REDUCED-COMPLEXITY SIGNAL PROCESSING TECHNIQUES FOR MULTIPLE-INPUT MULTIPLE-OUTPUT STORAGE AND WIRELESS COMMUNICATION SYSTMES

LI HUANG

NATIONAL UNIVERSITY OF SINGAPORE

2007

REDUCED-COMPLEXITY SIGNAL PROCESSING TECHNIQUES FOR MULTIPLE-INPUT MULTIPLE-OUTPUT STORAGE AND WIRELESS COMMUNICATION SYSTMES

LI HUANG

(B. Eng, Huazhong University of Science & Technology)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2007

Acknowledgments

I would like to express my gratitude to my main supervisor, Prof. Chong Tow Chong, whose insights, perspective, and enthusiasm are a continual source of inspiration. I wish to thank my co-supervisor, Dr. George Mathew, for his considerable and sustained help and guidance in the area of signal processing for data storage systems. But for his help, I could not have reached so far. Dr. George Mathew critically reviewed the entire manuscript of this thesis and made numerous invaluable suggestions and polished this thesis in many ways. His resolute trust and constant encouragement, his extraordinary patience and cogent guidance facilitated my completion of this thesis. I want to thank my co-supervisor, Prof. Jan W. M. Bergmans, for leading me into the wonderful and challenging area of wireless communications. He provided indispensable advice by setting aside large amounts of his time for discussion and review of this thesis. I feel extremely fortunate to have him as my supervisor. I want to thank Prof. Frans M. J. Willems, for his counseling and expertise in multiple-input multiple-output (MIMO) systems. He has freely shared his time and insights with me and provided excellent guidance and comments during my stay in Technische Universiteit Eindhoven (TU/e). I also want to thank Mr. C. K. Ho for his direction in the area of channel estimation. Moreover, I owe a great deal to numerous people who provided me necessary support during the past four years. These are Prof. W. Ye, Dr. Y. Lin, Dr. K. S. Chan, Dr. Z. Qin, Ms. L. Chen, Mr. J. Riani, Mr. E. A. P. Habets, Mr. A. Martinez, Mr. A. Michael, Mr. X. Zou, Mr. B. W. Lim, Mr. E. M. Rachid and Ms. K. Cai. I would like to acknowledge Mr. C. Peh, Mr. H. van Meer, and H. M. Kuipers for their dependable and cheerful technical assistances. I would also like to acknowledge Ms. Y. E. M. Broers and Ms. A. Louis for their considerable administrative assistances. In addition, I am grateful to Ms. S. Sun and Mr. Y. Wu for their kind directions and guidance in the area of wireless communications during my short stay in Institute for Infocomm Research (I2R), Singapore. Even though I cannot list here all of the people who helped me accomplish this work, nevertheless I am indebted to all of them. In addition, I wish to thank National University of Singapore (NUS), Design Technology Institute (DTI) and TU/e for offering me the opportunity to pursue higher education. I am also thankful to Data Storage Institute (DSI) for providing all the necessary support and excellent research environment for my research work during the initial three years. Last but not least, my greatest thanks go to my family who have always been encouraging me throughout my studies. Without their love, support and sacrifice, my work would have been much more difficult.

Contents

Summary

Multiple-input multiple-output technology can provide many benefits and has been investigated for various digital communication systems. In this thesis, we explore reduced-complexity detection and channel estimation techniques to facilitate high-speed and high-quality data reception in two different systems with the multiple-input multiple-output technology. In Part I of the thesis, we concentrate on the development of reduced-complexity detection techniques to facilitate high-speed implementation of the two-dimensional optical storage (TwoDOS) system, which is expected to play a critical role in the development of the 4th generation optical storage system. Moreover, though the techniques we develop are for the TwoDOS system in which the bit-cells are arranged in a hexagonal structure, most of them are applicable to any multi-track data storage system with square or rectangular bit-cells. In Part II of the thesis, we study channel estimation techniques for multiple-input multiple-output systems where prior knowledge of the channel is not available. These channel estimation techniques perform noise filtering in the angle domain, where the channel model lends itself to a simple physical interpretation. To the best of our knowledge, this is the first work to systematically investigate these angle-domain channel estimation techniques. Though the techniques in this part are developed for multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems, they are applicable to other multiple-input multiple-output wireless communication systems as well.

