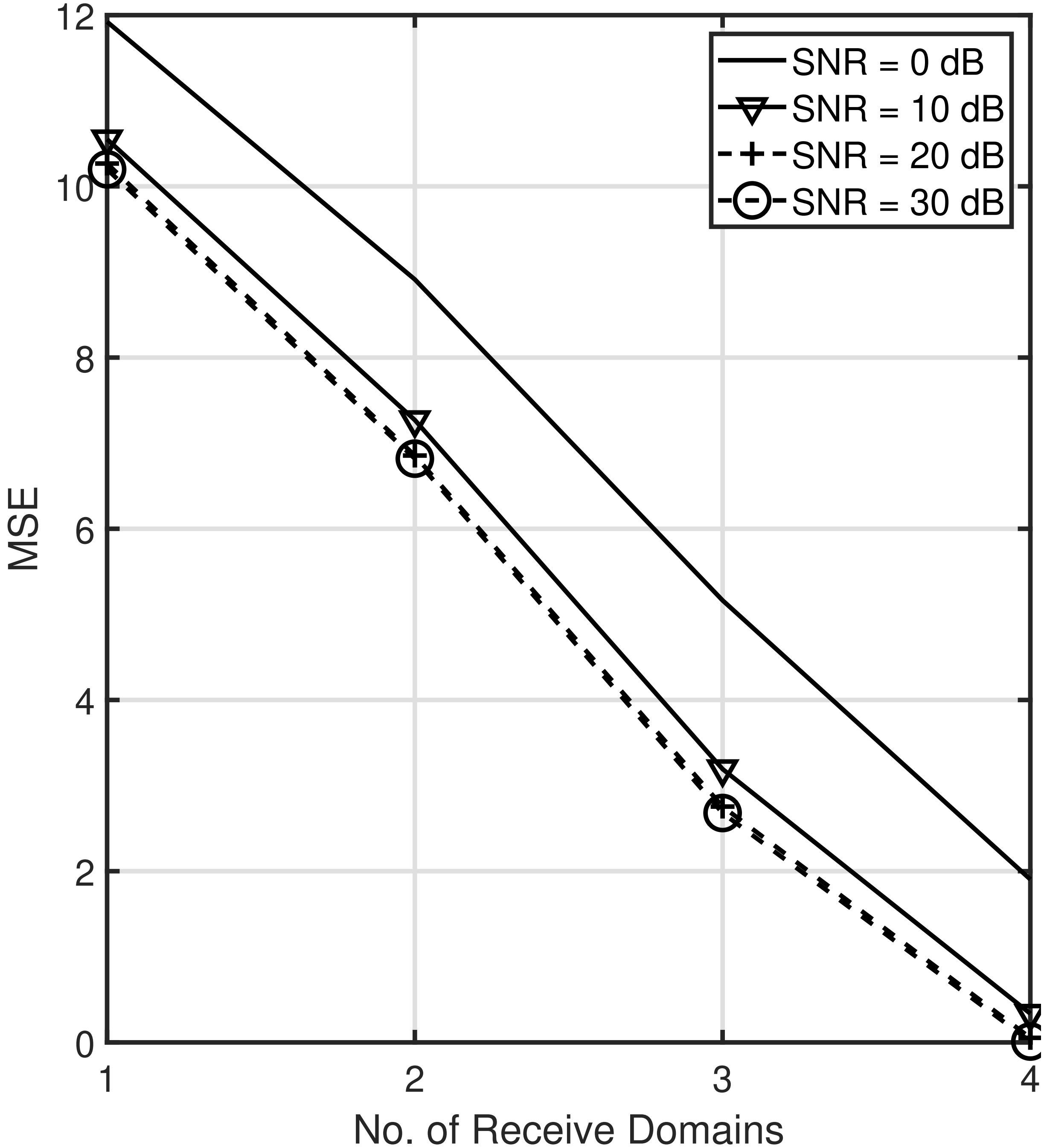


Fig. 4.13 MSE vs number of receive domains for different SNR

as the ratio of the energy per tensor symbol component E_s and the noise variance at the input to the equalizer where E_s is the total energy of the tensor divided by the number of components in the tensor. The other parameters used for the simulations are defined in Table 4.5

Table 4.5 Simulation parameters for MIMO GFDM

Description	parameter	value
No. of subcarriers	K	16
No. of time slots	M	5
pulse shaping filter	g	Raised Cosine (RC)
roll-off factor	α	1
No. of transmit,receive antennas	T, R	2,2
modulation order	μ	6 (64-QAM)

Fig. 4.14 shows the minimum MSE as a function of SNR. Results from 1000 different channel realizations were averaged to find the mean squared error using simulations. Fig. 4.15 shows the variation of the symbol error rate with SNR. The error rate is found by averaging MATLAB simulation results over 100 channel realizations, accumulating 250 errors at each SNR. Also shown in Figures 4.14 and 4.15 are the performance results for a Linear MMSE equalizer with the same number of equalizer taps as the feedforward length of the DF-equalizer and the same decision delay $\Delta = 0$. From Fig. 4.14 we can see that the decision feedback equalizer performs better than the linear equalizer and at higher SNRs, this gap in performance increases. This result is also seen for the error rate results of Fig. 4.15 as the gap in SER increases with SNR.

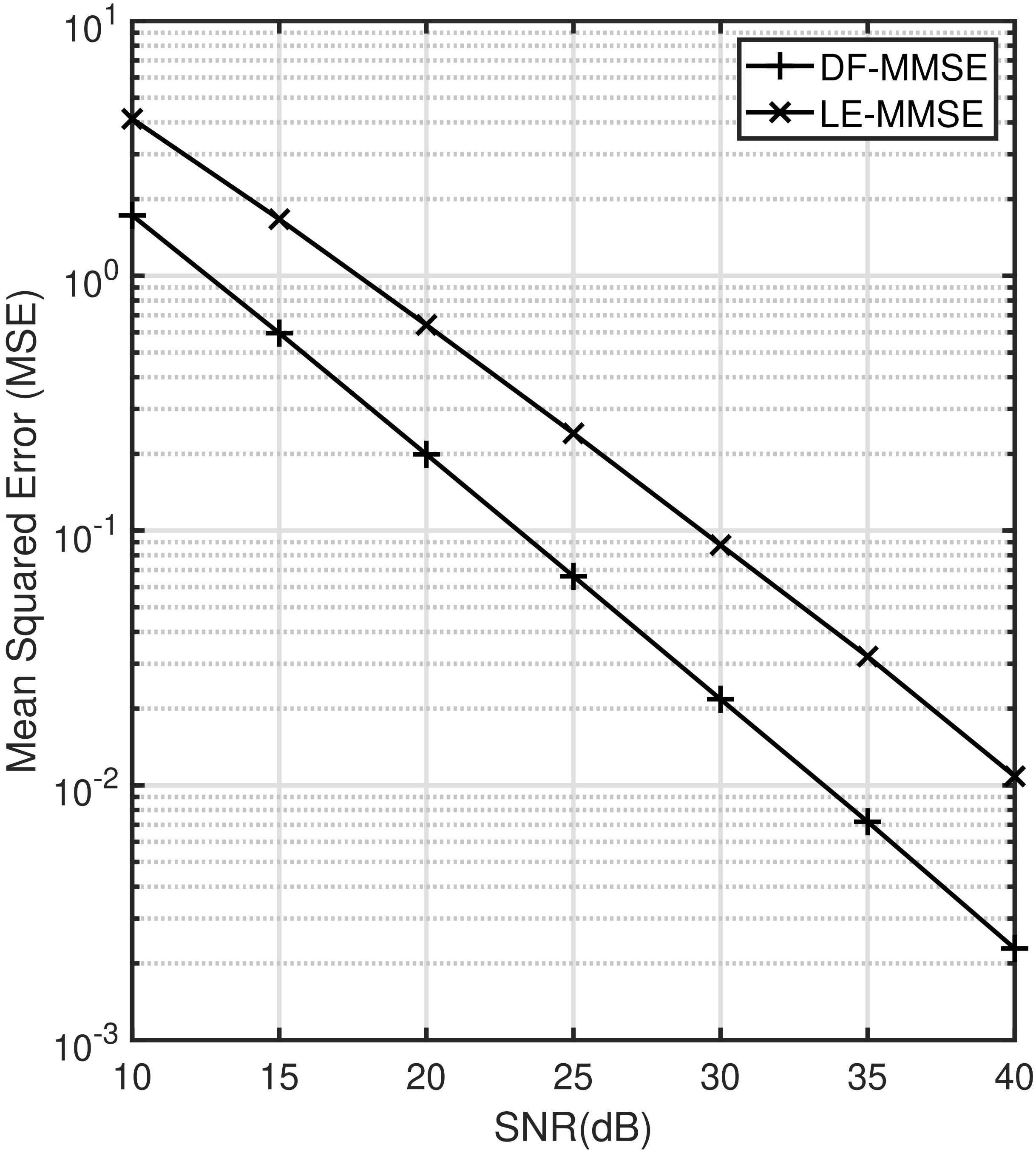


Fig. 4.14 MSE vs SNR for MIMO GFDM

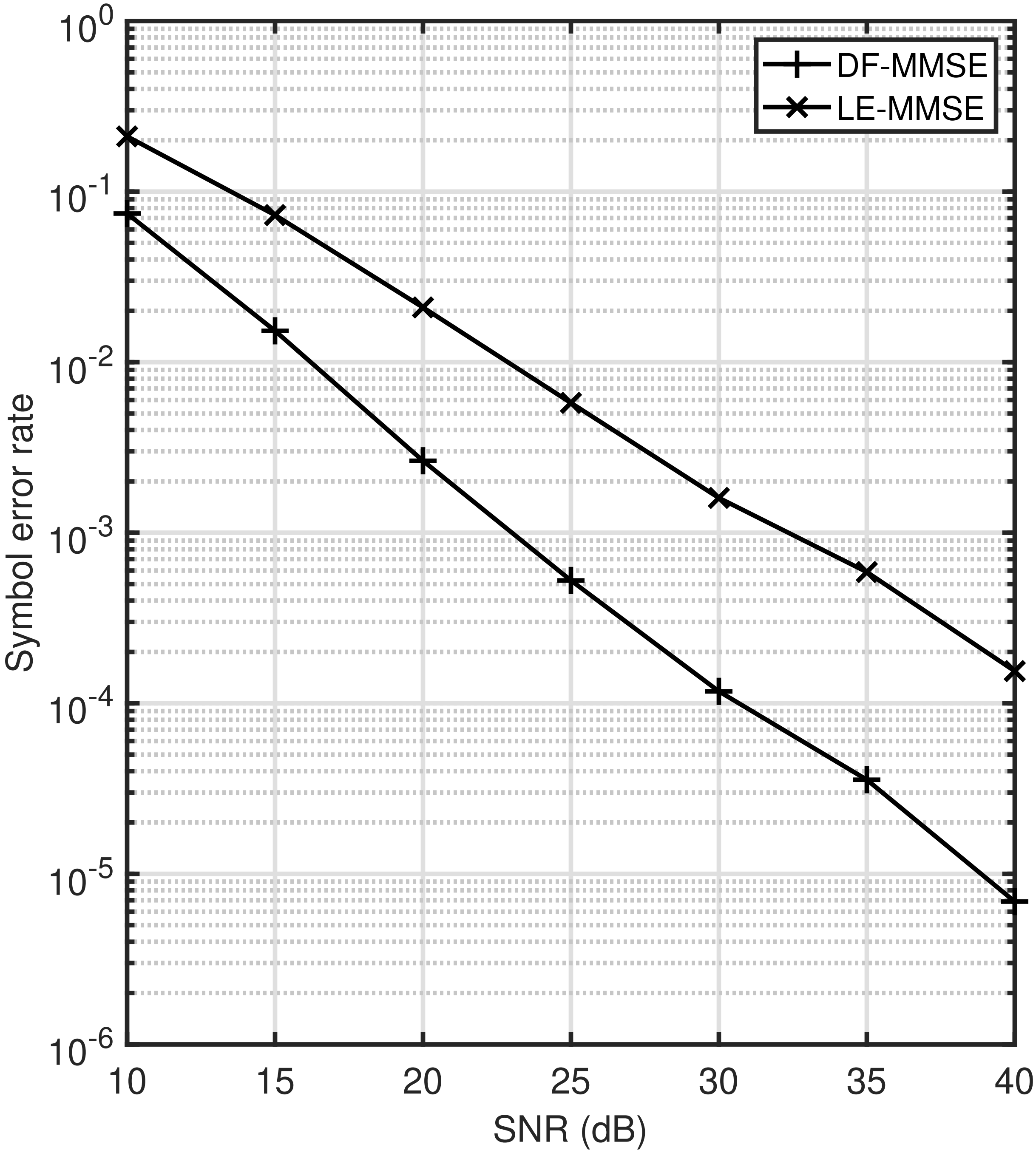


Fig. 4.15 SER vs SNR for MIMO GFDM

4.3.1 Infinite tap DFE

We consider the case where the receive system tensor $\boldsymbol{\mathcal{H}}_R(t)$ is matched to the combined channel and transmit system $\boldsymbol{\mathcal{C}}(t)$ such that $\boldsymbol{\mathcal{H}}_R(t) = \boldsymbol{\mathcal{C}}^H(-t)$. The overall channel is $\boldsymbol{\mathcal{H}}(t) = \{\boldsymbol{\mathcal{H}}_R(t) * \boldsymbol{\mathcal{C}}(t)\}_{(Q)} = \{\boldsymbol{\mathcal{C}}^H(-t) * \boldsymbol{\mathcal{C}}(t)\}_{(Q)}$. The input to the decision device is

$$\tilde{\boldsymbol{\mathcal{D}}}[k] = \sum_m \{\boldsymbol{\mathcal{W}}[m], \boldsymbol{\mathcal{Y}}[k-m]\}_{(N)} - \sum_j \{\boldsymbol{\mathcal{B}}[j], \hat{\boldsymbol{\mathcal{D}}}[k-j]\}_{(N)} \tag{4.208}$$

where $\hat{\boldsymbol{\mathcal{D}}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$ is the estimate of the data tensor $\boldsymbol{\mathcal{D}}[k] \in \mathbb{C}_k^{I_1 \times \ldots \times I_N}$. The D transform of (4.208) is

$$\breve{\tilde{\boldsymbol{\mathcal{D}}}}(D) = \{\breve{\boldsymbol{\mathcal{W}}}(D), \breve{\boldsymbol{\mathcal{Y}}}(D)\}_{(N)} - \{\breve{\boldsymbol{\mathcal{B}}}(D), \breve{\hat{\boldsymbol{\mathcal{D}}}}(D)\}_{(N)} \tag{4.209}$$

We assume that the components of $\tilde{\boldsymbol{\mathcal{D}}}[k]$ are estimated sequentially to produce $\hat{\boldsymbol{\mathcal{D}}}[k]$. Moreover, the decision $\hat{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]$ is aided by previously detected components of the same tensor. The ordering of the detection is as described in (4.37). We assume that previously detected symbols are correct. With these assumptions, the feedback system $\boldsymbol{\mathcal{B}}[k]$ thus has the form

$$\boldsymbol{\mathcal{B}}[k] = \boldsymbol{\mathcal{B}}[0] + \boldsymbol{\mathcal{B}}[1]D + \ldots \tag{4.210}$$

where $\boldsymbol{\mathcal{B}}[0]$ is a lower triangular tensor and $\boldsymbol{\mathcal{B}}[k]$ has an infinite number of taps. The equalizer described in this section is designed to select $\breve{\boldsymbol{\mathcal{W}}}(D)$ and $\breve{\boldsymbol{\mathcal{B}}}(D)$ such that the mean squared errors between the estimate and the transmitted data are minimised. The optimal MMSE-DFE feedforward and feedback systems, the complete derivation of which can be found in the appendix A.6, are given by

$$\breve{\boldsymbol{\mathcal{W}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{(d)}(0), \{(\breve{\boldsymbol{\mathcal{M}}}^{(d)}(0))^{-1}, \{\breve{\boldsymbol{\mathcal{M}}}^{-H}(D^{-1}), \breve{\boldsymbol{\mathcal{A}}}^{H}(D)\}_{(N)}\}_{(N)}\}_{(N)}\}_{(N)} \tag{4.211}$$

$$\breve{\boldsymbol{\mathcal{B}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{(d)}(0), \{(\breve{\boldsymbol{\mathcal{M}}}^{(d)}(0))^{-1}\{\breve{\boldsymbol{\mathcal{M}}}(D), \breve{\boldsymbol{\mathcal{A}}}^{-1}(D)\}_{(N)}\}_{(N)}\}_{(N)}\}_{(N)} - \boldsymbol{\mathcal{I}}_N \tag{4.212}$$

where $\breve{\boldsymbol{\mathcal{R}}}(D) = \{\{\breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1}), \breve{\boldsymbol{\mathcal{H}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathcal{A}}}(D)\}_{(N)} + N_0\boldsymbol{\mathcal{I}}_N$ with a spectral factorization $\breve{\boldsymbol{\mathcal{R}}}(D) = \{\breve{\boldsymbol{\mathcal{M}}}(D), \breve{\boldsymbol{\mathcal{M}}}^{H}(D^{-1})\}_{(N)}$ and $\breve{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\mathcal{D}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}(D), \breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1})\}_{(N)}$.

We define the $i_1, \ldots, i_N$th Mean squared error as

$$\begin{aligned}\epsilon_{i_1,\ldots,i_N} &= \mathbb{E}\big[\big|\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] - \tilde{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k]\big|^2\big] \\ &= \mathbb{E}\big[(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] - \tilde{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k])(\boldsymbol{\mathcal{D}}_{i_1,\ldots,i_N}[k] - \tilde{\boldsymbol{\mathcal{D}}}_{i_1,\ldots,i_N}[k])^*\big] \\ &= \boldsymbol{\mathcal{R}}_{(\boldsymbol{\mathcal{D}}-\tilde{\boldsymbol{\mathcal{D}}})_{i_1,\ldots i_N, i_1,\ldots,i_N}}[0]\end{aligned} \tag{4.213}$$

where $\boldsymbol{\mathfrak{R}}_{(\boldsymbol{\mathfrak{D}}-\tilde{\boldsymbol{\mathfrak{D}}})}[0]$ is the auto-correlation of the sequence $\boldsymbol{\mathfrak{D}}[k]-\tilde{\boldsymbol{\mathfrak{D}}}[k]$ at zero delay. Substituting for $\tilde{\breve{\boldsymbol{\mathfrak{D}}}}(D)$ from (4.209) and, assuming that past decisions are correct ($\breve{\boldsymbol{\mathfrak{D}}}(D) = \hat{\breve{\boldsymbol{\mathfrak{D}}}}(D)$), we get

$$\begin{aligned}\breve{\boldsymbol{\mathfrak{D}}}(D) - \tilde{\breve{\boldsymbol{\mathfrak{D}}}}(D) &= \breve{\boldsymbol{\mathfrak{D}}}(D) - (\{\breve{\boldsymbol{\mathfrak{W}}}(D), \breve{\boldsymbol{\mathfrak{Y}}}(D)\}_{(N)} - \{\breve{\boldsymbol{\mathfrak{B}}}(D), \hat{\breve{\boldsymbol{\mathfrak{D}}}}(D)\}_{(N)}) \\ &= \{\big(\boldsymbol{\mathfrak{I}}_N - \breve{\boldsymbol{\mathfrak{W}}}(D) + \breve{\boldsymbol{\mathfrak{B}}}(D)\big), \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)} - \{\breve{\boldsymbol{\mathfrak{W}}}(D), \breve{\boldsymbol{\mathfrak{V}}}(D)\}_{(N)} \\ &= \{\breve{\boldsymbol{\mathfrak{F}}}(D), \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)} + \breve{\boldsymbol{\mathfrak{Z}}}(D) \end{aligned} \tag{4.214}$$

where we have defined $\breve{\boldsymbol{\mathfrak{F}}}(D) = \boldsymbol{\mathfrak{I}}_N - \breve{\boldsymbol{\mathfrak{W}}}(D) + \breve{\boldsymbol{\mathfrak{B}}}(D)$ and $\breve{\boldsymbol{\mathfrak{Z}}}(D) = -\{\breve{\boldsymbol{\mathfrak{W}}}(D), \breve{\boldsymbol{\mathfrak{V}}}(D)\}_{(N)}$. Using (4.21) we can write the spectrum of $\breve{\boldsymbol{\mathfrak{Z}}}(D)$ as

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{Z}}}(D) = \{\{\breve{\boldsymbol{\mathfrak{W}}}(D), \breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{V}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathfrak{W}}}(D)\}_{(N)} = N_0\{\{\breve{\boldsymbol{\mathfrak{W}}}(D), \breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{V}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathfrak{W}}}(D)\}_{(N)} \tag{4.215}$$

The spectrum of (4.214) can be written as

$$\breve{\boldsymbol{\mathfrak{S}}}_{(\boldsymbol{\mathfrak{D}}-\tilde{\boldsymbol{\mathfrak{D}}})}(D) = \{\{\breve{\boldsymbol{\mathfrak{F}}}(D), \breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathfrak{F}}}(D)\}_{(N)} + N_0\{\{\breve{\boldsymbol{\mathfrak{W}}}(D), \breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{V}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathfrak{W}}}(D)\}_{(N)} \tag{4.216}$$

The $i_1, \ldots, i_N$th Mean squared error can be calculated by integrating the $i_1, \ldots, i_N$th pseudo-diagonal of $\breve{\boldsymbol{\mathfrak{S}}}_{(\boldsymbol{\mathfrak{D}}-\tilde{\boldsymbol{\mathfrak{D}}})}(D)$ over the unit-circle by setting $D = e^{j2\pi f}$. i.e.,

$$\int_0^1 \breve{\boldsymbol{\mathfrak{S}}}_{(\boldsymbol{\mathfrak{D}}-\tilde{\boldsymbol{\mathfrak{D}}})_{i_1,\ldots i_N, i_1,\ldots,i_N}}(e^{j2\pi f})df = \boldsymbol{\mathfrak{R}}_{(\boldsymbol{\mathfrak{D}}-\tilde{\boldsymbol{\mathfrak{D}}})_{i_1,\ldots i_N, i_1,\ldots,i_N}}[0] = \epsilon_{i_1,\ldots,i_N} \tag{4.217}$$

and the overall MSE can be calculated as

$$MSE = \frac{1}{I_1 \cdot I_2 \ldots I_N} \sum_{i_1,\ldots,i_N} \epsilon_{i_1,\ldots,i_N} \tag{4.218}$$

A comparison between the linear MMSE equalier and the decision feedback MMSE equalizer is plotted in figures 4.16 and 4.17. The input $\breve{\boldsymbol{\mathfrak{D}}}(D) \in \mathbb{C}_D^{2\times 2}$ has a spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{D}}} = \boldsymbol{\mathfrak{I}}_N$. The equivalent channel $\breve{\boldsymbol{\mathfrak{H}}}(D)$ consists of two taps $\boldsymbol{\mathfrak{H}}[0]$ and $\boldsymbol{\mathfrak{H}}[1]$ whose components are Gaussian random variables. In Fig. 4.16, MSE is plotted against SNR for three specific realizations of the channel and Fig. 4.17 shows the MSE results averaged over 100 channel realizations. The individual mean squared errors $\epsilon_{i_1,\ldots,i_N}$ are calculated using (4.217) and the overall MSE is calculated using (4.218). The additive noise is has a spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{N}}}(D) = N_0\boldsymbol{\mathfrak{I}}_N$. For this case we have $\breve{\boldsymbol{\mathcal{A}}}(D) = \boldsymbol{\mathfrak{I}}_N$ and $\breve{\boldsymbol{\mathfrak{R}}}(D) = \breve{\boldsymbol{\mathfrak{H}}}(D)\}_{(N)} + N_0\boldsymbol{\mathfrak{I}}_N$. The feedforward equalizer from (4.211) becomes

$$\breve{\boldsymbol{\mathfrak{W}}}(D) = \{(\breve{\boldsymbol{\mathfrak{M}}}^{(d)}(0))^{-1}, \breve{\boldsymbol{\mathfrak{M}}}^{-H}(D^{-1})\}_{(N)} \tag{4.219}$$

and the feedback filter becomes

$$\breve{\boldsymbol{\mathcal{B}}}(D) = \{(\breve{\boldsymbol{\mathcal{M}}}^{(d)}(0))^{-1}, \breve{\boldsymbol{\mathcal{M}}}(D)\}_{(N)} - \boldsymbol{\mathcal{I}}_N \tag{4.220}$$

The averaged receive signal to noise ratio is defined as the ratio of the energy per tensor symbol component E_s and the noise variance at the input to the equalizer where E_s is the total energy of the tensor divided by the number of components in the tensor. The performance of the linear equalizer is worse than the decision feedback equalizer. Further, as the SNR increases, the difference in the performance between the two equalizer also increases. Although the results for the three channel realizations are different, the gap between the linear equalizer and the DFE increases with SNR for all the three cases. This is consistent with the results from Fig. 4.14 where the finite tap decision feedback equalizer was compared with the linear MMSE equalizer for MIMO GFDM.