In Part I of the thesis, we first present a channel model for the TwoDOS system in the presence of additive noise, domain bloom and transition jitter. We also propose a computationally efficient technique based on the 1D Hankel transform to simulate the channel model. Further, we develop an approximated model to simplify the signal generation process for the TwoDOS system with additive noise, domain bloom and transition jitter. The two-dimensional (2D) Viterbi detector (VD), which is the optimal 2D detector in the presence of additive white Gaussian noise, serves as the benchmark in terms of performance. Therefore, we develop techniques to reduce the complexity of the 2D VD in the temporal dimension in Chapter 3 and in the spatial dimension in Chapter 4 and

Chapter 5. We also develop a novel 2D target optimization technique and design several suitable targets to compensate for the detection performance loss due to the complexity reduction in both temporal and spatial dimensions.

In Part II of the thesis, we develop channel estimation techniques in the angle domain, where the channel model lends itself to a simple physical interpretation. All the angle-domain techniques proposed are flexible in implementation. They can either use conventional array-domain estimators as the coarse estimators and perform post-processing in the angle domain, or use the specifically designed pilots for the direct implementation. The applicability of these angle-domain techniques is highly dependent on the channel stochastic information (*e.g.* channel power or correlation) available to the receiver. For the situation where no channel stochastic information is available to the receiver, we develop the angle-frequency domain most significant taps (MST) selection technique, angle-time domain MST selection technique and angle-time domain approximated minimum mean square error (AMMSE) technique. For the situation where the channel power is known, we develop the angle-time domain channel power based AMMSE technique. For the situation where the channel correlation is known to the receiver, we develop the quasi one-dimensional (Q1D) linear minimum mean square error (LMMSE) technique that can further improve the performance. Our simulation results show that the Q1D LMMSE technique can perform similar to the 2D LMMSE technique yet with significantly lower complexity.

List of Tables

List of Figures

List of Abbreviations

The following abbreviations are adopted throughout this thesis.

1D	One-Dimensional
2D	Two-Dimensional
AMMSE	Approximated Minimum Mean Square Error
AoA	Angle of Arrival
AoD	Angle of Departure
AWGN	Additive White Gaussian Noise
BD	Blu-ray Disc
BER	Bit Error Rate
CCK	Complementary Code Keying
CD	Compact Disc
CP	Cyclic Prefix
DFE	Decision Feedback Equalizer
DFT	Discrete Fourier Transform
DVD	Digital Versatile Disc
ECC	Error Correction Codes
ETSI	European Telecommunication Standard Institute
FCC	Federal Communications Commission
FDTS	Fixed-Delay Tree Search
FDTS/DF	Fixed-Delay Tree Search with Decision Feedback
FFT	Fast Fourier Transform
GB	Gigabyte
Gb	Gigabit
HIPERLAN	High Performance Radio Local Area Network
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IFFT	Inverse Discrete Fourier Transform
ISI	Intersymbol Interference
ISM	Industrial Scientific and Medical
ITI	Intertrack Interference
Kb	Kilobit
LAN	Local Area Network
LMMSE	Linear Minimum Mean Square Error

LOS	Line-of-Sight
LPF	Low-Pass Filter
LS	Least Squares
MB	Megabyte
Mb	Megabit
MIMO	Multiple-Input Multiple-Output
ML	Maximum-Likelihood
MLSD	Maximum-Likelihood Sequence Detector
MMSE	Minimum Mean Square Error
MRD	Missing-Run Detector
MSE	Mean Square Error
MST	Most Significant Taps
MTF	Modulation Transfer Function
NA	Numerical Aperture
NLOS	Non-Line-of-Sight
OFDM	Orthogonal Frequency Division Multiplexing
PR	Partial Response
PD	Peak Detector
PDA	Personal Digital Assistant
Q1D	Quasi One-Dimensional
RD	Runlength Detector
RF	Radio Frequency
RLL	Runlength-Limited
SDM	Space-Division Multiplexing
SNR	Signal-to-Noise Ratio
$\mathrm{SNR}_{\mathrm{eff}}$	Effective Signal-to-Noise Ratio
SVD	Singular Value Decomposition
SWVD	Stripe-Wise Viterbi Detector
TD	Threshold Detector
TwoDOS	Two-Dimensional Optical Storage
UNII	Unlicensed National Information Infrastructure
VD	Viterbi Detector
VLP	Video Long Play
WLAN	Wireless Local Area Network

List of Symbols

In Part I of the thesis, the following symbols are often used.