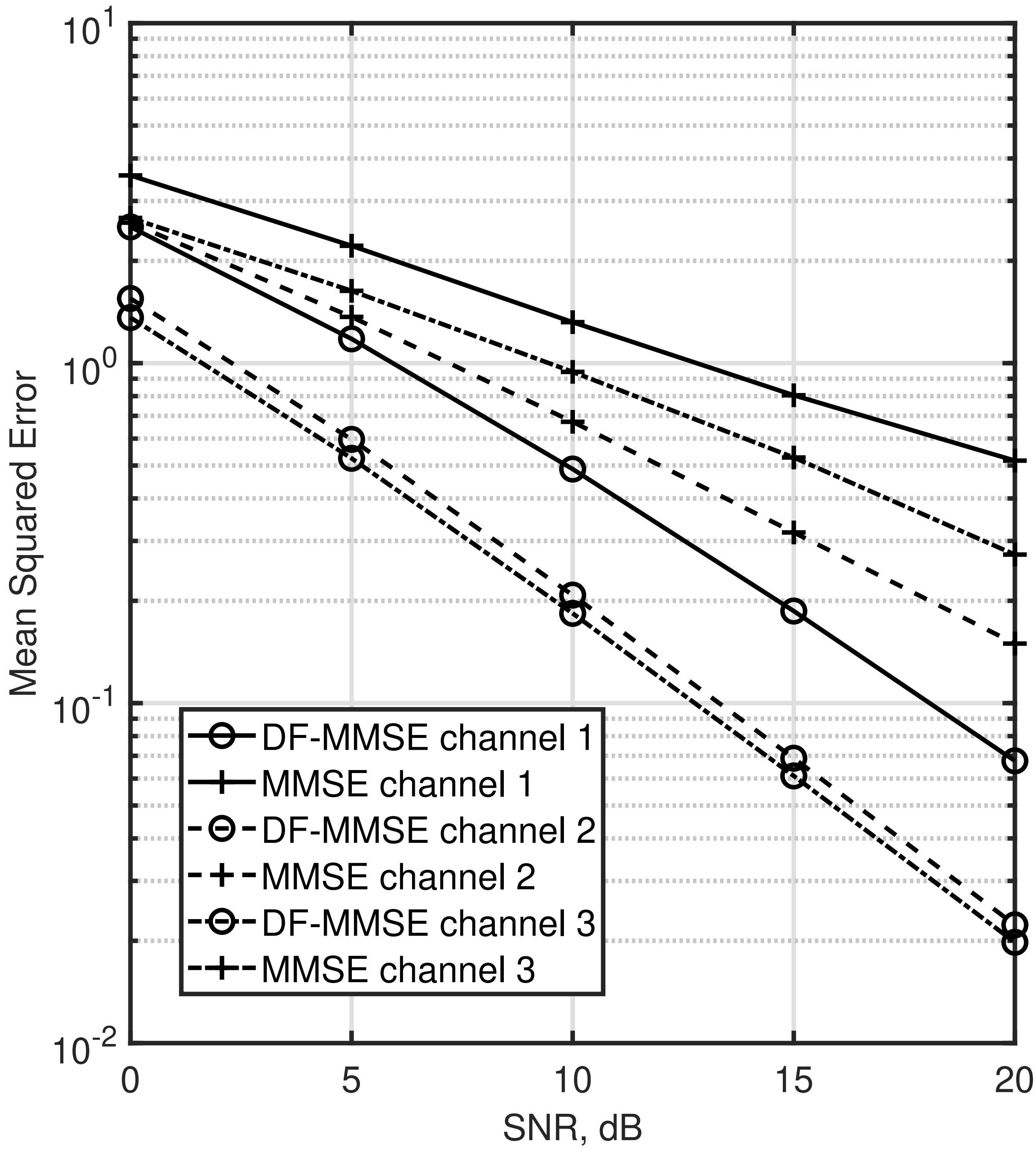


Fig. 4.16 MSE vs SNR for specific channels for DFE-MMSE and MMSE equalizers

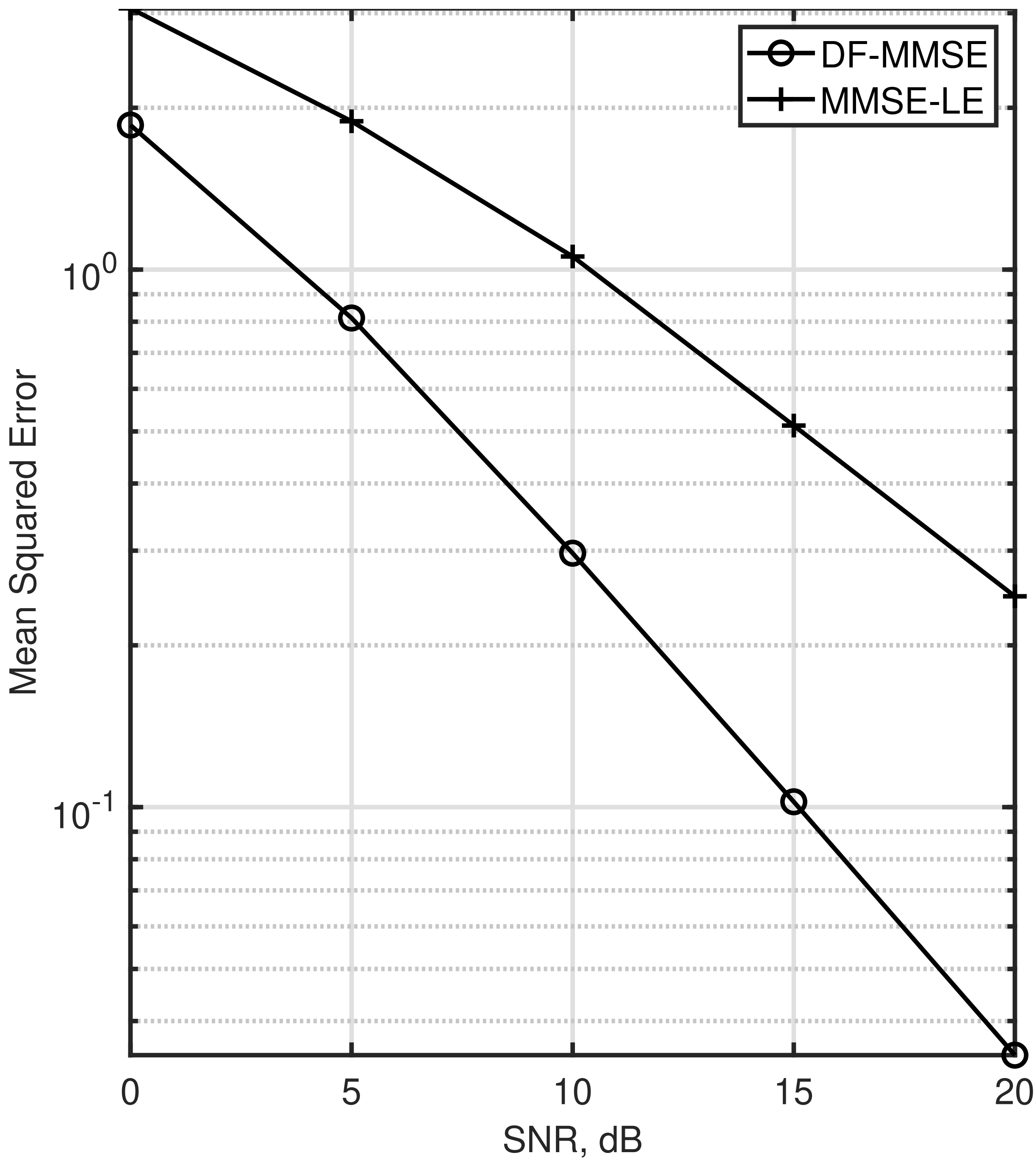


Fig. 4.17 MSE vs SNR averaged over 100 channel realizations for DFE-MMSE and MMSE equalizers

Chapter 5

Tensor Correlative Coding

5.1 Introduction

Classical Nyquist signaling schemes completely eliminate inter-symbol interference by design. Correlative Coding, otherwise known as Partial Response Signaling (PRS), is a transmission method correlation is introduced between successive transmitted symbols by allowing a controlled amount of inter-symbol interference. The objective is to shape the spectrum of the transmitted signal by using correlative codes to achieve desirable properties.

The concept of correlative coding dates back to the 1960s. Lender [64] describes duobinary partial response signaling as a transmission method. Several different correlative codes were introduced and categorized based on different characteristics such as speed tolerance and SNR degradation by Kretzmer [65]. Pasupathy et al. [36, 37] propose a general PRS framework where the waveform generation is divided into two parts.
In this chapter we describe a method to allow controlled interference in multiple domains using the tensor framework. Such a scheme will be called Tensor Partial Response Signaling (TPRS) or tensor correlative coding.

5.2 System Model

Consider a complex linear time invariant system $\boldsymbol{\mathcal{H}}(t) \in \mathbb{C}_t^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$. Let the input to this system be a sequence of uncorrelated data tensors, spaced at intervals of T, $\sum_n \boldsymbol{\mathcal{D}}[n]\delta(t-nT) \in \mathbb{C}_T^{I_1\times\ldots\times I_N}$ with a spectrum $\breve{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\mathcal{D}}}(D) \in \mathbb{C}_D^{I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N}$. The output of this system is

$$\boldsymbol{\mathcal{X}}(t) = \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(t-nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)} \tag{5.1}$$

which, when sampled at intervals of T gives

$$\boldsymbol{\mathcal{X}}[k] = \boldsymbol{\mathcal{X}}(kT) = \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}(kT-nT), \boldsymbol{\mathcal{D}}[n]\}_{(N)} = \sum_{n=-\infty}^{+\infty} \{\boldsymbol{\mathcal{H}}[k-n], \boldsymbol{\mathcal{D}}[n]\}_{(N)} \tag{5.2}$$

with components

$$\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_N}[k] = \sum_{n=0}^{N-1} \sum_{i'_1,\ldots,i'_N} \boldsymbol{\mathcal{F}}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}[n] \boldsymbol{\mathcal{D}}_{i'_1,\ldots,i'_N}[k-n] \tag{5.3}$$

In the D domain, (5.2) becomes

$$\breve{\boldsymbol{\mathcal{X}}}(D) = \{\breve{\boldsymbol{\mathcal{H}}}(D), \breve{\boldsymbol{\mathcal{D}}}(D)\}_{(N)} \tag{5.4}$$

Let the system $\boldsymbol{\mathcal{H}}(t)$ be a cascade of two systems $\boldsymbol{\mathcal{F}}(t)$ and $\boldsymbol{\mathcal{G}}(t)$. The system $\boldsymbol{\mathcal{F}}(t)$ is a tensor tapped delay line with N taps, such that $\boldsymbol{\mathcal{F}}(t) = \sum_{n=0}^{N-1} \boldsymbol{\mathcal{F}}[n]\delta(t-nT)$. The D-transform of the tensor tapped delay line $\boldsymbol{\mathcal{F}}[n]$ is

$$\breve{\boldsymbol{\mathcal{F}}}(D) = \sum_{n=0}^{N-1} \boldsymbol{\mathcal{F}}[n] D^n \tag{5.5}$$

Proposition 1. If the system $\boldsymbol{\mathcal{G}}(t)$ follows the strict tensor Nyquist criterion then the samples of $\boldsymbol{\mathcal{H}}(t)$ are $\boldsymbol{\mathcal{F}}[n]$. i.e.,

$$\boldsymbol{\mathcal{H}}(nT) = \boldsymbol{\mathcal{F}}[n] \tag{5.6}$$

Proof. The tensor tapped delay line may be denoted as a sum of N impulses as:

$$\boldsymbol{\mathcal{F}}(t) = \sum_{n=0}^{N-1} \boldsymbol{\mathcal{F}}[n]\delta(t-nT) \tag{5.7}$$

The system tensor $\boldsymbol{\mathcal{H}}(t)$ is

$$\boldsymbol{\mathcal{H}}(t) = \{\boldsymbol{\mathcal{G}}(t) * \boldsymbol{\mathcal{F}}(t)\}_{(N)}$$

$$= \{\boldsymbol{\mathfrak{G}}(t) * \left(\sum_{n=0}^{N-1} \boldsymbol{\mathfrak{F}}[n]\delta(t - nT)\right)\}_{(N)}$$
$$= \sum_{n=0}^{N-1} \{\boldsymbol{\mathfrak{G}}(t - nT), \boldsymbol{\mathfrak{F}}[n]\}_{(N)} \tag{5.8}$$

Since $\boldsymbol{\mathfrak{G}}(t)$ follows the strict Tensor Nyquist Criterion, sampling (5.8) at intervals of T we have

$$\boldsymbol{\mathfrak{H}}(kT) = \sum_{n=0}^{N-1} \{\boldsymbol{\mathfrak{G}}(kT - nT), \boldsymbol{\mathfrak{F}}[n]\}_{(N)}$$
$$= \boldsymbol{\mathfrak{F}}[k] \tag{5.9}$$

□

The D-transorm of $\boldsymbol{\mathfrak{H}}[k]$ is $\breve{\boldsymbol{\mathfrak{H}}}(D) = \sum_i \boldsymbol{\mathfrak{H}}[k]D^i = \boldsymbol{\mathfrak{F}}(D)$. The relation (5.4) thus becomes $\breve{\boldsymbol{\mathfrak{X}}}(D) = \{\breve{\boldsymbol{\mathfrak{H}}}(D), \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)}$. Using (4.21), the relation between the spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D)$ and the system $\breve{\boldsymbol{\mathfrak{F}}}(D)$ is

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D) = \{\breve{\boldsymbol{\mathfrak{F}}}(D), \{\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{D}}}(D), \breve{\boldsymbol{\mathfrak{F}}}^H(D^{-1})\}_{(N)}\}_{(N)} = \{\breve{\boldsymbol{\mathfrak{F}}}(D), \breve{\boldsymbol{\mathfrak{F}}}^H(D^{-1})\}_{(N)} \tag{5.10}$$

where the last step is because the data tensor $\boldsymbol{\mathfrak{D}}[k]$ has spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{D}}}(D) = \boldsymbol{\mathfrak{I}}_N$. This means that the spectrum of $\breve{\boldsymbol{\mathfrak{X}}}(D)$ depends only on the tapped delay line $\breve{\boldsymbol{\mathfrak{F}}}(D)$. The TPRS system thus has two parts. The tapped delay line $\breve{\boldsymbol{\mathfrak{F}}}(D)$ that is used to shape the spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D)$ and the system $\boldsymbol{\mathfrak{G}}(t)$ which is used to band-limit the resulting system function $\boldsymbol{\mathfrak{H}}(t)$. For a given $\breve{\boldsymbol{\mathfrak{F}}}(D)$, different choices of $\boldsymbol{\mathfrak{G}}(t)$ result in different overall system functions $\boldsymbol{\mathfrak{H}}(t)$. However, as long as $\boldsymbol{\mathfrak{G}}(t)$ meets the generalized Nyquist criterion, $\boldsymbol{\mathfrak{H}}[k] = \boldsymbol{\mathfrak{F}}[k]$. Fig. 5.1 shows the system model of the TPRS system.

5.3 Tensor Correlative Codes

5.3.1 Structure of the TPRS polynomial

The cross-spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D)$ can be controlled by changing the structure of the TPRS system $\breve{\boldsymbol{\mathfrak{F}}}(D)$. In its most general form, the TPRS system introduces correlation among all the components of the data tensor and also the components of previous data tensors. A compo-

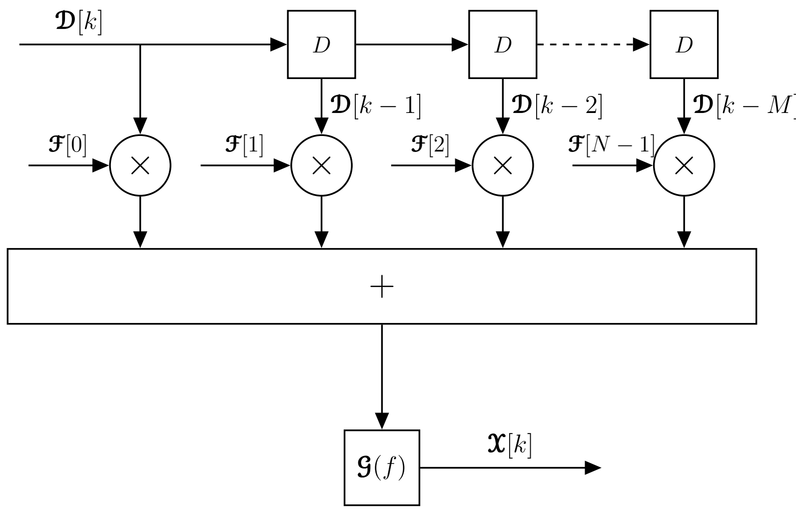


Fig. 5.1 TPRS System Model

nent $\mathfrak{X}_{i_1,\ldots,i_N}[k]$ hence contains controlled intra-tensor interference from $\mathfrak{D}_{i'_1,\ldots,i'_N}[k]$ as well as inter-tensor interference from $N-1$ preceding data tensors $\mathfrak{D}[k-1],\ldots,\mathfrak{D}[k-N+1]$. By using different TPRS systems $\breve{\mathfrak{F}}(D)$ and imposing restrictions such as, for example, only allowing intra-tensor interference in the output $\mathfrak{X}[k]$, the shape of the spectrum $\breve{\mathfrak{S}}_{\mathfrak{X}}(D)$ can be manipulated. Denote the frequency response of the spectrum by $\breve{\mathfrak{S}}_{\mathfrak{X}}(\omega) = \breve{\mathfrak{S}}_{\mathfrak{X}}(D)|_{D=e^{j\omega T}}$. Using this, we can write the components of the frequency response as

$$\breve{\mathfrak{S}}_{\mathfrak{X}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}}(\omega) = \sum_{j_1,\ldots,j_N} \breve{\mathfrak{F}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}(D)\breve{\mathfrak{F}}^H_{j_1,\ldots,j_N,i'_1,\ldots,i'_N}(D^{-1})\Big|_{D=e^{j\omega T}} \tag{5.11}$$

Spectral nulls at different values of ω are desirable for specific applications. To create a spectral null at a frequency ω' in component $(i_1,\ldots,i_N,i'_1,\ldots,i'_N)$ of $\breve{\mathfrak{S}}(\omega)$, we require that $\sum_{j_1,\ldots,j_N} \breve{\mathfrak{F}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}(D)\breve{\mathfrak{F}}^H_{j_1,\ldots,j_N,i'_1,\ldots,i'_N}(D^{-1})|_{D=e^{j\omega' T}} = 0$. In this section, we present different TPRS systems and illustrate them using the specific case of a fourth order TPRS polynomial $\breve{\mathfrak{F}}(D) \in \mathbb{C}^{2\times2\times2\times2}$ whose input is an uncorrelated sequence $\breve{\mathfrak{D}}(D) \in \mathbb{C}_D^{2\times2}$ and output is $\breve{\mathfrak{X}}(D) \in \mathbb{C}_D^{2\times2}$ with the desired spectrum $\breve{\mathfrak{S}}_{\mathfrak{X}}(D) \in \mathbb{C}_D^{2\times2\times2\times2}$.

Class1: Degenerate TPRS

The first class of TPRS systems that we consider is one where $\boldsymbol{\mathcal{F}}[k]$ containts only one non-zero tap $\boldsymbol{\mathcal{F}}[0]$. i.e., $\boldsymbol{\mathcal{F}}[k] = \mathbf{0}_T$ for $k \neq 0$. This implies that the components within a tensor $\boldsymbol{\mathcal{X}}[k]$ are correlated but successive tensors are uncorrelated. The tensor $\boldsymbol{\mathcal{X}}[k] = \{\boldsymbol{\mathcal{F}}[0], \boldsymbol{\mathcal{D}}[k]\}_{(N)}$ has components

$$\mathcal{X}_{i_1,\ldots,i_N}[k] = \sum_{i'_1,\ldots,i'_N} \mathcal{F}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}[0]\mathcal{D}_{i'_1,\ldots,i'_N}[k] \tag{5.12}$$

The D transform of $\boldsymbol{\mathcal{F}}[k]$ is $\breve{\boldsymbol{\mathcal{F}}}(D) = \boldsymbol{\mathcal{F}}[0]$ and the spectrum $\breve{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\mathcal{X}}}(D)$ has components

$$\breve{\mathcal{S}}_{\mathcal{X}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}}(D) = \sum_{j_1,\ldots,j_N} \mathcal{F}_{i_1,\ldots,i_N,j_1,\ldots,j_N}\mathcal{F}^H_{j_1,\ldots,j_N,i'_1,\ldots,i'_N} \tag{5.13}$$

This means that the spectrum does not depend on D and the frequency response $\breve{\boldsymbol{\mathcal{S}}}(w)$ is flat. Further, by changing the structure of $\boldsymbol{\mathcal{F}}[0]$, the correlation introduced may be restricted to certain domains. We illustrate this with two example TPRS systems. Figures 5.2 and 5.3 show the spectrum of $\breve{\boldsymbol{\mathcal{X}}}(D)$ and the components of the TPRS system used are listed in Table.5.1. The code used in Fig. 5.2 is such that each component of $\boldsymbol{\mathcal{X}}[k]$ is a linear combination of all the components of $\boldsymbol{\mathcal{D}}[k]$. In Fig. 5.3, the code used restricts the correlation to a single domain. Here, the component $\mathcal{X}_{i_1,i_2}[k] = \sum_{j_2=1}^{2} \mathcal{F}_{i_1,i_2,i_1,j_2}[0]\mathcal{D}_{i_1,j_2}[k]$ and the TPRS code is

$$\mathcal{F}_{i_1,i_2,j_1,j_2}[0] = \begin{cases} 0 & \text{if } i_1 \neq j_1 \\ 1 & \text{if } i_1 = j_1, i_2 = j_2 \\ 0.5 & \text{otherwise} \end{cases} \tag{5.14}$$

An example of restricting the correlation to a single domain is when frequency domain correlation is introduced in a MIMO multi-carrier transmission system. Such a code has been shown to suppress inter-carrier interference caused by Doppler frequency shift in MIMO OFDM systems [66]. Such a system has a TPRS system as described by the code used for Fig. 5.3 when there are two antennas and two sub-carriers.

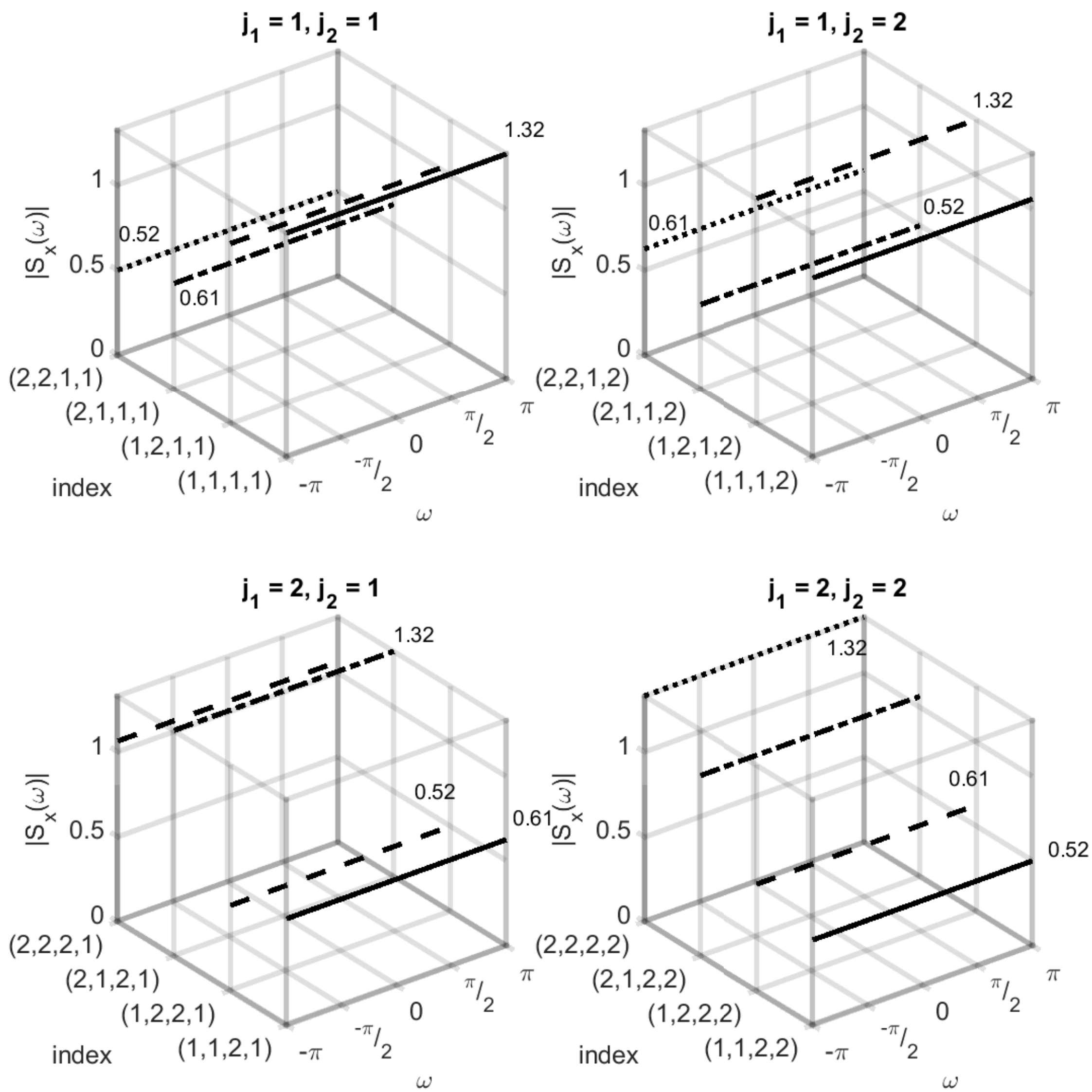


Fig. 5.2 Spectrum for no inter-tensor interference, case 1

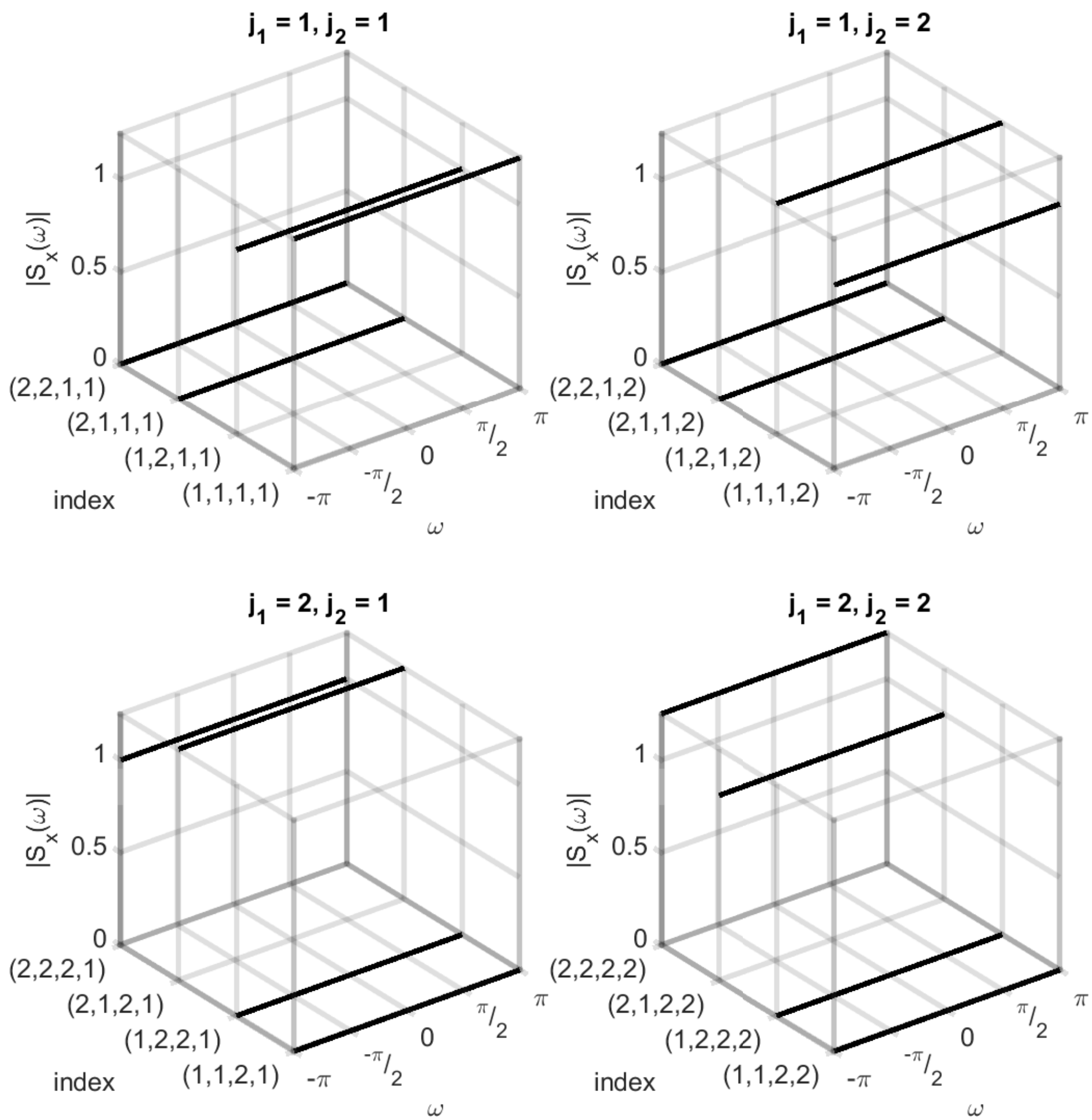


Fig. 5.3 Spectrum for no inter-tensor interference, case 2

Table 5.1 Structure of the Class 1 TPRS System

Component	Fig. 5.2	Fig. 5.3
$\mathfrak{F}_{1,1,1,1}$	1	1
$\mathfrak{F}_{1,1,1,2}$	0.5	0.5
$\mathfrak{F}_{1,1,2,1}$	0.25	0
$\mathfrak{F}_{1,1,2,2}$	0.125	0
$\mathfrak{F}_{1,2,1,1}$	0.5	0.5
$\mathfrak{F}_{1,2,1,2}$	1	1
$\mathfrak{F}_{1,2,2,1}$	0.125	0
$\mathfrak{F}_{1,2,2,2}$	0.25	0
$\mathfrak{F}_{2,1,1,1}$	0.25	0
$\mathfrak{F}_{2,1,1,2}$	0.125	0
$\mathfrak{F}_{2,1,2,1}$	1	1
$\mathfrak{F}_{2,1,2,2}$	0.5	0.5
$\mathfrak{F}_{2,2,1,1}$	0.125	0
$\mathfrak{F}_{2,2,1,2}$	0.25	0
$\mathfrak{F}_{2,2,2,1}$	0.5	0.5
$\mathfrak{F}_{2,2,2,2}$	1	1

Class 2: Pseudo-diagonal TPRS

In the second class of TPRS systems considered, $\breve{\mathfrak{F}}(D)$ is pseudo-diagonal. All the taps $\mathfrak{F}[1], \ldots, \mathfrak{F}[N-1]$ are hence pseudo-diagonal tensors. A component $\mathfrak{X}_{i_1,\ldots,i_N}[k]$ only contains interference from components of $\mathfrak{D}[k-1], \ldots, \mathfrak{D}[k-N+1]$. This means that the correlation introduced is restricted to the same component in successive data tensors. The components of $\mathfrak{X}[k]$ are

$$\mathfrak{X}_{i_1,\ldots,i_N}[k] = \sum_{n=0}^{N-1} \mathfrak{F}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[n] \mathfrak{D}_{i_1,\ldots,i_N}[k-n] \tag{5.15}$$

For such a system, the spectrum of $\breve{\mathfrak{X}}(D)$ from (5.10) has components

$$\breve{\mathfrak{S}}_{\mathfrak{X}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}}(D) = \sum_{j_1,\ldots,j_N} \breve{\mathfrak{F}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}(D) \breve{\mathfrak{F}}^H_{j_1,\ldots,j_N,i'_1,\ldots,i'_N}(D^{-1}) \tag{5.16}$$

We know that $\boldsymbol{\mathfrak{F}}_{i_1,\dots,i_N,j_1,\dots j_N}(D) = 0$ for $i_1 \neq j_1, \dots, i_N \neq j_N$ since $\boldsymbol{\mathfrak{F}}(D)$ is a pseudo-diagonal tensor. The components of the spectrum from (5.16) hence become

$$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}_{i_1,\dots,i_N,i'_1,\dots,i'_N}}(D) = \begin{cases} \breve{\boldsymbol{\mathfrak{F}}}_{i_1,\dots,i_N,i_1,\dots,i_N}(D)\breve{\boldsymbol{\mathfrak{F}}}^*_{i_1,\dots,i_N,i_1,\dots,i_N}(D^{-1}) & \text{if } i_1 = i'_1, \dots, i_N = i'_N \\ 0 \quad \text{otherwise} \end{cases} \tag{5.17}$$

Equation (5.17) implies that $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D)$ is a pseudo-diagonal tensor. This means that the cross-spectrum of $\boldsymbol{\mathfrak{X}}_{i_1,\dots,i_N}(D)$ and $\boldsymbol{\mathfrak{X}}_{i'_1,\dots,i'_N}(D)$ is 0. Further, the spectrum of the individual components $\boldsymbol{\mathfrak{X}}_{i_1,\dots,i_N}(D)$ depends only on $\breve{\boldsymbol{\mathfrak{F}}}_{i_1,\dots,i_N,i_1,\dots,i_N}(D)$ and can be shaped by varying it. For example, consider a fourth order pseudo-diagonal TPRS system $\breve{\boldsymbol{\mathfrak{F}}}(D)$ with pseudo-diagonal components

$$\breve{\boldsymbol{\mathfrak{F}}}_{1,1,1,1}(D) = 1 + D \tag{5.18}$$

$$\breve{\boldsymbol{\mathfrak{F}}}_{2,1,2,1}(D) = 1 - D^2 \tag{5.19}$$

$$\breve{\boldsymbol{\mathfrak{F}}}_{1,2,1,2}(D) = (1 + D)^2 \tag{5.20}$$

$$\breve{\boldsymbol{\mathfrak{F}}}_{2,2,2,2}(D) = 1 - D^4 \tag{5.21}$$

Let the output of this system be $\breve{\boldsymbol{\mathfrak{X}}}(D) \in \mathbb{C}^{2\times 2}$. Since $\breve{\boldsymbol{\mathfrak{F}}}(D)$ is a pseudo-diagonal tensor, the spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(D)$ is also pseudo-diagonal. Table. 5.2 summarizes the pseudo-diagonal components of this TPRS system and the corresponding pseudo-diagonal components of the spectrum. Denote the frequency response of the spectrum by $\breve{\boldsymbol{\mathfrak{S}}}(\omega) = \breve{\boldsymbol{\mathfrak{S}}}(D)\big|_{D=e^{-j\omega T}}$. Shown in Fig. 5.4 are the components $|\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}}(\omega)|$ for $T = 1$. The components of $\breve{\boldsymbol{\mathfrak{F}}}(D)$ were

Table 5.2 Structure of the Class 2 TPRS system

Component (i_1, i_2, j_1, j_2)	$\breve{\boldsymbol{\mathfrak{F}}}_{i_1,i_2,j_1,j_2}(D)$	$\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}_{i_1,i_2,j_1,j_2}}(D)$
$1, 1, 1, 1$	$1 + D$	$(1 + D)(1 + \frac{1}{D^*})$
$1, 2, 1, 2$	$1 - D^2$	$(1 - D^2)(1 - (\frac{1}{D*})^2)$
$2, 1, 2, 1$	$(1 + D)^2$	$(1 + D)^2(1 + \frac{1}{D*})^2$
$2, 2, 2, 2$	$1 - D^4$	$(1 - D^4)(1 - (\frac{1}{D*})^4)$

chosen such that all the pseudo-diagonal components of the spectrum contain nulls at $\pm\pi$. We can see that the pseudo-diagonal elements $|\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}_{1,1,1,1}}(\omega)|$ and $|\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{X}}_{2,1,2,1}}(\omega)|$ have different roll-off rates. This is because $\breve{\mathfrak{F}}_{2,1,2,1}(D)$ contains two zeros at $\omega = \pm\pi$. $|\breve{\mathfrak{S}}_{\mathfrak{X}_{1,2,1,2}}(\omega)|$ and $|\breve{\mathfrak{S}}_{\mathfrak{X}_{2,2,2,2}}(\omega)|$ contain additional spectral nulls at $\omega = 0$ and $\omega = 0, \pm\frac{\pi}{2}$ respectively.

Class3: TPRS

As we have seen, class 1 systems can be used to manipulate the level of the spectrum and cross-spectrum components of $\breve{\mathfrak{S}}_{\mathfrak{X}}(D)$ and class 2 systems can be used to shape the pseudo-diagonal components of $\breve{\mathfrak{S}}_{\mathfrak{X}}(D)$. The third class of TPRS systems is the most general and there are no restrictions on the structure of the taps $\mathfrak{F}[k]$. Such a TPRS system can be used to simultaneously shape the spectrum of $\breve{\mathfrak{X}}_{i_1,\ldots,i_N}(D)$ and the cross spectrum of $\breve{\mathfrak{X}}_{i_1,\ldots,i_N}(D)$ and $\breve{\mathfrak{X}}_{i'_1,\ldots,i'_N}(D)$. The frequency response of $\breve{\mathfrak{S}}_{\mathfrak{X}}(D)$ is

$$\breve{\mathfrak{S}}_{\mathfrak{X}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}}(\omega) = \sum_{j_1,\ldots,j_N} \breve{\mathfrak{F}}_{i_1,\ldots,i_N,j_1,\ldots,j_N}(D)\breve{\mathfrak{F}}^{H}_{j_1,\ldots,j_N,i'_1,\ldots,i'_N}(D^{-1})\Big|_{D=e^{j\omega T}} \tag{5.22}$$

As an example, the spectrum of the output of a TPRS system $\breve{\mathfrak{F}}(D) \in \mathbb{C}^{2\times2\times2\times2}$ is shown in Fig. 5.5. The components of $\breve{\mathfrak{S}}_{\mathfrak{X}}(\omega)$ that correspond to the cross-spectrum are represented by dashed lines and the components that correspond to the spectrum are represented by solid lines. The system used in this case is designed such that the spectrum of $\mathfrak{X}_{1,1,1,1}[k]$ has a null at $\omega = \pm\pi$, the spectrum of $\mathfrak{X}_{1,2,1,2}[k]$ has a null at $\omega = 0$, the spectrum of $\mathfrak{X}_{2,1,2,1}[k]$ has spectral nulls at $\omega = \pm\pi, 0$ and the spectrum of $\mathfrak{X}_{2,2,2,2}[k]$ has spectral nulls at $\omega = 0, \pm\frac{\pi}{2}, \pm\pi$.

It is important to note that different systems can be designed to have the same spectral terms but different cross-spectrum terms. To this end, shown in Fig. 5.6 is the output of a different TPRS system $\breve{\mathfrak{F}}_2(D)$. As we can see, the pseudo-diagonal elements of $\breve{\mathfrak{S}}_{\mathfrak{X}}(D)$ have the same frequency response for both the cases but the other components (which correspond to the cross-spectrum) are vastly different. This means that it is possible to manipulate the cross-spectrum without having to compromise on the shape of the spectrum itself. Such a feature is useful in situations where different applications require the same spectral shape but different cross-spectra. The TPRS systems used in figures 5.5 and 5.6 are detailed in Table. 5.3.