$\mathbf{a}(n)$	channel input vector per group at time index n
$\hat{\mathbf{a}}(n)$	detected channel input vector per group at time index n
$\tilde{\mathbf{a}}(n)$	modified channel input vector due to domain bloom and transition jitter at time index n
$\mathbf{C}_s(\theta, r)$	2D linear pulse modulator in the spatial-temporal domain for a single spot
$\mathbf{d}_n$	target output vector at time index n
D	delay from the detector input to the detector output
$\mathbf{e}(n)$	error vector between the actual and detected channel input vectors at time index n
$\mathbf{F}_{sf}(\phi, \rho)$	2D modulation transfer function in the spatial-frequency domain for a single spot
$\mathbf{g}$	1D target vector transformed from the 2D target matrices
$\mathbf{G}_k$	2D target matrix at time delay k
$\mathbf{H}_{sf}(\phi, \rho)$	2D symbol response in the spatial-frequency domain for a single spot
$\tilde{\mathbf{H}}_{sf}(\rho)$	radially symmetric 2D symbol response in the spatial-frequency domain for a single spot
$\mathbf{H}_s(\theta, r)$	2D symbol response in the spatial-temporal domain for a single spot
$\tilde{\mathbf{H}}_s(r)$	radially symmetric 2D symbol response in the spatial-temporal domain for a single spot
$\mathbf{H}_k$	2D channel matrix at time delay k
$J_0(x)$	Bessel function of the first kind and zero order of x
$J_1(x)$	Bessel function of the first kind and first order of x
m_0	delay from the channel input to the equalizer output
$\mathbf{M}_k$	2D residual channel matrix at time delay k
N_e	error event length
N_g	target length
N_h	temporal span of the channel
N_r	number of tracks per group
N_s	number of data points in each dimension
N_{sr}	number of tracks per subgroup
N_w	equalizer length

NA	numerical aperture of lens
$O(x)$	order of x (used in complexity comparison)
r	radial coordinate in the 2D spatial-temporal plane
R	radius of the pit hole
$\underline{\mathbf{R}}_{aa}$	autocorrelation matrix of the channel input vector
$\underline{\mathbf{R}}_{z}$	autocorrelation matrix of the channel output vector
$\underline{\mathbf{R}}_{za}$	cross-correlation matrix between the channel output and input vectors
T	center-to-center distance between adjacent bits
$\mathbf{W}_k$	2D equalizer matrix at time delay k
$\mathbf{x}(n)$	equalizer output vector at time index n
$\mathbf{z}(n)$	channel output vector at time index n
Δ_b	normalized degree of domain bloom
$\boldsymbol{\varepsilon}(n)$	error vector between the equalizer and target output vectors at time index n
Δ_t	normalized degree of transition shift
λ	wave length of laser diodes
$\boldsymbol{\theta}(n)$	noise vector at time index n
ρ	spatial angular frequency in the 2D spatial-frequency plane
ρ_c	angular cut-off frequency in the 2D spatial-frequency plane
σ^2	variance of the additive white Gaussian noise
σ_t^2	variance of the normalized transition shift

In Part II of the thesis, the following symbols are often used.