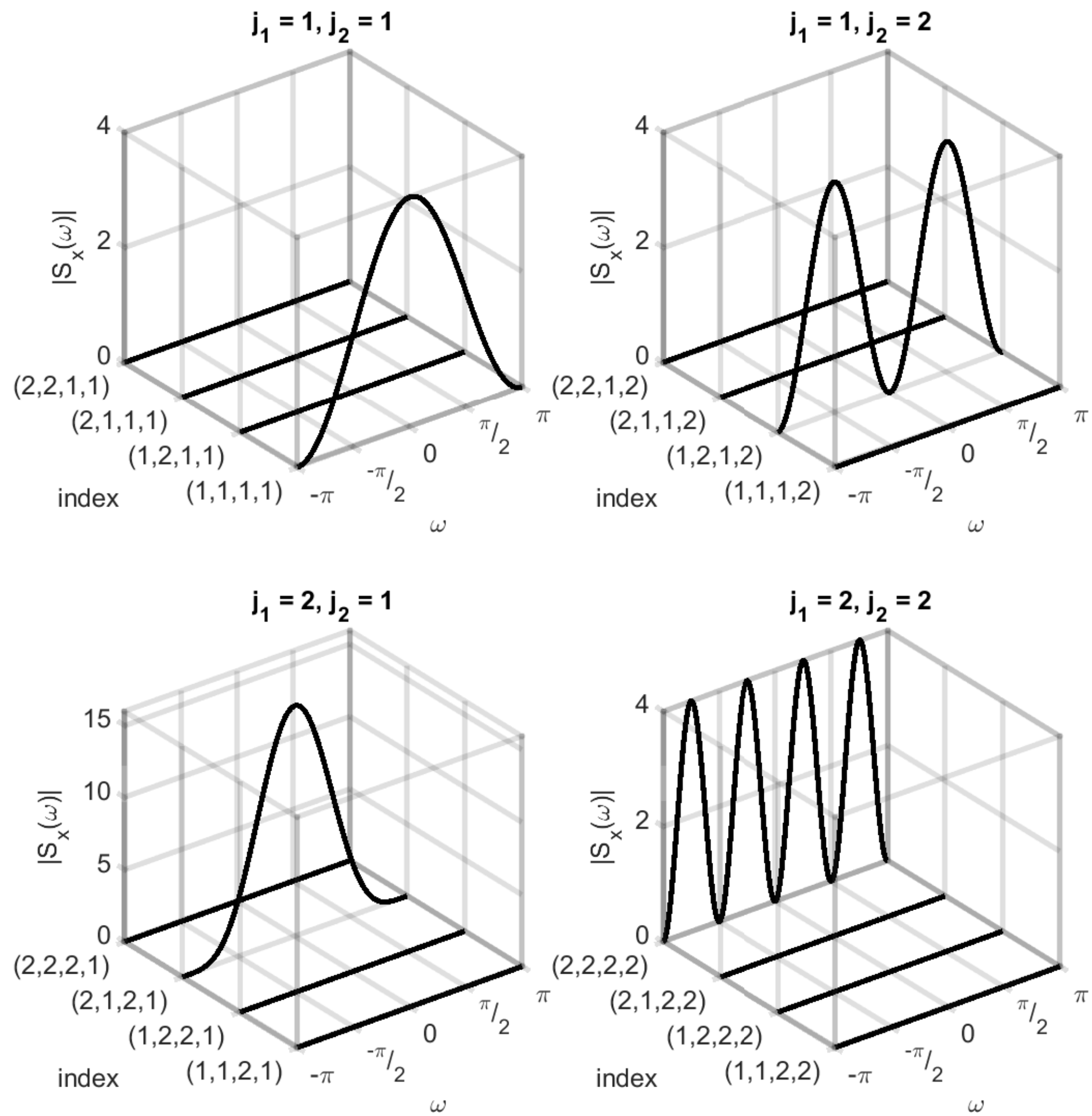


Fig. 5.4 Spectrum of $\breve{\mathfrak{X}}$ for pseuo-diagonal $\mathfrak{F}(\breve{D})$

Table 5.3 Structure of the Class 3 TPRS System

Component	Fig. 5.5	Fig. 5.6
$\mathfrak{F}_{1,1,1,1}$	$1+D$	$1+D$
$\mathfrak{F}_{1,1,1,2}$	0	$0.5(1+D)$
$\mathfrak{F}_{1,1,2,1}$	$0.5(1+D)$	0
$\mathfrak{F}_{1,1,2,2}$	0	0
$\mathfrak{F}_{1,2,1,1}$	$0.5(1-D)$	$0.5(1-D)$
$\mathfrak{F}_{1,2,1,2}$	$(1-D)$	$(1-D)$
$\mathfrak{F}_{1,2,2,1}$	0.125	0
$\mathfrak{F}_{1,2,2,2}$	0.25	0
$\mathfrak{F}_{2,1,1,1}$	0.25	0
$\mathfrak{F}_{2,1,1,2}$	0.125	0
$\mathfrak{F}_{2,1,2,1}$	$(1-D^2)$	$(1-D^2)$
$\mathfrak{F}_{2,1,2,2}$	$0.5(1-D^2)$	$0.5(1-D^2)$
$\mathfrak{F}_{2,2,1,1}$	0	0
$\mathfrak{F}_{2,2,1,2}$	$0.5(1-D^4)$	0
$\mathfrak{F}_{2,2,2,1}$	0	$0.5(1-D^4)$
$\mathfrak{F}_{2,2,2,2}$	$(1-D^4)$	$(1-D^4)$

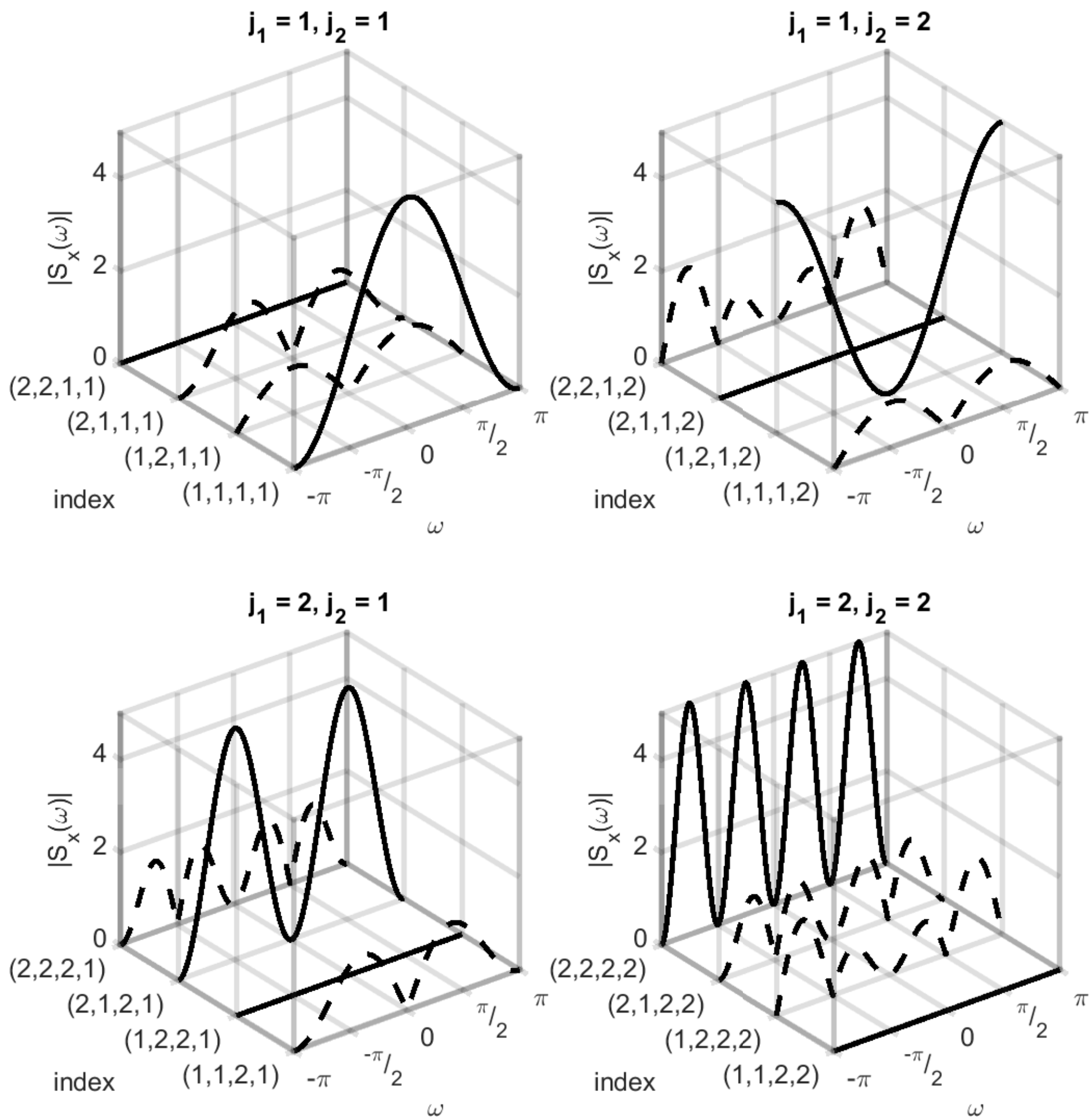


Fig. 5.5 Spectrum and Cross-spectrum components of $\breve{\mathfrak{s}}_{\mathfrak{X}}(D)$

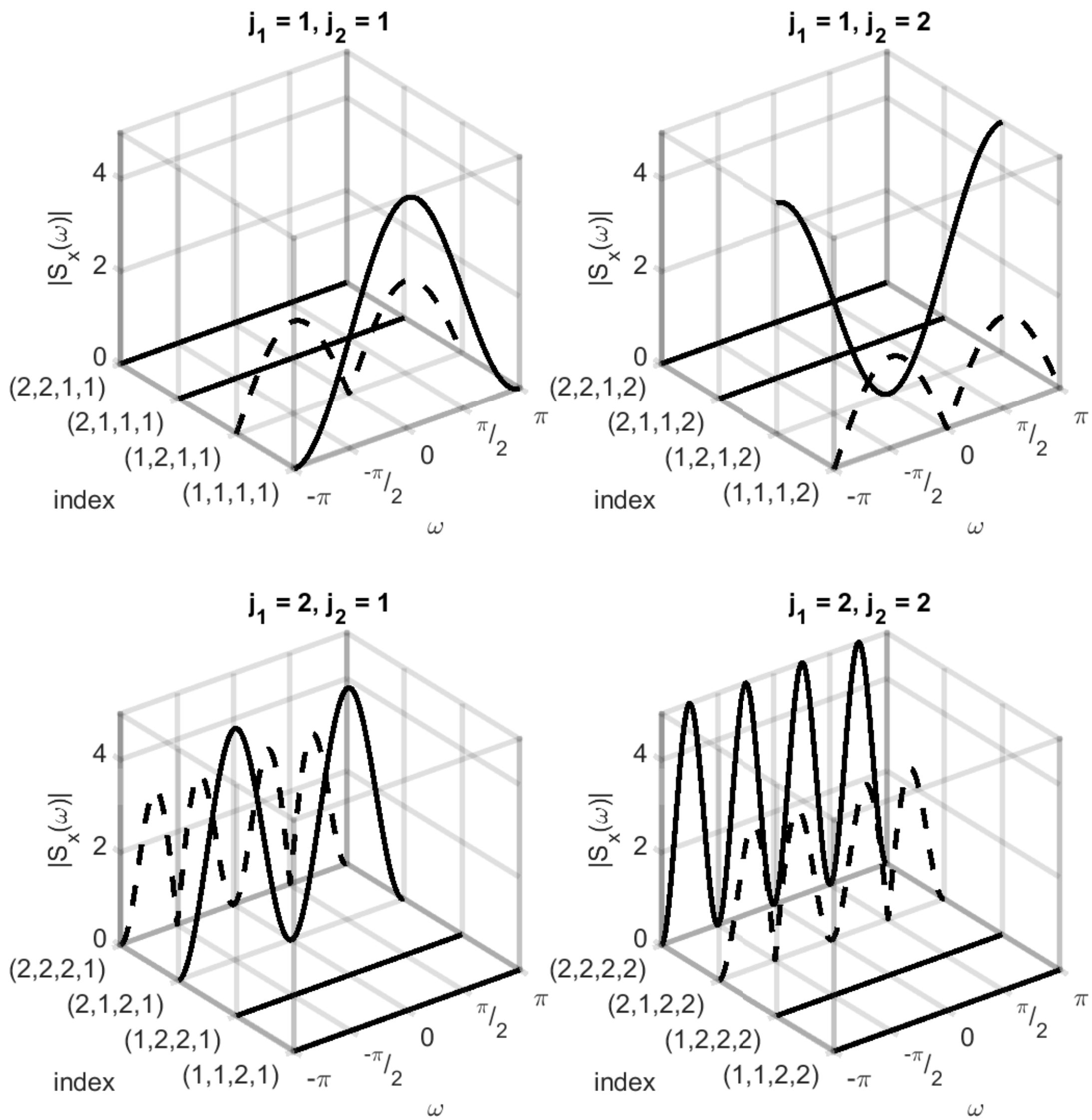


Fig. 5.6 Modified TPRS system that preserves the spectrum while changing the cross-spectrum

Chapter 6

Conclusion

This thesis presented a unified tensor framework, which can be used to represent, design and analyse communication systems that span several domains. No distinctions have been assigned to the domains of the systems and the general framework presented can be used for a myriad of communication systems. The transmitted signals are represented by Nth order signal tensors which are coupled, using a system tensor of order $N + M$, with the received signals which are represented by another signal tensor of order M through the contracted convolution. The notion of a tensor of functions forms the basis for the definition of signal and system tensors. A generalization of the Nyquist's criterion for zero inter symbol interference was derived which allows unifying treatment of interference from several domains, dubbed multi domain interference (MDI). It was shown that for the tensor case, a relaxation of the Nyquist Criterion is possible that allows recovery of data symbols even in the presence of intra-tensor interference. The tensor framework was used to model existing systems such as OFDM, GFDM and FBMC. Using the tensor framework, an example higher domain extension for GFDM and FBMC was derived where different filters are used at the analysis and synthesis filter banks of each antenna. These examples demonstrate the utility of our tensor framework.

Further, linear and non-linear tensor based equalizers were derived for different criterion such as peak distortion and minimum mean squared error. The performance of the minimum mean squared linear equalizer was found to be better than that of the zero forcing equalizer, which is consistent with the scalar case. For the linear MMSE, the mean squared error decreases as the number of domains at the receiver are increased while keeping the size of each domain the same. This implies that domain diversity can be leveraged to improve performance in cases where there are restrictions of the size of any particular domain, as is often the case in practice due to bandwidth limitations (frequency domain) or restrictions on the number of antennas (space domain). The performance of the Decision Feedback Minimum Mean Squared (DF-MMSE) equalizer is better than the linear MMSE equalizer. Further, the mean squared error decreases as the number of domains at the receiver are increased while keeping the size of each domain the same. Finally, the notion of partial response signaling was extended to multi-domain systems in the form of Tensor Partial Response Signaling (TPRS). Here, a multi-domain tensor tapped delay line is used to shape the spectrum of the transmitted signal to achieve desirable spectral properties. It was found that controlled interference from within the same tensor (intra-tensor interference) changes the level of the spectrum while maintaining a flat frequency response. Controlled interference from successive data tensors allows the placement of spectral nulls. Combining the two allows for manipulation of both the shape and the level of the spectrum of the transmitted signal. Moreover, it was found that the cross-spectrum of the components of the transmitted signal can be shaped independently of the spectrum and different TPRS polynomials can be designed that result in the same spectrum but different cross-spectrum.

Appendix A

Some Proofs

A.1 Proof of Theorem 2

Consider an input $\mathcal{X}(t) = \sum_{n=-\infty}^{+\infty} \mathcal{X}[n]\delta(t-nT) \in \mathbb{C}_t^{I_1\times\ldots\times I_N}$ to a system tensor $\mathcal{A}(t) \in \mathbb{C}_t^{J_1\times\ldots\times J_M\times I_1\times\ldots\times I_N}$. The output of this filter, corrupted by additive white Gaussian noise, is $\mathcal{R}(t) = \{\mathcal{A}(t)*\mathcal{X}(t)\}_{(N)}+\mathcal{N}(t)$ where $\mathcal{N}(t)$ is a tensor whose components are white Gaussian noise processes. Let $\mathcal{R}(t)$ be the input to a system tensor $\mathcal{B}(t) \in \mathbb{C}_t^{I_1\times\ldots\times I_N\times J_1\times\ldots\times J_M}$ with an output $\mathcal{Y}(t) \in \mathbb{C}_t^{I_1\times\ldots\times I_N}$. The per component SNR of the samples $\mathcal{Y}(kT) = \mathcal{Y}[k]$ is maximized when $\mathcal{B}(t) = \mathcal{A}^H(-t)$.

Proof. The tensor $\mathcal{Y}(t)$ can be written as

$$\begin{aligned}\mathcal{Y}(t) &= \{\mathcal{B}(t) * \{\mathcal{A}(t) * \mathcal{X}(t)\}_{(N)}\}_{(M)} + \{\mathcal{B}(t) * \mathcal{N}(t)\}_{(M)}\\ &= \sum_{n=-\infty}^{+\infty} \{\mathcal{C}(t-nT), \mathcal{X}[n]\}_{(N)} + \mathcal{V}(t) \end{aligned} \tag{A.1}$$

where $\mathcal{C}(t) = \{\mathcal{B}(t) * \mathcal{A}(t)\}_{(M)}$ and $\mathcal{V}(t) = \{\mathcal{B}(t) * \mathcal{N}(t)\}_{(M)}$. Sampling at intervals of kT gives

$$\mathcal{Y}[k] = \sum_{n=-\infty}^{+\infty} \{\mathcal{C}[k-n], \mathcal{X}[n]\}_{(N)} + \mathcal{V}[k] \tag{A.2}$$

with components

$$\mathcal{Y}_{i_1,\ldots,i_N}[k] = \mathcal{C}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[0]\mathcal{X}_{i_1,\ldots,i_N}[k] + \mathcal{I}_{i_1,\ldots,i_N}[k] + \mathcal{V}_{i_1,\ldots,i_N}[k] \tag{A.3}$$

where

$$\mathbf{\mathfrak{I}}_{i_1,\ldots,i_N}[k] = \sum_{i'_1,\ldots,i'_N \neq i_1,\ldots,i_N} \mathbf{\mathfrak{C}}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}[0]\mathbf{\mathfrak{X}}_{i'_1,\ldots,i'_N}[k] + \sum_{\substack{n=-\infty \\ n\neq k}}^{+\infty} \sum_{i'_1,\ldots,i'_N} \mathbf{\mathfrak{C}}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}[k-n]\mathbf{\mathfrak{X}}_{i'_1,\ldots,i'_N}[n] \tag{A.4}$$

We define the per component SNR $\gamma_{i_1,\ldots,i_N}$ at the sampled output of the system $\mathbf{\mathfrak{B}}(t)$ as the ratio of the power of the desired symbol $\mathbf{\mathfrak{X}}_{i_1,\ldots,i_N}[k]$ and the power of the noise $\mathbf{\mathfrak{V}}_{i_1,\ldots,i_N}[k]$. The intra-tensor and inter-tensor interference is contained in $\mathbf{\mathfrak{I}}_{i_1,\ldots,i_N}[k]$ and is not considered in this definition of the SNR. We have

$$\gamma_{i_1,\ldots,i_N} = \frac{\left|\mathbf{\mathfrak{C}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[0]\right|^2 E_{i_1,\ldots,i_N}}{\mathbb{E}[\mathbf{\mathfrak{V}}_{i_1,\ldots,i_N}[k]\mathbf{\mathfrak{V}}^*_{i_1,\ldots,i_N}[k]]} \tag{A.5}$$

where $E_{i_1,\ldots,i_N} = \mathbb{E}[\mathbf{\mathfrak{X}}_{i_1,\ldots,i_N}\mathbf{\mathfrak{X}}^*_{i_1,\ldots,i_N}]$. We may expand $\left|\mathbf{\mathfrak{C}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[0]\right|^2$ as

$$\left|\mathbf{\mathfrak{C}}_{i_1,\ldots,i_N,i_1,\ldots,i_N}[0]\right|^2 = \Big| \sum_{j_1,\ldots,j_M} \int_{-\infty}^{+\infty} \mathbf{\mathfrak{B}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}(t)\mathbf{\mathfrak{A}}_{j_1,\ldots,j_M,i_1,\ldots,i_N}(-t)dt\Big|^2 \tag{A.6}$$

The denominator of (A.5) can be expanded, using $\mathbb{E}[N_{j_1,\ldots,j_M}(t)N_{j'_1,\ldots,j'_M}(p)] = N_0\delta(t-p)$ when $(j_1,\ldots,j_M) = (j'_1,\ldots,j'_M)$ and 0 otherwise, as

$$\begin{aligned}
&\mathbb{E}[\mathbf{\mathfrak{V}}_{i_1,\ldots,i_N}[k]\mathbf{\mathfrak{V}}^*_{i_1,\ldots,i_N}[k]] \\
&= \mathbb{E}[\sum_{j_1\ldots j_M}\int_{-\infty}^{+\infty} \mathbf{\mathfrak{B}}_{i_1\ldots i_N j_1\ldots j_M}(t)\mathbf{\mathfrak{N}}_{j_1\ldots j_M}(kT-t)dt \sum_{j'_1\ldots j'_M}\int_{-\infty}^{+\infty} \mathbf{\mathfrak{B}}^*_{i_1\ldots i_N,j'_1\ldots j'_M}(p)\mathbf{\mathfrak{N}}^*_{j'_1\ldots j'_M}(kT-p)dp] \\
&= N_0 \sum_{j_1,\ldots,j_M}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathbf{\mathfrak{B}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}(t)\mathbf{\mathfrak{B}}^*_{i_1,\ldots,i_N,j_1,\ldots,j_M}(p)\delta(t-p)dtdp \\
&= N_0 \sum_{j_1,\ldots,j_M}\int_{-\infty}^{+\infty} |\mathbf{\mathfrak{B}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}(t)|^2 dt
\end{aligned} \tag{A.7}$$

Using (A.6) and (A.7) in (A.5) we get

$$\gamma_{i_1,\ldots,i_N} = \frac{\Big|\sum_{j_1,\ldots,j_M}\int_{-\infty}^{+\infty} \mathbf{\mathfrak{B}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}(t)\mathbf{\mathfrak{A}}_{j_1,\ldots,j_M,i_1,\ldots,i_N}(-t)dt\Big|^2 E_{i_1,\ldots,i_N}}{N_0\sum_{j_1,\ldots,j_M}\int_{-\infty}^{+\infty} |\mathbf{\mathfrak{B}}_{i_1,\ldots,i_N,j_1,\ldots,j_M}(t)|^2 dt} \tag{A.8}$$

Using the Cauchy-Schwartz inequality from Appendix A.3 this becomes

$$\gamma_{i_1,\ldots,i_N} \leq \frac{(\sum\limits_{j_1\ldots j_M}\int\limits_{-\infty}^{+\infty}\left|\mathcal{A}_{i_1\ldots i_N j_1\ldots j_M}(t)\right|^2 dt \sum\limits_{j_1\ldots j_M}\int\limits_{-\infty}^{+\infty}\left|\mathcal{B}_{j_1\ldots j_M k_1\ldots k_P}(t)\right|^2 dt)E_{i_1,\ldots,i_N}}{N_0\sum\limits_{j_1,\ldots,j_M}\int_{-\infty}^{+\infty}\left|\mathcal{B}_{i_1,\ldots,i_N,j_1,\ldots,j_M}(t)\right|^2 dt}$$

$$= \frac{\sum\limits_{j_1\ldots j_M}\int\limits_{-\infty}^{+\infty}\left|\mathcal{A}_{i_1\ldots i_N j_1\ldots j_M}(t)\right|^2 dt E_{i_1,\ldots,i_N}}{N_0} \tag{A.9}$$

with equality when $\mathcal{B}_{i_1,\ldots,i_N,j_1,\ldots,j_M}(t) = \mathcal{A}^*_{j_1,\ldots,j_M,i_1,\ldots,i_N}(-t)$. This means that the SNR attains its maximum value when $\mathcal{B}(t) = \mathcal{A}^H(-t)$. □

A.2 Proof of Theorem 3

Consider the general system where a data sequence $\mathcal{D}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N}$ is the input to a channel $\mathcal{H}[k] \in \mathbb{C}_k^{L_1\times\ldots\times L_M\times I_1\times\ldots\times I_N}$ and is corrupted by additive noise $\mathcal{V}[k] \in \mathbb{C}_k^{L_1\times\ldots\times L_M}$. The observation $\mathcal{Y}[k] \in \mathbb{C}_k^{L_1\times\ldots\times L_M}$ is

$$\mathcal{Y}[k] = \sum_{m=-\infty}^{+\infty}\{\mathcal{H}[m], \mathcal{D}[k-m]\}_{(N)} + \mathcal{V}[k] \tag{A.10}$$

The estimate of the data sequence is

$$\hat{\mathcal{D}}[k] = \sum_{m=0}^{+\infty}\{\mathcal{G}[m], \mathcal{Y}[k-m]\}_{(M)} \tag{A.11}$$

where $\mathcal{G}[m] \in \mathbb{C}^{I_1\times\ldots\times I_N\times L_1\times\ldots\times L_M}$. Denote the error by $\mathcal{E}[k] = \hat{\mathcal{D}}[k] - \mathcal{D}[k]$ We wish to prove that for the $\mathcal{G}[m]$ that minimizes the mean squared error between the estimate and the data, the error is uncorrelated with the observation. i.e., we wish to show that if

$$\mathbb{E}\left[\mathcal{E}[k] \circ \mathcal{Y}^*[k-i]\right] = \mathbf{0}_T \quad \text{for all } i \tag{A.12}$$

then the error is minimized in the mean squared sense. Consider the general system where a data sequence $\mathcal{D}[k] \in \mathbb{C}_k^{I_1\times\ldots\times I_N}$ is the input to a channel $\mathcal{H}[k] \in \mathbb{C}_k^{L_1\times\ldots\times L_M\times I_1\times\ldots\times I_N}$ and is corrupted by additive noise $\mathcal{V}[k] \in \mathbb{C}_k^{L_1\times\ldots\times L_M}$. The observation $\mathcal{Y}[k] \in \mathbb{C}_k^{L_1\times\ldots\times L_M}$ is

$$\mathcal{Y}[k] = \sum_{m=-\infty}^{+\infty}\{\mathcal{H}[m], \mathcal{D}[k-m]\}_{(N)} + \mathcal{V}[k] \tag{A.13}$$

The estimate of the data sequence is

$$\hat{\boldsymbol{\mathcal{D}}}[k] = \sum_{m=0}^{+\infty} \{\boldsymbol{\mathcal{G}}[m], \boldsymbol{\mathcal{Y}}[k-m]\}_{(M)} \tag{A.14}$$

where $\boldsymbol{\mathcal{G}}[m] \in \mathbb{C}^{I_1 \times \ldots \times I_N \times L_1 \times \ldots \times L_M}$. Denote the error by $\boldsymbol{\mathcal{E}}[k] = \hat{\boldsymbol{\mathcal{D}}}[k] - \boldsymbol{\mathcal{D}}[k]$ We wish to prove that for the $\boldsymbol{\mathcal{G}}[m]$ that minimizes the mean squared error between the estimate and the data, the error is uncorrelated with the observation. i.e., we wish to show that if

$$\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{Y}}^*[k-i]\right] = \mathbf{0}_T \quad \text{for all } i \tag{A.15}$$

then the error is minimized in the mean squared sense.

Proof. Assuming that the equalizer co-efficients are complex, the equalizer tensor may be written as $\boldsymbol{\mathcal{G}}[m] = \boldsymbol{\mathcal{A}}[m] + j\boldsymbol{\mathcal{B}}[m]$. Extending the gradient vector in [63], we define a corresponding tensor gradient operator ∇, with components

$$\nabla_{i_1,\ldots,i_N,m,l_1,\ldots,l_M} = \frac{\partial}{\partial \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_N,l_1,\ldots,l_M}[m']} + j\frac{\partial}{\partial \boldsymbol{\mathcal{B}}_{i_1,\ldots,i_N,l_1,\ldots,l_M}[m']} \tag{A.16}$$

Define the cost function

$$J(\boldsymbol{\mathcal{G}}) = \operatorname{trace}(\mathbb{E}[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{E}}^*[k]]) = \sum_{i_1} \ldots \sum_{i_N} \mathbb{E}\left[\boldsymbol{\mathcal{E}}_{i_1 \ldots i_N}[k] \boldsymbol{\mathcal{E}}^*_{i_1 \ldots i_N}[k]\right] \tag{A.17}$$

Due to the quadratic nature of the error surface, finding a stationary point assures global optimization of the cost function [63]. Minimizing the cost function is thus a convex unconstrained optimization problem [67] and can be solved be equating each component of the gradient tensor of the cost function to zero:

$$\nabla_{i_1,\ldots,i_N,m,l_1,\ldots,l_M} J(\boldsymbol{\mathcal{G}}^{opt}) = 0 \tag{A.18}$$

Let us consider one particular component of the gradient tensor where the indices have values $i'_1, \ldots, i'_N, m', l'_1, \ldots, l'_M$. From (A.17) we get

$$\begin{aligned}
\nabla_{i'_1,\ldots,i'_N,m',l'_1,\ldots,l'_M} J(\boldsymbol{\mathcal{G}}) &= \nabla_{i'_1,\ldots,i'_N,m',l'_1,\ldots,l'_M} \mathbb{E}[\sum_{i_1,\ldots,i_N} \boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N}[k] \boldsymbol{\mathcal{E}}^*_{i_1 \ldots i_N}[k]] \\
&= \frac{\partial\, \mathbb{E}[\sum_{i_1,\ldots,i_N} \boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N}[k] \boldsymbol{\mathcal{E}}^*_{i_1 \ldots i_N}[k]]}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\ldots,i'_N,l'_1,\ldots,l'_M}[m']} \\
&\quad + j\frac{\partial\, \mathbb{E}[\sum_{i_1,\ldots,i_N} \boldsymbol{\mathcal{E}}_{i_1,\ldots,i_N}[k] \boldsymbol{\mathcal{E}}^*_{i_1 \ldots i_N}[k]]}{\partial \boldsymbol{\mathcal{B}}_{i'_1,\ldots,i'_N,l'_1,\ldots,l'_M}[m']}
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[\sum_{i_1,\dots,i_N} \frac{\partial \boldsymbol{\mathcal{E}}_{i_1,\dots,i_N}[k]\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k]}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']} + j\frac{\partial \boldsymbol{\mathcal{E}}_{i_1,\dots,i_N}\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}}{\partial \boldsymbol{\mathcal{B}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}] \\
&= \mathbb{E}[\sum_{i_1,\dots,i_N} \frac{\partial \boldsymbol{\mathcal{E}}_{i_1,\dots,i_N}[k]}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k] + \frac{\partial \boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k]}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}\boldsymbol{\mathcal{E}}_{i_1\dots i_N}[k] \\
&\quad + \frac{\partial \boldsymbol{\mathcal{E}}_{i_1\dots i_N}[k]}{\partial \boldsymbol{\mathcal{B}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}j\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k] + \frac{\partial \boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k]}{\partial \boldsymbol{\mathcal{B}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}j\boldsymbol{\mathcal{E}}_{i_1\dots i_N}[k]]
\end{aligned} \tag{A.19}
$$

The first term on the right hand side of (A.19) can be expanded as

$$
\begin{aligned}
&\sum_{i_1,\dots,i_N} \frac{\partial \boldsymbol{\mathcal{E}}_{i_1,\dots,i_N}[k]}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k] \\
&\quad = \sum_{i_1,\dots,i_N} \frac{\partial\{\sum_m \sum_{l_1\dots l_M} \boldsymbol{\mathcal{G}}_{i_1\dots i_N l_1\dots l_M}[m]\boldsymbol{\mathcal{Y}}_{l_1\dots l_M}[k-m] - \boldsymbol{\mathcal{D}}_{i_1\dots i_N}[k]\}}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k] \\
&\quad = \sum_{i_1,\dots,i_N}\sum_m\sum_{l_1\dots l_M} \frac{\partial\{(\boldsymbol{\mathcal{A}}_{i_1\dots i_N l_1\dots l_M}[m] + j\boldsymbol{\mathcal{B}}_{i_1\dots i_N l_1\dots l_M}[m])\boldsymbol{\mathcal{Y}}_{l_1\dots l_M}[k-m] - \boldsymbol{\mathcal{D}}_{i_1\dots i_N}[k]\}}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k] \\
&\quad = \boldsymbol{\mathcal{Y}}_{l'_1\dots l'_M}[k-m']\boldsymbol{\mathcal{E}}^*_{i'_1\dots i'_N}[k]
\end{aligned} \tag{A.20}
$$

Similarly we have

$$
\frac{\partial \boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k]}{\partial \boldsymbol{\mathcal{A}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}\boldsymbol{\mathcal{E}}_{i_1\dots i_N}[k] = \boldsymbol{\mathcal{Y}}^*_{l'_1\dots l'_M}[k-m']\boldsymbol{\mathcal{E}}_{i'_1\dots i'_N}[k] \tag{A.21}
$$

,

$$
\begin{aligned}
\frac{\partial \boldsymbol{\mathcal{E}}_{i_1\dots i_N}[k]}{\partial \boldsymbol{\mathcal{B}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}j\boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k] &= j\boldsymbol{\mathcal{Y}}_{l'_1\dots l'_M}[k-m']j\boldsymbol{\mathcal{E}}^*_{i'_1\dots i'_N}[k] \\
&= -\boldsymbol{\mathcal{Y}}_{l'_1\dots l'_M}[k-m']\boldsymbol{\mathcal{E}}^*_{i'_1\dots i'_N}[k]
\end{aligned} \tag{A.22}
$$

and,

$$
\begin{aligned}
\sum_{i_1\dots i_N} \frac{\partial \boldsymbol{\mathcal{E}}^*_{i_1\dots i_N}[k]}{\partial \boldsymbol{\mathcal{B}}_{i'_1,\dots,i'_N,l'_1,\dots,l'_M}[m']}j\boldsymbol{\mathcal{E}}_{i_1\dots i_N}[k] &= -j\boldsymbol{\mathcal{Y}}^*_{l'_1\dots l'_M}[k-m']j\boldsymbol{\mathcal{E}}_{i'_1\dots i'_N}[k] \\
&= \boldsymbol{\mathcal{Y}}^*_{l'_1\dots l'_M}[k-m']\boldsymbol{\mathcal{E}}_{i'_1\dots i'_N}[k]
\end{aligned} \tag{A.23}
$$

Substituting these in (A.19) we get

$$
\nabla_{i'_1,\dots,i'_N,m',l'_1,\dots,l'_M} J(\boldsymbol{\mathcal{G}}) = \mathbb{E}\Big[\boldsymbol{\mathcal{Y}}_{l'_1\dots l'_M}[k-m']\boldsymbol{\mathcal{E}}^*_{i'_1\dots i'_N}[k] + \boldsymbol{\mathcal{Y}}^*_{l'_1\dots l'_M}[k-m']\boldsymbol{\mathcal{E}}_{i'_1\dots i'_N}[k]
$$

$$
\begin{aligned}
&\left. - \boldsymbol{\mathcal{Y}}_{l'_1\ldots l'_M}[k-m']\boldsymbol{\mathcal{E}}^*_{i'_1\ldots i'_N}[k] + \boldsymbol{\mathcal{Y}}^*_{l'_1\ldots l'_M}[k-m']\boldsymbol{\mathcal{E}}_{i'_1\ldots i'_N}[k]\right] \\
&= \mathbb{E}\left[2\boldsymbol{\mathcal{Y}}^*_{l'_1\ldots l'_M}[k-m']\boldsymbol{\mathcal{E}}_{i'_1\ldots i'_N}[k]\right] \\
&= 2\,\mathbb{E}\left[\boldsymbol{\mathcal{Y}}^*_{l'_1\ldots l'_M}[k-m']\boldsymbol{\mathcal{E}}_{i'_1\ldots i'_N}[k]\right] \qquad \text{(A.24)}
\end{aligned}
$$

The optimal $\boldsymbol{\mathcal{G}}[m]$ is found by equating (A.24) to 0 for all values of $i'_1,\ldots,i'_N,m',l'_1,\ldots,l'_M$. This gives

$$
\mathbb{E}\left[\boldsymbol{\mathcal{Y}}^*_{l'_1,\ldots,l'_M}[k-m']\boldsymbol{\mathcal{E}}_{i'_1\ldots i'_N}[k]\right] = 0 \quad \text{for all } i'_1,\ldots,i'_N,m',l'_1,\ldots,l'_M \qquad \text{(A.25)}
$$

We can see that the LHS of (A.25) is the auto-correlation between the error and the observation. We have

$$
\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}},\boldsymbol{\mathcal{Y}}_{i'_1,\ldots,i'_N,m',l'_1,\ldots,l'_M}}[-m'] = \mathbb{E}\left[\boldsymbol{\mathcal{E}}_{l'_1,\ldots,l'_M}[k]\boldsymbol{\mathcal{Y}}^*_{l'_1,\ldots,l'_M}[k-m']\right] = 0 \quad \text{for all } i'_1,\ldots,i'_N,m',l'_1,\ldots,l'_M \qquad \text{(A.26)}
$$

Since (A.26) holds for all values of m', we can write this in tensor notation as

$$
\begin{aligned}
\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}},\boldsymbol{\mathcal{Y}}}[m] &= \mathbb{E}\left[\boldsymbol{\mathcal{E}}[k]\circ\boldsymbol{\mathcal{Y}}^*[k-m]\right] \\
&= \mathbf{0}_T \qquad \text{(A.27)}
\end{aligned}
$$

Showing that for the optimal $\boldsymbol{\mathcal{G}}[m]$, the error is uncorrelated with the observation. This can be considered as a tensor orthogonality condition. □

A.3 A Cauchy-Schwartz Inequality

For two tensors $\boldsymbol{\mathcal{A}}(t) \in \mathbb{C}_t^{J_1\times\ldots\times J_M}$ and $\boldsymbol{\mathcal{B}}(t) \in \mathbb{C}_t^{J_1\times\ldots\times J_M}$ the following inequality holds:

$$
\left|\sum_{j_1\ldots j_M}\int_{-\infty}^{+\infty}\boldsymbol{\mathcal{A}}_{j_1\ldots j_M}(t)\boldsymbol{\mathcal{B}}_{j_1\ldots j_M}(t)dt\right|^2 \leq \sum_{j_1\ldots j_M}\int_{-\infty}^{+\infty}\left|\boldsymbol{\mathcal{A}}_{j_1\ldots j_M}(t)\right|^2 dt \sum_{j_1\ldots j_M}\int_{-\infty}^{+\infty}\left|\boldsymbol{\mathcal{B}}_{j_1\ldots j_M}(t)\right|^2 dt \qquad \text{(A.28)}
$$

Proof. Let λ be a complex scalar. We have

$$
0 \leq \sum_{j_1\ldots j_M}\int_{-\infty}^{+\infty}\left|\boldsymbol{\mathcal{A}}_{j_1\ldots j_M}(t)+\lambda\boldsymbol{\mathcal{B}}_{j_1\ldots j_M}(t)\right|^2 dt
$$

$$= \sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \left|\mathcal{A}_{j_1\dots j_M}(t)\right|^2 dt + |\lambda|^2 \sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \left|\mathcal{B}_{j_1\dots j_M}(t)\right|^2 dt + 2\Re\Big\{\lambda^* \sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \mathcal{A}_{j_1\dots j_M}(t)\mathcal{B}^*_{j_1\dots j_M}(t)dt\Big\} \tag{A.29}$$

Choose $\lambda = y\sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \mathcal{A}_{j_1\dots j_M}(t)\mathcal{B}^*_{j_1\dots j_M}(t)dt$ where $y \in \mathbb{R}$. This gives

$$\begin{aligned} 0 \le &\Big(\Big|\sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \mathcal{A}_{j_1\dots j_M}(t)\mathcal{B}^*_{j_1\dots j_M}(t)dt\Big|^2 \sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \left|\mathcal{B}_{j_1\dots j_M}(t)\right|^2 dt\Big)y^2 \\ &+ 2y\Big(\Big|\sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \mathcal{A}_{j_1\dots j_M}(t)\mathcal{B}^*_{j_1\dots j_M}(t)dt\Big|^2\Big) + \sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \left|\mathcal{A}_{j_1\dots j_M}(t)\right|^2 dt \end{aligned} \tag{A.30}$$

Let $p = \Big|\sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \mathcal{A}_{j_1\dots j_M}(t)\mathcal{B}^*_{j_1\dots j_M}(t)dt\Big|^2$, $r = \sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \left|\mathcal{B}_{j_1\dots j_M}(t)\right|^2 dt$ and $q = \sum_{j_1\dots j_M} \int_{-\infty}^{+\infty} \left|\mathcal{A}_{j_1\dots j_M}(t)\right|^2 dt$ such that

$$0 \le pry^2 + 2py + q \tag{A.31}$$

Since (A.31) is a non-negative quadratic polynomial, the discriminant is non-positive. i.e., $(2p)^2 - 4prq \le 0$ or $p \le rq$. Substituting for p,q and r gives (A.28) with equality when $\mathcal{B}_{i_1,\dots,i_N,j_1,\dots,j_M}(t) = \mathcal{A}^*_{j_1,\dots,j_M,i_1,\dots,i_N}(-t)$. □

A.4 Proof of Equation 4.148

We wish to show that for a finite length equalizer with N taps, (4.145) tends to (4.137) in the limit as N tends to infinity. i.e.,

$$\lim_{N\to\infty} \mathcal{R}^{\text{finite}}_{\mathcal{E},\min} = \mathcal{R}^{\text{inf}}_{\mathcal{E},\min} \tag{A.32}$$

where $\mathcal{R}^{\text{finite}}_{\mathcal{E},\min}$ is the error correlation tensor for the finite tap equalizer and $\mathcal{R}^{\text{inf}}_{\mathcal{E},\min}$ is the error correlation tensor for the infinite tap equalizer.

Proof. Denote the RHS of (4.148) by $\mathcal{R}^{\text{finite}}_{\mathcal{E},\min}$ and the finite tap equalizer by $\bar{\mathcal{G}}^{\text{finite}}$. The components of this tensor are

$$\mathfrak{R}^{\text{finite}}_{\mathfrak{E},\min_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} = \mathfrak{R}_{\mathfrak{D}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} - \sum_{m=1}^{M}\sum_{l_1,\ldots,l_P} \bar{\mathfrak{G}}^{\text{finite}}_{i_1,\ldots,i_N,m,l_1,\ldots,l_P}\,\mathbb{E}\big[\bar{\mathfrak{Y}}_{m,l_1,\ldots,l_P}[k]\mathfrak{D}_{i'_1,\ldots,i'_N}\big]$$

Writing the components of $\bar{\mathfrak{G}}^{\text{finite}}$ and $\bar{\mathfrak{Y}}[k]$ in terms of the components of $\mathfrak{G}^{\text{finite}}$ and $\mathfrak{Y}[k]$ gives

$$\mathfrak{R}^{\text{finite}}_{\mathfrak{E},\min_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} = \mathfrak{R}_{\mathfrak{D}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} - \sum_{m=0}^{M-1}\sum_{l_1,\ldots,l_P} \mathfrak{G}^{\text{finite}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[m]\,\mathbb{E}\big[\mathfrak{Y}_{l_1,\ldots,l_P}[k-m]\mathfrak{D}_{i'_1,\ldots,i'_N}[k]\big] \tag{A.33}$$

As the number of taps M tends to infinity, this becomes

$$\begin{aligned}\mathfrak{R}^{\text{finite}}_{\mathfrak{E},\min_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} &= \mathfrak{R}_{\mathfrak{D}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} - \sum_{m=0}^{\infty}\sum_{l_1,\ldots,l_P} \mathfrak{G}^{\text{finite}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[m]\,\mathbb{E}\big[\mathfrak{Y}_{l_1,\ldots,l_P}[k-m]\mathfrak{D}_{i'_1,\ldots,i'_N}[k]\big]\\ &= \mathfrak{R}_{\mathfrak{D}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}} - \mathbb{E}\big[\sum_{m=0}^{\infty}\sum_{l_1,\ldots,l_P} \mathfrak{G}^{\text{finite}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[m]\mathfrak{Y}_{l_1,\ldots,l_P}[k-m]\mathfrak{D}_{i'_1,\ldots,i'_N}[k]\big]\end{aligned} \tag{A.34}$$

Writing the components of either side of (4.131) gives

$$\sum_{m,l_1,\ldots,l_P} \bar{\mathfrak{G}}^{finite}_{i_1,\ldots,i_N,m,l_1,\ldots,l_P}\mathfrak{R}_{\bar{\mathfrak{Y}}_{m,l_1,\ldots,l_P,m',l'_1,\ldots,l'_P}} = \mathfrak{R}_{\mathfrak{D},\bar{\mathfrak{Y}}_{i_1,\ldots,i_N,m',l'_1,\ldots,l'_P}} \tag{A.35}$$

which can be written in terms of $\mathfrak{G}^{\text{finite}}[m]$,$\mathfrak{Y}[m]$ and $\mathfrak{D}[m]$ as

$$\sum_{m,l_1,\ldots,l_P} \mathfrak{G}^{\text{finite}}_{i_1,\ldots,i_N,l_1,\ldots,l_P}[m]\,\mathbb{E}[\mathfrak{Y}_{l_1,\ldots,l_P}[k-m]\mathfrak{Y}_{l'_1,\ldots,l'_P}[k-m']] = \mathbb{E}[\mathfrak{D}_{i_1,\ldots,i_N}[k]\mathfrak{Y}_{l_1,\ldots,l_P}[k-m']] \tag{A.36}$$