$\mathbf{0}_{N_1 \times N_2}$	$(N_1 \times N_2)$ zero matrix
$\mathbf{C}(l)$	array-time domain channel matrix at time delay l
$\mathbf{C}^a(l)$	angle-time domain channel matrix at time delay l
$\mathbf{F}$	unitary Fourier matrix
$\mathbf{H}(k)$	array-frequency domain channel matrix channel matrix at kth subcarrier
$\mathbf{H}^a(k)$	angle-frequency domain channel matrix channel matrix at kth subcarrier
$\mathbf{I}_N$	$(N \times N)$ identity matrix
$J_0(x)$	Bessel function of the first kind and zero order of x
N_c	number of consecutive OFDM symbols for estimation
N_d	total number of subcarriers
N_g	cyclic prefix length
N_h	temporal span of the channel
$\tilde{N}_h$	estimated temporal span of the channel
N_p	number of pilot subcarriers
N_t	number of transmit antennas
N_r	number of receive antennas
$O(x)$	order of x (used in complexity comparison)
$\mathbf{R}$	2D channel correlation in the array-frequency domain
$\mathbf{R}^{\mathrm{a}}$	2D channel correlation in the angle-frequency domain
$\mathbf{s}(l,n)$	array-time domain channel input vector at lth sample and nth OFDM symbol
$\mathbf{s}^{\mathrm{a}}(l,n)$	angle-time domain channel input vector at lth sample and nth OFDM symbol
$\mathbf{x}(k,n)$	array-frequency domain channel input vector at kth subcarrier and nth OFDM symbol
$\mathbf{x}^{\mathrm{a}}(k,n)$	angle-frequency domain channel input vector at kth subcarrier and nth OFDM symbol
$\mathbf{y}(k,n)$	array-frequency domain channel output vector at kth subcarrier and nth OFDM symbol
$\mathbf{y}^{\mathrm{a}}(k,n)$	angle-frequency domain channel output vector at kth subcarrier and nth OFDM symbol
$\mathbf{z}(l,n)$	array-time domain channel output vector at lth sample and nth OFDM symbol
$\mathbf{z}^{\mathrm{a}}(l,n)$	angle-time domain channel output vector at lth sample and nth OFDM symbol

η	threshold
λ_{i-1}	ith largest eigenvalue of the 2D channel correlation matrix
Λ	diagonal matrix containing eigenvalues of the 2D channel correlation matrix
σ_f^2	variance of the additive white Gaussian noise
σ_i^2	power of the ith angle-time domain channel coefficient

Chapter 1

Introduction

1.1 Motivation

Multiple-input multiple-output technology can provide many benefits and has been investigated for various digital communication systems. For example, multi-track optical storage systems with parallel read-out can increase the data rate and storage density relative to single-track systems [44, 103, 135, 156]. Wireless communication systems with multiple transmit and receive antennas can increase the data rate and link reliability relative to systems with single transmit and receive antennas [13, 68, 175–177]. The main challenge in multiple-input multiple-output technology is the high computational complexity and the associated hardware complexity. Against this background, the scope of our research work in this thesis is the development of reduced-complexity signal processing techniques to facilitate high-speed and high-quality data reception in systems with multiple-input multiple-output technology.

Fig. 1.1 shows the main functional blocks that constitute a system with multiple-input multiple-output technology. The figure includes references to the chapters in this thesis that are devoted to each of the building blocks of the system. The blocks before and after the channel and additive noise form the transmitter and receiver, respectively. As shown, the input signal is first converted in the source encoder block into an efficient digital representation so as to facilitate transmission or storage. Then, redundant information is added to the source-encoded signal for the purpose of improving the resilience to errors