Notice that $\mathbb{E}[\mathfrak{Y}_{l_1,\ldots,l_P}[k-m]\mathfrak{Y}_{l'_1,\ldots,l'_P}[k-m']] = \mathfrak{R}_{\mathfrak{Y}_{l_1,\ldots,l_P,l'_1,\ldots,l'_P}}[m'-m]$ and $\mathbb{E}[\mathfrak{D}_{i_1,\ldots,i_N}[k]\mathfrak{Y}_{l_1,\ldots,l_P}[k-m']] = \mathfrak{R}_{\mathfrak{D},\mathfrak{Y}_{l_1,\ldots,l_P,l'_1,\ldots,l'_P}}[m']$. Using this, we may write (A.36) in tensor form as $\sum_m\{\mathfrak{G}^{\text{finite}}[m'],\mathfrak{R}_{\mathfrak{Y}}[m'-m]\}_{(P)} = \mathfrak{R}_{\mathfrak{D},\mathfrak{Y}}[m']$. The D-transform of this equation is

$$\{\breve{\mathfrak{G}}^{\text{finite}}(D),\breve{\mathfrak{S}}_{\mathfrak{Y}}(D)\}_{(P)} = \breve{\mathfrak{S}}_{\mathfrak{D},\mathfrak{Y}}(D) \tag{A.37}$$

Comparing (A.37) with (4.95), we can see that when there are an infinite number of taps, the finite length equalizer $\mathfrak{G}^{\text{finite}}[m] = \mathfrak{G}^{\text{inf}}[m]$ where $\mathfrak{G}^{\text{inf}}[m]$ denotes the infinite length equalizer. Substituting this in (A.34) gives

$$\mathcal{R}^{\text{finite}}_{\mathcal{E},\min_{i_1,\dots,i_N,i'_1,\dots,i'_N}} = \mathcal{R}_{\mathcal{D}_{i_1,\dots,i_N,i'_1,\dots,i'_N}} - \mathbb{E}\Big[\sum_{m=0}^{\infty}\sum_{l_1,\dots,l_P} \mathcal{G}^{\text{inf}}_{i_1,\dots,i_N,l_1,\dots,l_P}[m]\mathcal{Y}_{l_1,\dots,l_P}[k-m]\mathcal{D}_{i'_1,\dots,i'_N}[k]\Big] \tag{A.38}$$

Since $\sum_{m=0}^{\infty}\sum_{l_1,\dots,l_P} \mathcal{G}^{\text{inf}}_{i_1,\dots,i_N,l_1,\dots,l_P}[m]\mathcal{Y}_{l_1,\dots,l_P}[k-m] = \hat{\mathcal{D}}_{i_1,\dots,i_N}[k]$, (A.38) becomes

$$\mathcal{R}^{\text{finite}}_{\mathcal{E},\min_{i_1,\dots,i_N,i'_1,\dots,i'_N}} = \mathcal{R}_{\mathcal{D}_{i_1,\dots,i_N,i'_1,\dots,i'_N}} - \mathbb{E}\big[\hat{\mathcal{D}}_{i_1,\dots,i_N}\mathcal{D}_{i'_1,\dots,i'_N}[k]\big] = \mathcal{R}_{\mathcal{D}_{i_1,\dots,i_N,i'_1,\dots,i'_N}} - \mathcal{R}_{\hat{\mathcal{D}}[k],\mathcal{D}_{i_1,\dots,i_N,i'_1,\dots,i'_N}} \tag{A.39}$$

which can be written, in tensor notation, as $\mathcal{R}^{\text{finite}}_{\mathcal{E},\min} = \mathcal{R}_{\mathcal{D}} - \mathcal{R}_{\hat{\mathcal{D}},\mathcal{D}}$. Since $\mathcal{R}_{\mathcal{D}}$ and $\mathcal{R}_{\mathcal{D},\mathcal{D}}$ are short-hand for $\mathcal{R}_{\mathcal{D}}[0]$ and $\mathcal{R}_{\mathcal{D},\mathcal{D}}[0]$ respectively, the RHS of this is the same as the RHS of (4.137). □

A.5 A derivation for the Finite Tap DFE

Starting from the RHS of (4.182) and using (4.181), we have

$$\begin{aligned}\mathbb{E}\left[\mathcal{E}[k]\circ\mathcal{E}[k]^*\right] &= \mathbb{E}\left[(\{\bar{\mathcal{W}},\bar{\mathcal{Y}}[k]\}_{(P+1)} - \{\bar{\mathcal{B}},\bar{\mathcal{D}}[k]\}_{(N+1)})\circ(\{\bar{\mathcal{W}},\bar{\mathcal{Y}}[k]\}_{(P+1)} - \{\bar{\mathcal{B}},\bar{\mathcal{D}}[k]\}_{(N+1)})^*\right]\\ &= \mathbb{E}\left[\{\bar{\mathcal{W}},\bar{\mathcal{Y}}[k]\}_{(P+1)}\circ\{\bar{\mathcal{W}}^*,\bar{\mathcal{Y}}^*[k]\}_{(P+1)} - \{\bar{\mathcal{W}},\bar{\mathcal{Y}}[k]\}_{(P+1)}\circ\{\bar{\mathcal{B}}^*,\bar{\mathcal{D}}^*[k]\}_{(N+1)}\right.\\ &\quad\left. - \{\bar{\mathcal{B}},\bar{\mathcal{D}}[k]\}_{(N+1)}\circ\{\bar{\mathcal{W}}^*,\bar{\mathcal{Y}}^*[k]\}_{(P+1)} + \{\bar{\mathcal{B}},\bar{\mathcal{D}}[k]\}_{(N+1)}\circ\{\bar{\mathcal{B}}^*,\bar{\mathcal{D}}^*[k]\}_{(N+1)}\right]\end{aligned} \tag{A.40}$$

Using (2.35) in (A.40) we get

$$\begin{aligned}\mathbb{E}\left[\mathcal{E}[k]\circ\mathcal{E}[k]^*\right] &= \mathbb{E}\left[\{\bar{\mathcal{W}},\bar{\mathcal{Y}}[k]\}_{(P+1)}\circ\{\bar{\mathcal{Y}}^*,\bar{\mathcal{W}}^H\}_{(P+1)} - \{\bar{\mathcal{W}},\bar{\mathcal{Y}}[k]\}_{(P+1)}\circ\{\bar{\mathcal{D}}[k]^*,\bar{\mathcal{B}}^H\}_{(N+1)}\right.\\ &\quad\left. - \{\bar{\mathcal{B}},\bar{\mathcal{D}}[k]\}_{(N+1)}\circ\{\bar{\mathcal{Y}}^*[k],\bar{\mathcal{W}}^H\}_{(P+1)} + \{\bar{\mathcal{B}},\bar{\mathcal{D}}[k]\}_{(N+1)}\circ\{\bar{\mathcal{D}}^*[k],\bar{\mathcal{B}}^H\}_{(N+1)}\right]\end{aligned} \tag{A.41}$$

Using the associativitiy property, (A.41) becomes

$$\begin{aligned}\mathbb{E}\left[\mathcal{E}[k]\circ\mathcal{E}[k]^*\right] &= \left\{\bar{\mathcal{W}},\left\{\mathbb{E}\left[\bar{\mathcal{Y}}[k]\circ\bar{\mathcal{Y}}^*[k]\right],\bar{\mathcal{W}}^H\right\}_{(P+1)}\right\}_{(P+1)} - \left\{\bar{\mathcal{W}},\left\{\mathbb{E}\left[\bar{\mathcal{Y}}[k]\circ\bar{\mathcal{D}}^*[k]\right],\bar{\mathcal{B}}^H\right\}_{(N+1)}\right\}_{(P+1)}\\ &\quad - \left\{\bar{\mathcal{B}},\left\{\mathbb{E}\left[\bar{\mathcal{D}}[k]\circ\bar{\mathcal{Y}}^*[k]\right],\bar{\mathcal{W}}^H\right\}_{(P+1)}\right\}_{(N+1)}\\ &\quad + \left\{\bar{\mathcal{B}},\left\{\mathbb{E}\left[\bar{\mathcal{D}}[k]\circ\bar{\mathcal{D}}^*[k]\right],\bar{\mathcal{B}}^H\right\}_{(N+1)}\right\}_{(N+1)}\end{aligned}$$

$$
\begin{aligned}
&= \{\bar{\boldsymbol{\mathcal{W}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}^{H}\}_{(P+1)}\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{W}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(P+1)} \\
&\quad - \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}^{H}\}_{(P+1)}\}_{(N+1)} + \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(N+1)}
\end{aligned}
\tag{A.42}
$$

which can be written, by adding and subtracting $\{\bar{\boldsymbol{\mathcal{B}}}, \{\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}^{-1},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(N+1)}$ from (A.42), as

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{E}}[k]^{*}\right] &= \{\bar{\boldsymbol{\mathcal{W}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}^{H}\}_{(P+1)}\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{W}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(P+1)} \\
&\quad - \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}^{H}\}_{(P+1)}\}_{(N+1)} + \{\boldsymbol{\mathcal{B}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(N+1)} \\
&\quad + \{\bar{\boldsymbol{\mathcal{B}}}, \{\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(N+1)} \\
&\quad - \{\bar{\boldsymbol{\mathcal{B}}}, \{\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(N+1)} \\
&= \left\{\bar{\boldsymbol{\mathcal{B}}}, \left\{\left(\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}} - \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}\right), \bar{\boldsymbol{\mathcal{B}}}^{H}\right\}_{(N+1)}\right\}_{(N+1)} \\
&+ \{\bar{\boldsymbol{\mathcal{W}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}^{H}\}_{(P+1)}\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{W}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(P+1)} \\
&- \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}^{H}\}_{(P+1)}\}_{(N+1)} + \{\bar{\boldsymbol{\mathcal{B}}}, \{\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\}_{(N+1)}
\end{aligned}
\tag{A.43}
$$

Taking out the common terms gives

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{E}}[k]^{*}\right] &= \left\{\bar{\boldsymbol{\mathcal{B}}}, \left\{\left(\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}} - \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}\right), \bar{\boldsymbol{\mathcal{B}}}^{H}\right\}_{(N+1)}\right\}_{(N+1)} \\
&\quad + \{\bar{\boldsymbol{\mathcal{W}}}, \left(\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}^{H}\}_{(P+1)} - \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\right)\}_{(P+1)} \\
&\quad - \{\{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}\}_{(P+1)}\}_{N+1}, \left(\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}, \bar{\boldsymbol{\mathcal{W}}}\}_{(P+1)} + \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}^{H}\}_{(N+1)}\right)\}_{(P+1)}
\end{aligned}
\tag{A.44}
$$

which on simplifying gives

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{E}}[k]^{*}\right] &= \left\{\bar{\boldsymbol{\mathcal{B}}}, \left\{\left(\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}}} - \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}\}_{(P+1)}\}_{(P+1)}\right), \bar{\boldsymbol{\mathcal{B}}}^{H}\right\}_{(N+1)}\right\}_{(N+1)} \\
+ \Bigg\{\left(\bar{\boldsymbol{\mathcal{W}}} - \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{\mathcal{D}}},\bar{\boldsymbol{y}}}, \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}\}_{(P+1)}\}_{(N+1)}\right)&, \left\{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}, \left(\bar{\boldsymbol{\mathcal{W}}}^{H} - \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}}}^{-1}, \{\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{y}},\bar{\boldsymbol{\mathcal{D}}}}, \bar{\boldsymbol{\mathcal{B}}}\}_{(N+1)}\}_{(P+1)}\right)\right\}_{(P+1)}\Bigg\}_{(P+1)}
\end{aligned}
\tag{A.45}
$$

Using $\boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}^H = \boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}$ and $\boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}},\bar{\mathcal{D}}}^H = \boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}$ we get

$$\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \boldsymbol{\mathcal{E}}[k]^*\right] = \left\{\bar{\boldsymbol{\mathcal{B}}}, \left\{\left(\boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}}} - \{\boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}^{-1}, \boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}},\bar{\mathcal{D}}}\}_{(P+1)}\}_{(P+1)}\right), \bar{\boldsymbol{\mathcal{B}}}^H\right\}_{(N+1)}\right\}_{(N+1)}$$
$$+\left\{\left(\bar{\boldsymbol{\mathcal{W}}} - \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}^{-1}\}_{(P+1)}\}_{(N+1)}\right)\left\{\boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}, \left(\bar{\boldsymbol{\mathcal{W}}} - \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}^{-1}\}_{(P+1)}\}_{(N+1)}\right)^H\right\}_{(P+1)}\right\}_{(P+1)} \tag{A.46}$$

We wish to find tap coefficient tensors that minimise mean squared error (MSE) which is the trace of $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{E}}}$. To find the optimal tap co-efficient tensors, using Theorem 3, we set the error uncorrelated to the observation (i.e., $\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \bar{\boldsymbol{\mathcal{Y}}}[k]^*\right] = \mathbf{0}_T$) and using (4.181) we get

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\mathcal{E}}[k] \circ \bar{\boldsymbol{\mathcal{Y}}}[k]^*\right] &= \mathbb{E}\left[\left(\{\bar{\boldsymbol{\mathcal{W}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{B}}}, \bar{\boldsymbol{\mathcal{D}}}[k]\}_{(N+1)}\right) \circ \bar{\boldsymbol{\mathcal{Y}}}[k]^*\right] \\
&= \mathbb{E}\left[\{\bar{\boldsymbol{\mathcal{W}}}, \bar{\boldsymbol{\mathcal{Y}}}[k]\}_{(P+1)} \circ \bar{\boldsymbol{\mathcal{Y}}}[k]^*\right] - \mathbb{E}\left[\{\bar{\boldsymbol{\mathcal{B}}}, \bar{\boldsymbol{\mathcal{D}}}[k]\}_{(N+1)} \circ \bar{\boldsymbol{\mathcal{Y}}}[k]^*\right] \\
&= \{\bar{\boldsymbol{\mathcal{W}}}, \mathbb{E}\left[\bar{\boldsymbol{\mathcal{Y}}}[k] \circ \bar{\boldsymbol{\mathcal{Y}}}[k]^*\right]\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{B}}}, \mathbb{E}\left[\bar{\boldsymbol{\mathcal{D}}}[k] \circ \bar{\boldsymbol{\mathcal{Y}}}[k]^*\right]\}_{(N+1)} \\
&= \{\bar{\boldsymbol{\mathcal{W}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}\}_{(P+1)} - \{\bar{\boldsymbol{\mathcal{B}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}\}_{(N+1)} \\
&= \mathbf{0}_T
\end{aligned} \tag{A.47}$$

which gives

$$\bar{\boldsymbol{\mathcal{W}}} = \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}^{-1}\}_{(P+1)}\}_{(N+1)} \tag{A.48}$$

From (A.48) we have $\left(\bar{\boldsymbol{\mathcal{W}}} - \{\bar{\boldsymbol{\mathcal{B}}}, \{\boldsymbol{\mathcal{R}}_{\bar{\mathcal{D}},\bar{\mathcal{Y}}}, \boldsymbol{\mathcal{R}}_{\bar{\mathcal{Y}}}^{-1}\}_{(P+1)}\}_{(N+1)}\right) = \mathbf{0}_T$. By substituting this in (A.46) we get (4.183).

A.6 Decision Feedback Equalizer

The derivation of the optimal feedforward and feedback systems of the infitnite tap DFE depends on the factorization of spectral tensors. We assume that the spectrum of the input $\breve{\boldsymbol{\mathcal{D}}}(D)$ has a factorization

$$\breve{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\mathcal{D}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}(D), \breve{\boldsymbol{\mathcal{A}}}^H(D^{-1})\}_{(N)} \tag{A.49}$$

such that $\breve{\boldsymbol{\mathcal{A}}}(D)$ is causal with a stable and causal inverse. We can see that if $\breve{\boldsymbol{\mathfrak{D}}}(D)$ is the input to a system $\breve{\boldsymbol{\mathcal{A}}}^{-1}(D)$, the output

$$\breve{\boldsymbol{\mathfrak{U}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{-1}(D), \breve{\boldsymbol{\mathfrak{D}}}(D)\}_{(N)} \tag{A.50}$$

is WSS and white, and using (4.21), we get the spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{U}}}(D) = \boldsymbol{\mathfrak{I}}_N$. To simplify the derivation we now re-state the input-output relation (4.52), using (A.50), as

$$\breve{\boldsymbol{\mathfrak{Y}}}(D) = \{\breve{\boldsymbol{\mathfrak{H}}}(D), \{\breve{\boldsymbol{\mathcal{A}}}(D), \breve{\boldsymbol{\mathfrak{U}}}(D)\}_{(N)}\}_{(N)} + \breve{\boldsymbol{\mathfrak{V}}}(D) \tag{A.51}$$

Further, to find the dependence of the input to the decision device $\tilde{\breve{\boldsymbol{\mathfrak{D}}}}(D)$ on $\breve{\boldsymbol{\mathfrak{U}}}(D)$ we decompose the feedforward and feedback systems into a cascade of systems such that

$$\breve{\boldsymbol{\mathfrak{W}}}(D) = \{\breve{\boldsymbol{\Gamma}}(D), \breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1})\}_{(N)} \tag{A.52}$$

$$\breve{\boldsymbol{\mathfrak{B}}}(D) = \{\breve{\boldsymbol{\mathfrak{C}}}(D), \breve{\boldsymbol{\mathcal{A}}}^{-1}(D)\}_{(N)} \tag{A.53}$$

The input to the decision device (4.209) can now be re-written as

$$\tilde{\breve{\boldsymbol{\mathfrak{D}}}}(D) = \{\breve{\boldsymbol{\Gamma}}(D), \tilde{\breve{\boldsymbol{\mathfrak{Y}}}}(D)\}_{(N)} - \{\breve{\boldsymbol{\mathfrak{C}}}(D), \hat{\breve{\boldsymbol{\mathfrak{U}}}}(D)\}_{(N)} \tag{A.54}$$

where

$$\tilde{\breve{\boldsymbol{\mathfrak{Y}}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1}), \breve{\boldsymbol{\mathfrak{Y}}}(D)\}_{(N)} \tag{A.55}$$

and $\hat{\breve{\boldsymbol{\mathfrak{U}}}}(D) = \breve{\boldsymbol{\mathcal{A}}}^{-1}(D), \hat{\breve{\boldsymbol{\mathfrak{D}}}}(D)$ is the sequence of estimates of the past components of $\breve{\boldsymbol{\mathfrak{U}}}(D)$. The equalization problem now changes to finding a feedforward system $\breve{\boldsymbol{\Gamma}}(D)$ and a feedback system $\breve{\boldsymbol{\mathfrak{C}}}(D)$ such that the mean squared error between the estimate $\hat{\breve{\boldsymbol{\mathfrak{U}}}}(D)$ and the input $\breve{\boldsymbol{\mathfrak{U}}}(D)$ is minimized. Since $\breve{\boldsymbol{\mathcal{A}}}^{-1}(D)$ is a causal system and $\breve{\boldsymbol{\mathfrak{B}}}(D)$ is a purely causal system, $\breve{\boldsymbol{\mathfrak{C}}}(D)$ has to be a purely causal system. Assuming that past decisions are correct, we have

$$\{\breve{\boldsymbol{\mathfrak{C}}}(D), \hat{\breve{\boldsymbol{\mathfrak{U}}}}(D)\}_{(N)} = \{\breve{\boldsymbol{\mathfrak{C}}}(D), \breve{\boldsymbol{\mathfrak{U}}}(D)\}_{(N)} \tag{A.56}$$

Substituting (A.51) in (A.54) and using (A.56) we have the estimate $\hat{\breve{\boldsymbol{\mathfrak{D}}}}(D)$ in terms of $\breve{\boldsymbol{\mathfrak{U}}}(D)$ given by

$$\begin{aligned}
\tilde{\breve{\boldsymbol{\mathfrak{D}}}}(D) &= \{\Big(\{\{\{\breve{\boldsymbol{\Gamma}}(D), \breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1})\}_{(N)}, \breve{\boldsymbol{\mathfrak{H}}}(D)\}_{(N)}, \breve{\boldsymbol{\mathcal{A}}}(D)\}_{(N)} - \breve{\boldsymbol{\mathfrak{C}}}(D)\Big), \breve{\boldsymbol{\mathfrak{U}}}(D)\}_{(N)} \\
&\quad + \{\breve{\boldsymbol{\Gamma}}(D), \{\breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1}), \breve{\boldsymbol{\mathfrak{V}}}(D)\}_{(N)}\}_{(N)} \\
&= \{\Big(\breve{\boldsymbol{\mathfrak{K}}}(D) - \breve{\boldsymbol{\mathfrak{C}}}(D)\Big), \breve{\boldsymbol{\mathfrak{U}}}(D)\}_{(N)} + \tilde{\breve{\boldsymbol{\mathfrak{V}}}}(D)
\end{aligned} \tag{A.57}$$

where

$$\breve{\mathcal{K}}(D) = \{\{\{\breve{\boldsymbol{\Gamma}}(D), \breve{\mathcal{A}}^H(D^{-1})\}_{(N)}, \breve{\mathcal{H}}(D)\}_{(N)}, \breve{\mathcal{A}}(D)\}_{(N)} \tag{A.58}$$

and the noise sequence $\tilde{\breve{\mathcal{V}}}(D) = \{\breve{\boldsymbol{\Gamma}}(D), \{\breve{\mathcal{A}}^H(D^{-1}), \breve{\mathcal{V}}(D)\}_{(N)}\}_{(N)}$. Notice that $\tilde{\breve{\mathcal{V}}}(D)$ is the output of a system $\{\breve{\boldsymbol{\Gamma}}(D), \breve{\mathcal{A}}^H(D^{-1})\}_{(N)}$ for an input $\breve{\mathcal{V}}(D)$. Hence, using (4.21), we get the spectrum $\breve{\mathcal{S}}_{\tilde{\mathcal{V}}}(D)$ as

$$\breve{\mathcal{S}}_{\tilde{\mathcal{V}}}(D) = \{\{\left(\{\breve{\boldsymbol{\Gamma}}(D), \breve{\mathcal{A}}^H(D^{-1})\}_{(N)}\right), \breve{\mathcal{S}}_{\mathcal{V}}\}_{(N)}, \left(\{\breve{\boldsymbol{\Gamma}}(D^{-1}), \breve{\mathcal{A}}^H(D)\}_{(N)}\right)^H\}_{(N)} \tag{A.59}$$

We know, from (4.59), that the spectrum $\breve{\mathcal{S}}_{\mathcal{V}} = N_0\breve{\mathcal{H}}(D)$. Substituting this in (A.59) and using the associativity property gives

$$\begin{aligned}
\breve{\mathcal{S}}_{\tilde{\mathcal{V}}}(D) &= \{\{\left(\{\breve{\boldsymbol{\Gamma}}(D), \breve{\mathcal{A}}^H(D^{-1})\}_{(N)}\right), N_0\breve{\mathcal{H}}(D)\}_{(N)}, \left(\{\breve{\boldsymbol{\Gamma}}(D^{-1}), \breve{\mathcal{A}}^H(D)\}_{(N)}\right)^H\}_{(N)} \\
&= N_0\{\{\left(\{\breve{\boldsymbol{\Gamma}}(D), \breve{\mathcal{A}}^H(D^{-1})\}_{(N)}\right), \breve{\mathcal{H}}(D)\}_{(N)}, \left(\{\breve{\mathcal{A}}(D), \breve{\boldsymbol{\Gamma}}^H(D^{-1})\}_{(N)}\right)^H\}_{(N)} \\
&= N_0\{\left(\{\{\breve{\boldsymbol{\Gamma}}(D), \breve{\mathcal{A}}^H(D^{-1})\}_{(N)}, \breve{\mathcal{H}}(D)\}_{(N)}, \breve{\mathcal{A}}(D)\}_{(N)}\right), \breve{\boldsymbol{\Gamma}}^H(D^{-1})\}_{(N)} \\
&= N_0\{\breve{\mathcal{K}}(D), \breve{\boldsymbol{\Gamma}}^H(D^{-1})\}
\end{aligned} \tag{A.60}$$

The difference between the estimate and the input is

$$\hat{\breve{\mathcal{D}}}(D) - \breve{\mathcal{D}}(D) = \{(\breve{\mathcal{K}}(D) - \breve{\mathcal{C}}(D) - \breve{\mathcal{A}}(D)), \breve{\mathcal{U}}(D)\}_{(N)} + \{\breve{\boldsymbol{\Gamma}}(D), \{\breve{\mathcal{A}}^H(D^{-1}), \breve{\mathcal{V}}(D)\}_{(N)}\}_{(N)} \tag{A.61}$$

Using Theorem 3, for the mean square error to be minimized we require

$$\mathbb{E}[(\hat{\mathcal{D}}[k] - \mathcal{D}[k]) \circ \tilde{\mathcal{Y}}^*[k-i]] = 0 \quad \text{for all } i \tag{A.62}$$

and

$$\mathcal{R}_{(\hat{\mathcal{D}}-\mathcal{D}),\mathcal{U}}[i] = \mathbb{E}[(\hat{\mathcal{D}}[k] - \mathcal{D}[k]) \circ \mathcal{U}^*[k-i]] = 0_{\mathbf{T}} \quad \text{for } i > 0 \tag{A.63}$$

$$\mathcal{R}_{(\hat{\mathcal{D}}-\mathcal{D}),\mathcal{U}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}}[i] = \mathbb{E}[(\hat{\mathcal{D}}_{i_1,\ldots,i_N}[k] - \mathcal{D}_{i_1,\ldots,i_N}[k]) \circ \mathcal{U}^H_{i'_1,\ldots,i'_N}[k-i]] = 0 \quad \text{for } I_d < I_u \tag{A.64}$$

where $I_d = \left(i'_1 + \sum_{k=2}^{N}(i'_k - 1)\prod_{l=k-1} I_l\right)$ and $I_u = \left(i_1 + \sum_{k=2}^{N}(i_k - 1)\prod_{l=k-1} I_l\right)$. Comparing (A.64) with (4.38) and (4.42), we can see that this constraint is equivalent to requiring that

$$\boldsymbol{\mathcal{R}}^{+}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}}[0] = 0_{\mathbf{T}} \tag{A.65}$$

where $\boldsymbol{\mathcal{R}}^{+}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}}[0]$ is the purely causal part of $\boldsymbol{\mathcal{R}}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}}[0]$. From (4.41), we get that the purely causal part of the cross-spectrum $\breve{\boldsymbol{\mathfrak{S}}}^{+}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}} = \boldsymbol{\mathcal{R}}^{+}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}}[0] + \boldsymbol{\mathcal{R}}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}}[1] + \ldots$. Using (A.63) and (A.64), we get $\breve{\boldsymbol{\mathfrak{S}}}^{+}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}} = \mathbf{0}_T$. Notice that (A.61) describes the input-output relation for an input $\breve{\boldsymbol{\mathfrak{U}}}(D)$ to a system $\breve{\boldsymbol{\mathfrak{K}}}(D) - \breve{\boldsymbol{\mathfrak{C}}}(D) - \breve{\boldsymbol{\mathcal{A}}}(D)$ with an output $\hat{\breve{\boldsymbol{\mathfrak{D}}}}(D) - \breve{\boldsymbol{\mathfrak{D}}}(D)$. Using (4.22), we can find the cross-spectrum between the output and input of such a system as $\breve{\boldsymbol{\mathfrak{S}}}_{(\hat{\boldsymbol{\mathfrak{D}}}-\boldsymbol{\mathfrak{D}}),\boldsymbol{\mathfrak{U}}}(D) = \{\left(\breve{\boldsymbol{\mathfrak{K}}}(D) - \breve{\boldsymbol{\mathfrak{C}}}(D) - \breve{\boldsymbol{\mathcal{A}}}(D)\right), \breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{U}}}(D)\}_{(N)}$. Since $\breve{\boldsymbol{\mathfrak{U}}}(D)$ is has a spectrum $\breve{\boldsymbol{\mathfrak{S}}}_{\boldsymbol{\mathfrak{U}}}(D) = \boldsymbol{\mathfrak{I}}_N$, setting the purely causal part of the cross-spectrum to zero is equivalent to the purely causal part of $\breve{\boldsymbol{\mathfrak{K}}}(D) - \breve{\boldsymbol{\mathfrak{C}}}(D) - \breve{\boldsymbol{\mathcal{A}}}(D)$ being set to zero. Thus the constraint is equivalent to

$$(\breve{\boldsymbol{\mathfrak{K}}}(D) - \breve{\boldsymbol{\mathfrak{C}}}(D) - \breve{\boldsymbol{\mathcal{A}}}(D))^{+} = \mathbf{0}_T \tag{A.66}$$

From which, since $\breve{\boldsymbol{\mathfrak{C}}}(D)$ is purely causal, we get the feedback filter

$$\breve{\boldsymbol{\mathfrak{C}}}(D) = (\breve{\boldsymbol{\mathfrak{K}}}(D) - \breve{\boldsymbol{\mathcal{A}}}(D))^{+} \tag{A.67}$$

To find the feedforward filter $\breve{\boldsymbol{\Gamma}}(D)$ we substitute (A.50),(A.55) and (A.57) into the orthogonality constraint (A.62). This gives

$$\tilde{\breve{\boldsymbol{\mathfrak{Y}}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1}), \{\breve{\boldsymbol{\mathfrak{H}}}(D), \{\breve{\boldsymbol{\mathcal{A}}}(D), \breve{\boldsymbol{\mathfrak{U}}}(D)\}_{(N)}\}_{(N)}\}_{(N)} + \{\breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1}), \breve{\boldsymbol{\mathfrak{V}}}(D)\}_{(N)} \tag{A.68}$$

Define systems $\boldsymbol{\mathfrak{P}}[k]$ and $\boldsymbol{\mathfrak{X}}[k]$ such that their D-transforms are $\{\breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1}), \{\breve{\boldsymbol{\mathfrak{H}}}(D), \breve{\boldsymbol{\mathcal{A}}}(D)\}_{(N)}\}_{(N)}$ and $\breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1})$ respectively. $\boldsymbol{\mathfrak{Q}}[k]$ and respectively. We know that from (4.19) the contracted product of two tensors in the D-domain is the contracted convolution in the time domain. Using this, we write the inverse D-transform of (A.68) as

$$\tilde{\boldsymbol{\mathfrak{Y}}}[k] = \sum_{m} \boldsymbol{\mathfrak{P}}[m]\boldsymbol{\mathfrak{U}}[k-m] + \sum_{n} \boldsymbol{\mathfrak{X}}[n]\boldsymbol{\mathfrak{V}}[k-n] \tag{A.69}$$

Further, Define systems $\boldsymbol{\mathfrak{Q}}[k]$ and $\boldsymbol{\mathfrak{O}}[k]$ such that their D-transforms are $\{(\breve{\boldsymbol{\mathfrak{K}}}(D) - \breve{\boldsymbol{\mathfrak{C}}}(D) - \breve{\boldsymbol{\mathcal{A}}}(D))$ and $\{\breve{\boldsymbol{\Gamma}}(D), \breve{\boldsymbol{\mathcal{A}}}^{H}(D^{-1})\}_{(N)}$ respectively. We can write the inverse D-transform of

(A.61) as

$$\hat{\mathfrak{D}}[k] - \mathfrak{D}[k] = \sum_m \mathfrak{Q}[m]\mathfrak{U}[k-m] + \sum_n \mathfrak{O}[n]\mathfrak{V}[k-n] \tag{A.70}$$

The criterion of (A.65) then becomes

$$\mathbb{E}[(\sum_m \mathfrak{Q}[m]\mathfrak{U}[k-m]+\sum_n \mathfrak{O}[n]\mathfrak{V}[k-n])\circ(\sum_{m,} \mathfrak{P}[m']\mathfrak{U}[k-m'-i]+\sum_{n'} \mathfrak{X}[n']\mathfrak{V}[k-n'-i])] = \mathbf{0}_T \tag{A.71}$$

On expanding and noticing that $\mathfrak{U}[k]$ and $\mathfrak{V}[k]$ are uncorrelated we get

$$\sum_{m,m'} \mathfrak{Q}[m]\mathfrak{R}_{\mathfrak{U}}[m'+i-m]\mathfrak{P}^H[m'] + \sum_{m,m'} \mathfrak{O}[m]\mathfrak{R}_{\mathfrak{U}}[m'+i-m]\mathfrak{X}^H[m'] = \mathbf{0}_T \tag{A.72}$$

Since the spectrum of $\mathfrak{U}[k]$ is $\breve{\mathfrak{S}}_{\mathfrak{U}}(D) = \mathfrak{I}_N$, we have $\mathfrak{R}_{\mathfrak{U}}[i] = \mathbf{0}_T$ for $i \neq 0$ and $\mathfrak{R}_{\mathfrak{U}}[0] = \mathfrak{I}_N$. Using this, we get

$$\sum_m \{\mathfrak{Q}[m]\mathfrak{P}^H[i+m]\}_{(N)} + \sum_m \sum_{m'} \{\{\mathfrak{O}[m], \mathfrak{R}_{\mathfrak{U}}[m'+i-m], \mathfrak{X}^H[i+m]\}_{(N)}\}_{(N)} = \mathbf{0}_T \tag{A.73}$$

Taking the D-transform of (A.73) and using $\breve{\mathfrak{S}}_{\mathfrak{V}} = N_0\breve{\mathfrak{H}}(D)$ and $\breve{\mathfrak{X}}(D) = \breve{\mathfrak{A}}^H(D^{-1})$ gives

$$\breve{\mathfrak{S}}_{(\mathfrak{D}-\mathfrak{D}),\tilde{\mathfrak{y}}} = \{\breve{\mathfrak{Q}}(D), \breve{\mathfrak{P}}^H(D^{-1})\}_{(N)} + N_0\{\breve{\mathfrak{O}}(D)\{\breve{\mathfrak{H}}(D), \breve{\mathfrak{A}}(D)\}_{(N)}\}_{(N)} = \mathbf{0}_T \tag{A.74}$$

Replacing the values of $\breve{\mathfrak{P}}(D)$ and $\breve{\mathfrak{Q}}(D)$ gives

$$\begin{aligned}\breve{\mathfrak{S}}_{(\mathfrak{D}-\mathfrak{D}),\tilde{\mathfrak{y}}} &= \{\breve{\mathfrak{Q}}(D), \breve{\mathfrak{P}}^H(D^{-1})\}_{(N)} + N_0\{\breve{\mathbf{\Gamma}}(D), \{\breve{\mathfrak{A}}^H(D^{-1}), \{\breve{\mathfrak{H}}(D)\breve{\mathfrak{A}}^H(D^{-1})\}_{(N)}\}_{(N)}\}_{(N)} \\ &= \{\breve{\mathfrak{Q}}(D), \breve{\mathfrak{P}}^H(D^{-1})\}_{(N)} + N_0\{\breve{\mathbf{\Gamma}}(D), \breve{\mathfrak{P}}^H(D^{-1})\}_{(N)}\}_{(N)}\end{aligned} \tag{A.75}$$

Multiplying both sides by $(\breve{\mathfrak{P}}^H(D^{-1}))^{-1}$ gives

$$\breve{\mathfrak{Q}}(D) + N_0\breve{\mathbf{\Gamma}}(D) = \breve{\mathfrak{K}}(D) - \breve{\mathfrak{C}}(D) - \breve{\mathfrak{A}}(D) + N_0\breve{\mathbf{\Gamma}}(D) = \mathbf{0}_T \tag{A.76}$$

Substituting (A.67) in (A.76) and using the property $\breve{\mathfrak{Z}}(D) = \breve{\mathfrak{Z}}^+(D) + \breve{\mathfrak{Z}}^-(D)$ gives

$$\begin{aligned}&\breve{\mathfrak{K}}(D) - \breve{\mathfrak{C}}(D) - \breve{\mathfrak{A}}(D) + N_0\breve{\mathbf{\Gamma}}(D) \\ &= \breve{\mathfrak{K}}(D) - (\breve{\mathfrak{K}}(D) - \breve{\mathfrak{A}}(D))^+ - \breve{\mathfrak{A}}(D) + N_0\breve{\mathbf{\Gamma}}(D) \\ &= \breve{\mathfrak{K}}(D) - \breve{\mathfrak{K}}^+(D) + \breve{\mathfrak{A}}^+(D) - \breve{\mathfrak{A}}(D) + N_0\breve{\mathbf{\Gamma}}(D) \\ &= \breve{\mathfrak{K}}^-(D) - \breve{\mathfrak{A}}^-(D) + N_0\breve{\mathbf{\Gamma}}(D)\end{aligned} \tag{A.77}$$

Substituting (A.58) in (A.77) gives

$$\breve{\mathfrak{K}}^-(D) - \breve{\mathfrak{A}}^-(D) + N_0\breve{\mathbf{\Gamma}}(D)$$

$$= (\{\{\{\breve{\mathbf{\Gamma}}(D), \breve{\mathfrak{A}}^{H}(D^{-1})\}_{(N)}, \breve{\mathfrak{H}}(D)\}_{(N)}, \breve{\mathfrak{A}}(D)\}_{(N)})^{-} - \breve{\mathfrak{A}}^{-}(D) + N_0(\breve{\mathbf{\Gamma}}^{+}(D) + \breve{\mathbf{\Gamma}}^{-}(D))$$

$$= \left(\{\breve{\mathbf{\Gamma}}(D), \breve{\mathfrak{R}}(D)\}_{(N)}\right)^{-} + -\breve{\mathfrak{A}}^{-}(D) + N_0(\breve{\mathbf{\Gamma}}^{+}(D) \tag{A.78}$$

where $\breve{\mathfrak{R}}(D) = \{\{\breve{\mathfrak{A}}^{H}(D^{-1}), \breve{\mathfrak{H}}(D)\}_{(N)}, \breve{\mathfrak{A}}(D)\}_{(N)} + N_0\mathfrak{I}_N$. Equating this to $\mathbf{0}_T$ gives

$$\left(\{\breve{\mathbf{\Gamma}}(D), \breve{\mathfrak{R}}(D)\}_{(N)}\right)^{-} = -N_0\breve{\mathbf{\Gamma}}(D)^{+} + \breve{\mathfrak{A}}(D)^{-} \tag{A.79}$$

Notice that if the system $\breve{\mathfrak{A}}(D)$ is excited by an input with spectrum $\breve{\mathfrak{H}}(D)$ and its output is corrupted by noise with spectrum $N_0\mathfrak{I}_N$, then the spectrum of the output is the tensor $\breve{\mathfrak{R}}(D)$. Since $\breve{\mathfrak{R}}(D)$ is a spectrum, it can be factored as

$$\breve{\mathfrak{R}}(D) = \{\breve{\mathfrak{M}}(D), \breve{\mathfrak{M}}^{H}(D^{-1})\}_{(N)} \tag{A.80}$$

where $\breve{\mathfrak{M}}(D)$ is causal with a causal and stable inverse. Denote the tensor containing the pseudo-diagonal components of a tensor $\breve{\mathfrak{S}}(D)$ by $\breve{\mathfrak{S}}^{(d)}(D)$. Since $\breve{\mathfrak{A}}(D)$ is causal, its anticausal component $\breve{\mathfrak{A}}(D)^{-}$ is $\breve{\mathfrak{A}}^{(d)}(0)$. This means that the purely causal component of $\breve{\mathbf{\Gamma}}(D)$ is zero since the left hand side of (A.79) is anticausal. We may thus re-write (A.79) as

$$\left(\{\breve{\mathbf{\Gamma}}(D), \breve{\mathfrak{R}}(D)\}_{(N)}\right)^{-} = \breve{\mathfrak{A}}^{(d)}(0) \tag{A.81}$$

Using (A.80) and the fact that $\breve{\mathfrak{R}}(D)$ and $\breve{\mathfrak{R}}^{-1}(D)$ are both causal and stable, we get

$$\breve{\mathbf{\Gamma}}(D) = \{\breve{\mathfrak{A}}^{(d)}(0), \{(\breve{\mathfrak{M}}^{(d)}(0))^{-1}, (\breve{\mathfrak{M}}^{H}(D^{-1}))^{-1}\}_{(N)}\}_{(N)} \tag{A.82}$$

We know, by definition, that $\breve{\mathfrak{R}}(D) = \{\{\breve{\mathfrak{A}}^{H}(D^{-1}), \breve{\mathfrak{H}}(D)\}_{(N)}, \breve{\mathfrak{A}}(D)\}_{(N)} + N_0\mathfrak{I}_N$. This can be written as

$$\breve{\mathfrak{R}}(D) - N_0\mathfrak{I}_N = \{\{\breve{\mathfrak{A}}^{H}(D^{-1}), \breve{\mathfrak{H}}(D)\}_{(N)}, \breve{\mathfrak{A}}(D)\}_{(N)} \tag{A.83}$$

Contracting both sides of (A.83) with $\breve{\mathbf{\Gamma}}(D)$ gives

$$\{\breve{\mathbf{\Gamma}}(D), (\breve{\mathfrak{R}}(D) - N_0\mathfrak{I}_N)\}_{(N)} = \{\breve{\mathbf{\Gamma}}, \{\{\breve{\mathfrak{A}}^{H}(D^{-1}), \breve{\mathfrak{H}}(D)\}_{(N)}, \breve{\mathfrak{A}}(D)\}_{(N)}\}_{(N)} \tag{A.84}$$

Comparing this to (A.58) we get $\breve{\mathfrak{K}}(D) = \{\breve{\mathbf{\Gamma}}(D), (\breve{\mathfrak{R}}(D) - N_0\mathfrak{I}_N)\}_{(N)}$. Substituting this in (A.67) gives

$$\breve{\mathfrak{C}}(D) = (\breve{\mathfrak{K}}(D) - \breve{\mathfrak{A}}(D))^{+}$$
$$= (\{\breve{\mathbf{\Gamma}}(D), (\breve{\mathfrak{R}}(D) - N_0\mathfrak{I}_N)\}_{(N)} - \breve{\mathfrak{A}}(D))^{+}$$

$$= (\{\breve{\mathbf{\Gamma}}(D), \breve{\boldsymbol{\mathfrak{R}}}(D)\}_{(N)})^{+} - N_0\breve{\mathbf{\Gamma}}^{+}(D) - \breve{\boldsymbol{\mathcal{A}}}^{+}(D) \tag{A.85}$$

Since $\breve{\mathbf{\Gamma}}(D)$ is anti-causal and $\breve{\boldsymbol{\mathcal{A}}}^{+}(D)$ is purely causal, we get

$$\breve{\boldsymbol{\mathfrak{C}}}(D) = (\{\breve{\mathbf{\Gamma}}(D), \breve{\boldsymbol{\mathfrak{R}}}(D)\}_{(N)})^{+} - \breve{\boldsymbol{\mathcal{A}}}(D) \tag{A.86}$$

Using (A.80) and (A.82) this can be written as

$$\breve{\boldsymbol{\mathfrak{C}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{(d)}(0), \{(\breve{\boldsymbol{\mathfrak{M}}}^{(d)}(0))^{-1}, (\breve{\boldsymbol{\mathfrak{M}}}(D))\}_{(N)}\}_{(N)} - \breve{\boldsymbol{\mathcal{A}}}^{+}(D) \tag{A.87}$$

The next step is to use $\breve{\boldsymbol{\mathfrak{C}}}(D)$ and $\breve{\mathbf{\Gamma}}(D)$ to find the actual feedback and feed-forward systems $\breve{\boldsymbol{\mathfrak{W}}}(D)$ and $\breve{\boldsymbol{\mathfrak{B}}}(D)$. Substituting (A.82) in (A.52) we get

$$\breve{\boldsymbol{\mathfrak{W}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{(d)}(0), \{(\breve{\boldsymbol{\mathfrak{M}}}^{(d)}(0))^{-1}, \{\breve{\boldsymbol{\mathfrak{M}}}^{-H}(D^{-1}), \breve{\boldsymbol{\mathcal{A}}}^{H}(D)\}_{(N)}\}_{(N)}\}_{(N)}\}_{(N)} \tag{A.88}$$

and substituting (A.87) in (A.53) we get

$$\breve{\boldsymbol{\mathfrak{B}}}(D) = \{\breve{\boldsymbol{\mathcal{A}}}^{(d)}(0), \{(\breve{\boldsymbol{\mathfrak{M}}}^{(d)}(0))^{-1}\{\breve{\boldsymbol{\mathfrak{M}}}(D), \breve{\boldsymbol{\mathcal{A}}}^{-1}(D)\}_{(N)}\}_{(N)}\}_{(N)}\}_{(N)} - \boldsymbol{\mathfrak{I}}_N \tag{A.89}$$

Appendix B

Miscellaneous

B.1 Spectral Factorization of a Tensor

As an example of spectral factorization, we show the factorization of a $2 \times 2 \times 2 \times 2$ tensor $\breve{\boldsymbol{\mathfrak{H}}}(D) = \{\breve{\boldsymbol{\mathfrak{Q}}}(D), \breve{\boldsymbol{\mathfrak{Q}}}^H(D^{-1})\}_{(2)}$. The component functions of $\breve{\boldsymbol{\mathfrak{H}}}(D)$ are given below followed by the components of the factor tensor $\breve{\boldsymbol{\mathfrak{Q}}}(D)$. A MATLAB script, which can be obtained from the author or his supervisor, was used to compute the factor tensor $\breve{\boldsymbol{\mathfrak{Q}}}(D)$. The factor tensor $\breve{\boldsymbol{\mathfrak{Q}}}(D)$ is computed iteratively using a generalization of the method detailed in [58] by contracting the original tensor $\breve{\boldsymbol{\mathfrak{H}}}(D)$ with transformation tensors until the result is an identity tensor. i.e., on the ith step we compute

$$\breve{\boldsymbol{\Phi}}_i(D) = \{\breve{\boldsymbol{\mathfrak{T}}}_i(D), \{\breve{\boldsymbol{\Phi}}_{i-1}(D), \breve{\boldsymbol{\mathfrak{T}}}_i^H(D^{-1})\}_{(2)}\}_{(2)} \tag{B.1}$$

where $\breve{\boldsymbol{\Phi}}_0(D) = \breve{\boldsymbol{\mathfrak{H}}}(D)$. The transformation tensors $\breve{\boldsymbol{\mathfrak{T}}}_i(D)$ are chosen to remove the poles from the components of the spectrum tensor, reduce it to a tensor with numerical elements (the components do not depend on D) and then finally to the identity tensor. This is accomplished with the help of the *spectralfact* command in the MATLAB Control Systems Toolbox. The factor tensor $\breve{\boldsymbol{\mathfrak{Q}}}(D)$ is computed using (4.47) as

$$\breve{\boldsymbol{\mathfrak{Q}}}(D) = \{\{\breve{\boldsymbol{\mathfrak{T}}}_1^{-1}(D), \breve{\boldsymbol{\mathfrak{T}}}_2^{-1}(D)\}_{(2)}, \ldots \breve{\boldsymbol{\mathfrak{T}}}_L^{-1}(D)\}_{(2)} \tag{B.2}$$

To verify that the factorization was performed correctly, we compare

$$E_{i,j,k,l} = \left| \boldsymbol{\mathfrak{H}}_{i,j,k,l}(D) - \left(\{\boldsymbol{\mathfrak{Q}}(D), \boldsymbol{\mathfrak{Q}}^H(D^{-1})\}_{(2)} \right)_{i,j,k,l} \right| \tag{B.3}$$

with a threshold $\epsilon = 10^{-6}$. This is done to allow for small errors that may occur due to the MATLAB floating point computations.

Below, we present the components of the spectrum $\breve{\boldsymbol{\mathfrak{H}}}(D)$ followed by the components of the factor tensor $\breve{\boldsymbol{\mathfrak{Q}}}(D)$. The substitution $D = z^{-1}$ is made in the following:

$\boldsymbol{\mathfrak{H}}_{[1111]} = \frac{(0-3i)z^4+(1+2i)z^3+(26+3.497e-15i)z^2+(1-2i)z-(1.665e-16-3i)}{z^2}$

$\boldsymbol{\mathfrak{H}}_{[1112]} = \frac{(1+1i)z^4+(5+4.441e-15i)z^3-z^2-(8.882e-15-6i)z+(3+3.997e-15i)}{z^2}$

$\boldsymbol{\mathfrak{H}}_{[1121]} = \frac{(0+2i)z^3+(7+1i)z^2+(14+4.774e-15i)z-(2.567e-16-1i)}{z}$

$\boldsymbol{\mathfrak{H}}_{[1122]} = \frac{(0+1i)z^4+(4+1i)z^3+(2+1i)z^2+(7+12i)z+(3-2i)}{z^2}$

$\boldsymbol{\mathfrak{H}}_{[1211]} = \frac{3z^4-(4.163e-16+6i)z^3-z^2+(5-3.393e-15i)z+(1-1i)}{z^2}$

$\boldsymbol{\mathfrak{H}}_{[1212]} = \frac{(2+2i)z^2+(11+8.882e-16i)z+(2-2i)}{z}$

$\boldsymbol{\mathfrak{H}}_{[1221]} = \frac{(0+1i)z^4+(2-1.776e-15i)z^3-(2.887e-15-4i)z^2+(14+2i)z+(6-1i)}{z}$

$\boldsymbol{\mathfrak{H}}_{[1222]} = \frac{(1+2i)z^2+(13+1i)z+(2-4i)}{z}$

$\boldsymbol{\mathfrak{H}}_{[2111]} = \frac{(0-1i)z^3+(14+2.859e-15i)z^2+(7-1i)z+(1.443e-15-2i)}{z^2}$

$\boldsymbol{\mathfrak{H}}_{[2112]} = \frac{(6+1i)z^4+(14-2i)z^3-(1.11e-15+4i)z^2+(2+3.22e-15i)z+(1.943e-16-1i)}{z^3}$

$\boldsymbol{\mathfrak{H}}_{[2121]} = \frac{(0+10i)z^2+(54+1.796e-16i)z+(9.352e-16-10i)}{z}$

$\boldsymbol{\mathfrak{H}}_{[2122]} = \frac{-(2-6i)z^4+(18-2i)z^3+(6-4i)z^2+(2+3i)z-(6.106e-16+1i)}{z^3}$

$\boldsymbol{\mathfrak{H}}_{[2211]} = \frac{(3+2i)z^4+(7-12i)z^3+(2-1i)z^2+(4-1i)z-(2.22e-16+1i)}{z^2}$

$\boldsymbol{\mathfrak{H}}_{[2212]} = \frac{(2+4i)z^2+(13-1i)z+(1-2i)}{z}$

$\boldsymbol{\mathfrak{H}}_{[2221]} = \frac{(0+1i)z^4+(2-3i)z^3+(6+4i)z^2+(18+2i)z-(2+6i)}{z}$

$\boldsymbol{\mathfrak{H}}_{[2222]} = \frac{-2z^4+(1+2i)z^3+(28+2.887e-15i)z^2+(1-2i)z-(2+5.829e-16i)}{z^2}$

The components of $\boldsymbol{\mathfrak{Q}}(D)$ are:

$\boldsymbol{\mathfrak{Q}}_{[1111]} = \frac{(1.244+0.9071i)z^2+(1.047+0.01788i)z-(0.2901-0.07506i)}{z^2}$

$\boldsymbol{\mathfrak{Q}}_{[1112]} = \frac{(0.6037-0.09844i)z^3-(0.662-0.1946i)z^2+(0.2456-0.04223i)z-(0.1873+0.05397i)}{z^3}$

$\boldsymbol{\mathfrak{Q}}_{[1121]} = \frac{(1.598-2.85i)z+(0.4016+0.85i)}{z}$

$\boldsymbol{\mathfrak{Q}}_{[1122]} = \frac{-(0.126-0.2595i)z^3-(0.1497+0.1795i)z^2+(1.463-0.02602i)z-(0.1873+0.05397i)}{z^3}$

$\boldsymbol{\mathfrak{Q}}_{[1211]} = \frac{(0.5452-3.605i)z^2-(0.08818-0.3898i)z+(0.543+0.2152i)}{z^2}$

$$\mathbf{Q}_{[1212]} = \frac{(0.9499+0.5713i)z^3+(1.009+0.7283i)z^2+(0.05367-0.3157i)z-(0.01227-0.01607i)}{z^3}$$
$$\mathbf{Q}_{[1221]} = \frac{(1.253-1.611i)z-(0.2525-1.611i)}{z}$$
$$\mathbf{Q}_{[1222]} = \frac{(1.705-0.03864i)z^3+(2.722+0.06879i)z^2-(0.4154+0.04622i)z-(0.01227-0.01607i)}{z^3}$$
$$\mathbf{Q}_{[2111]} = \frac{(2.507-0.1955i)z^2+(0.1541-0.1037i)z+(0.3394+0.2991i)}{z^2}$$
$$\mathbf{Q}_{[2112]} = \frac{-(0.7295+0.5761i)z^3+(0.9594+0.5259i)z^2+(0.6389+0.07285i)z+(0.1312-0.02266i)}{z^3}$$
$$\mathbf{Q}_{[2121]} = \frac{(0.1786+0.4215i)z+(1.821+0.5785i)}{z}$$
$$\mathbf{Q}_{[2122]} = \frac{-(0.5714+1.322i)z^3+(1.86+1.017i)z^2+(0.58+0.3281i)z+(0.1312-0.02266i)}{z^3}$$
$$\mathbf{Q}_{[2211]} = \frac{(0.08686+0.2953i)z^2+(1.16-0.007057i)z+(0.753-0.2882i)}{z^2}$$
$$\mathbf{Q}_{[2212]} = \frac{(1.51+0.6016i)z^3+(1.29-0.5906i)z^2+(0.07483-0.01868i)z+(0.1247+0.00769i)}{z^3}$$
$$\mathbf{Q}_{[2221]} = \frac{(1.224+2.649i)z+(4.776-0.6495i)}{z}$$
$$\mathbf{Q}_{[2222]} = \frac{(0.3229+2.176i)z^3+(1.572-0.8408i)z^2-(0.01939+0.3432i)z+(0.1247+0.00769i)}{z^3}$$

B.2 Simulation Program User Guide

This appendix details the software that was used to generate the simulation and analytical results in this thesis. The software can be obtained from the author or his supervisor. The software distribution consists of one main folder and two sub-folders. The first sub folder contains a library of general functions and the second sub folder contains matlab files that produce the figures of this thesis. The matlab file that produces a certain figure is named after the figure as it appears in this thesis. The tensor library is a directory that consists of functions that perform tensor operations and are used throughout all the simulations and analytical results. Each matlab file is titled based on the operation that it performs. For example, the matlab file titled "tensor_contraction.m" is a function that takes two tensors as its input and returns their contracted product over a specified number of domains as its output. These functions are required to run the matlab files that produce the figures and it is hence important to maintain the structure of the folders as they are. Most of the error rate results are obtained by simultaneously simulating for different SNRs on different instances of MATLAB. The final curves are obtained by collecting the results of these runs.

To reproduce a particular figure from the thesis, simply run the corresponding MATLAB file with the default settings. A comprehensive README document is provided along with the software distribution that details each matlab file with instructions on how to run them.

References

[1] Cisco Systems inc., “Visual networking index.” `https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html`, 2017.

[2] G. Fettweis and S. Alamouti, “5G: Personal mobile internet beyond what cellular did to telephony,” *IEEE Communications Magazine*, vol. 52, pp. 140–145, February 2014.

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

[4] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, “Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture,” *Electronics Letters*, vol. 35, pp. 14–16, Jan 1999.

[5] T. G. Kolda and B. W. Bader, “Tensor Decompositions and Applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[6] F. L. Hitchcock, “The expression of a tensor or a polyadic as a sum of products,” *Journal of Mathematics and Physics*, vol. 6, no. 1-4, pp. 164–189, 1927.

[7] R. B. Cattell, “Parallel proportional profiles” and other principles for determining the choice of factors by rotation,” *Psychometrika*, vol. 9, pp. 267–283, Dec 1944.

[8] L. R. Tucker, “Some mathematical notes on three-mode factor analysis,” *Psychometrika*, vol. 31, pp. 279–311, Sep 1966.

[9] J. D. Carroll and J.-J. Chang, “Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition,” *Psychometrika*, vol. 35, pp. 283–319, Sep 1970.

[10] R. Bro, “Review on multiway analysis in chemistry,” *Critical Reviews in Analytical Chemistry*, vol. 36, no. 3-4, pp. 279–293, 2006.

[11] C. J. Appellof and E. R. Davidson, “Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents,” *Analytical Chemistry*, vol. 53, no. 13, pp. 2053–2056, 1981.

[12] R. Bro, "PARAFAC. Tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 149 – 171, 1997.

[13] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[14] B. Chen, A. P. Petropulu, and L. De Lathauwer, "Blind identification of convolutive MIMO systems with 3 sources and 2 sensors," *EURASIP Journal on Advances in Signal Processing*, vol. 2002, pp. 432–606, May 2002.

[15] E. Acar, S. A. Çamtepe, M. S. Krishnamoorthy, and B. Yener, "Modeling and multiway analysis of chatroom tensors," in *Intelligence and Security Informatics* (P. Kantor, G. Muresan, F. Roberts, D. D. Zeng, F.-Y. Wang, H. Chen, and R. C. Merkle, eds.), (Berlin, Heidelberg), Springer Berlin Heidelberg, 2005.

[16] B. W. Bader, R. A. Harshman, and T. G. Kolda, "Temporal analysis of semantic graphs using ASALSAN," in *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pp. 33–42, Oct 2007.

[17] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Computer Vision — ECCV 2002* (A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, eds.), (Berlin, Heidelberg), pp. 447–460, Springer Berlin Heidelberg, 2002.

[18] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," in *Object recognition supported by user interaction for service robots*, vol. 2, pp. 511–514 vol.2, Aug 2002.

[19] D. Muti and S. Bourennane, "Multidimensional filtering based on a tensor approach," *Signal Processing*, vol. 85, no. 12, pp. 2338 – 2353, 2005.

[20] L. D. Lathauwer, J. Castaing, and J. F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2965–2973, June 2007.

[21] M. Brazell, N. Li, C. Navasca, and C. Tamon, "Solving multilinear systems via tensor inversion," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 2, pp. 542–570, 2013.

[22] L. Sun, B. Zheng, C. Bu, and Y. Wei, "MoorePenrose inverse of tensors via Einstein product," *Linear and Multilinear Algebra*, vol. 64, no. 4, pp. 686–698, 2016.

[23] R. Pan, "Tensor transpose and its properties," *CoRR*, vol. abs/1411.1503, 2014.

[24] "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, vol. 18, no. 2, pp. 95 – 138, 1977.

[25] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of n-way arrays," *Journal of Chemometrics*, vol. 14, no. 3, pp. 229–239, 2000.

[26] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Transactions on Signal Processing*, vol. 48, pp. 810–823, Mar 2000.

[27] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, pp. 2377–2388, Aug 2000.

[28] T. Jiang and N. D. Sidiropoulos, "A direct blind receiver for SIMO and MIMO OFDM systems subject to unknown frequency offset and multipath," in *2003 4th IEEE Workshop on Signal Processing Advances in Wireless Communications - SPAWC 2003 (IEEE Cat. No.03EX689)*, pp. 358–362, June 2003.

[29] N. Sidiropoulos and R. Budampati, "Khatri-Rao Space-time Codes," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2396–2407, Oct. 2002.

[30] A. L. F. de Almeida, G. Favier, and L. R. Ximenes, "Space-time-frequency (STF) MIMO communication systems with blind receiver based on a generalized PARATUCK2 Model," *IEEE Transactions on Signal Processing*, vol. 61, pp. 1895–1909, April 2013.

[31] K. Liu, J. P. C. L. da Costa, H. C. So, and A. L. F. de Almeida, "Semi-blind receivers for joint symbol and channel estimation in space-time-frequency MIMO-OFDM systems," *IEEE Transactions on Signal Processing*, vol. 61, pp. 5444–5457, Nov 2013.

[32] A. L. de Almeida, G. Favier, and J. C. M. Mota, "PARAFAC-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization," *Signal Processing*, vol. 87, no. 2, pp. 337 – 351, 2007. Tensor Signal Processing.

[33] A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "Space-time multiplexing codes: A tensor modeling approach," in *2006 IEEE 7th Workshop on Signal Processing Advances in Wireless Communications*, pp. 1–5, July 2006.

[34] G. Favier and A. L. F. de Almeida, "Tensor space-time-frequency coding with semi-blind receivers for mimo wireless communication systems," *IEEE Transactions on Signal Processing*, vol. 62, pp. 5987–6002, Nov 2014.

[35] M. Nazareth da Costa, G. Favier, and J. Romano, "Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers," *Signal Processing*, vol. 145, pp. 304–316, 04 2018.

[36] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Transactions on Communications*, vol. 23, pp. 921–934, Sep 1975.

[37] S. Pasupathy, "Correlative coding: A bandwidth-efficient signaling scheme," *IEEE Communications Society Magazine*, vol. 15, pp. 4–11, July 1977.

[38] M.-L. Liang, B. Zheng, and R.-J. Zhao, "Tensor inversion and its application to the tensor equations with Einstein product," *Linear and Multilinear Algebra*, vol. 67, no. 4, pp. 843–870, 2019.

[39] P. Yang, Y. Xiao, Y. L. Guan, M. Di Renzo, S. Li, and L. Hanzo, "Multidomain Index Modulation for Vehicular and Railway Communications: A Survey of Novel Techniques," *IEEE Vehicular Technology Magazine*, vol. 13, pp. 124–134, Sep. 2018.

[40] H. Gacanin, "Inter-domain bi-directional access in G.hn with network coding at the physical-layer," in *2012 IEEE International Symposium on Power Line Communications and Its Applications*, pp. 144–149, March 2012.

[41] V. Oksman and S. Galli, "G.hn: The new itu-t home networking standard," *IEEE Communications Magazine*, vol. 47, pp. 138–145, October 2009.

[42] A. Papoulis, *The Fourier Integral and its Applications.* New York: McGraw-Hill, 1962.

[43] D. A. Shnidman, "A generalized Nyquist criterion and an optimum linear receiver for a pulse modulation system," *The Bell System Technical Journal*, vol. 46, pp. 2163–277, Nov 1967.

[44] J. W. Smith, "A unified view of synchronous data transmission system design," *The Bell System Technical Journal*, vol. 47, pp. 273–300, March 1968.

[45] W. van Etten, "An optimum linear receiver for multiple channel digital transmission systems," *IEEE Transactions on Communications*, vol. 23, pp. 828–834, Aug 1975.

[46] E. Kofidis, C. Chatzichristos, and A. L. F. de Almeida, "Joint channel estimation / data detection in MIMO-FBMC/OQAM systems A tensor-based approach," in *25th European Signal Processing Conference (EUSIPCO)*, pp. 420–424, Aug 2017.

[47] K. Naskovska, S. A. Cheema, M. Haardt, B. Valeev, and Y. Evdokimov, "Iterative GFDM receiver based on the PARATUCK2 tensor decompo-sition," in *21st International ITG workshop on Smart Antennas*, pp. 420–424, 2017.

[48] "Physical Channels and Modulation," standard, European Telecommunications Standards Institute (ETSI), 2018.

[49] B. Farhang-Boroujeny, "OFDM Versus Filter Bank Multicarrier," *IEEE Signal Processing Magazine*, vol. 28, pp. 92–112, May 2011.

[50] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1170–1183, May 2002.

[51] C. Lélé, P. Siohan, and R. Legouable, "The Alamouti Scheme with CDMA-OFDM/OQAM," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 703–513, Jan 2010.

[52] N. Michailow, M. Matth, I. S. Gaspar, A. N. Caldevilla, L. L. Mendes, A. Festag, and G. Fettweis, "Generalized frequency division multiplexing for 5th generation cellular networks," *IEEE Transactions on Communications*, vol. 62, pp. 3045–3061, Sept 2014.

[53] J. G. Proakis, *Digital Communications, 5th. Edition.* McGraw-Hill, 1995.

[54] W. Etten, "Maximum likelihood receiver for multiple channel transmission systems," *Communications, IEEE Transactions on*, vol. 24, pp. 276 – 283, 03 1976.

[55] W. van Etten, "An optimum linear receiver for multiple channel digital transmission systems," *IEEE Transactions on Communications*, vol. 23, pp. 828–834, August 1975.

[56] A. Duel-Hallen, "Equalizers for multiple input/multiple output channels and PAM systems with cyclostationary input sequences," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 630–639, April 1992.

[57] B. A. Bjecke and J. G. Proakis, "Multiple transmit and receive antenna diversity techniques for wireless communications," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pp. 70–75, Oct 2000.

[58] M. Davis, "Factoring the spectral matrix," *IEEE Transactions on Automatic Control*, vol. 8, pp. 296–305, October 1963.

[59] P. Motyka and J. Cadzow, "The factorization of discrete-process spectral matrices," *IEEE Transactions on Automatic Control*, vol. 12, pp. 698–707, December 1967.

[60] John M. Cioffi, "Signal Processing and Detection." `http://web.stanford.edu/group/cioffi/`.

[61] N. Al-Dhahir and A. H. Sayed, "A computationally-efficient FIR MMSE-DFE for multiuser communications," in *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers (Cat. No.00CH37154)*, vol. 2, pp. 1867–1871 vol.2, Oct 2000.

[62] N. Michailow, S. Krone, M. Lentmaier, and G. Fettweis, "Bit error rate performance of generalized frequency division multiplexing," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sep. 2012.

[63] S. Haykin, *Adaptive Filter Theory (3rd Ed.).* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[64] A. Lender, "The duobinary technique for high-speed data transmission," *Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics*, vol. 82, pp. 214–218, May 1963.

[65] E. Kretzmer, "Generalization of a techinque for binary data communication," *IEEE Transactions on Communication Technology*, vol. 14, pp. 67–68, February 1966.

[66] Y. Zhang and H. Liu, "Frequency-domain correlative coding for mimo-ofdm systems over fast fading channels," *IEEE Communications Letters*, vol. 10, pp. 347–349, May 2006.

[67] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.