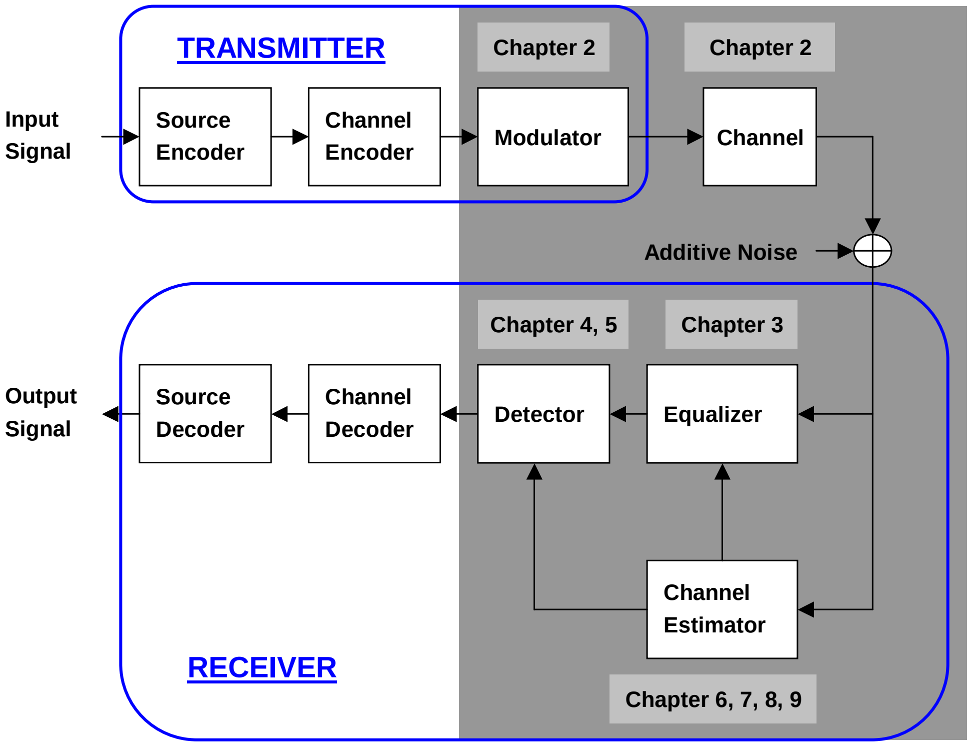

Figure 1.1: Block diagram of a system with multiple-input multiple-output technology.

caused by the channel. The modulator maps the channel-encoded signal into waveforms. This mapping operation may be followed by the modulation by a high-frequency carrier waveform. The modulated signal is now ready to pass through the channel by the use of multiple transmit antennas or a multi-track recording process. Throughout this thesis, the channel, which can represent the wireless propagation environment or storage medium, is considered as a link between the transmitted and received signals without additive noise. Usually, the channel is characterized by a set of coefficients, which are called channel coefficients. The additive noise is modeled as an additional component. Then, the signal is received by the multiple receive antennas or a parallel read-out configuration. At the receiver, the equalizer block acts to completely or partially undo the distortions caused by the channel. The detector block serves to make decisions on the signal from the equalizer output. Some of the channel distortions may also be accounted for in the detector block. Usually, the equalizer and detector blocks require knowledge of channel coefficients. These coefficients are estimated in the channel estimator block. Finally, the detected signal is passed through the channel decoder and source decoder to yield the output signal.

Fig. 1.1 correctly suggests that the receiver is, in general, more complex than the transmitter, both conceptually and in terms of hardware. Therefore, in this thesis, we focus on the design of reduced-complexity receivers for multiple-input multiple-output systems. In particular, we consider the detector and channel estimator block in which low-complexity and high-performance detection and estimation techniques, respectively, are developed. Detection and estimation techniques are two important topics in statistical signal processing for multiple-input multiple-output systems. Detection techniques serve to extract data embedded in noisy observations. As a prerequisite, they often require the estimation of unknown parameters (*e.g.* channel coefficients). On the other hand, estimation techniques serve to estimate unknown parameters from noisy observations, and often assume that detection-based preprocessing has been performed. There is, therefore, a close relationship between detection and estimation techniques. For this reason, we are concerned with both techniques in this thesis.

The techniques developed in this thesis pertain to two different systems: optical storage and wireless communication systems. Optical storage systems tend to have well defined channel characteristics because these characteristics are mainly defined by the optical light path and the employed storage medium, which are both manufactured within tight tolerances. For this reason, the channel estimation is comparatively unimportant and bit detection is the more challenging task. The application we consider is the two-dimensional optical storage (TwoDOS) system, which is a system using multiple-input multiple-output technology that is expected to increase the storage density with a factor of 2 and data rate with a factor of 10 [44] compared with the current blu-ray disc based third generation optical storage systems. As a background and basis of reference, we sketch in Section 1.2 the historic development of optical storage technology and the current state of the art in detection techniques for optical storage. On the other hand, in wireless communication systems, the channel characteristics are not known *a priori*, and can vary greatly because a system placed in different environments may experience completely different fading behaviors. In this sense, the estimation of channel coefficients is highly important in wireless communication systems. Therefore, we focus on developing channel estimation techniques for wireless communication systems. The application we consider is the multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) system, which has been exploited in the current Institute of Electrical and Electronics Engineers (IEEE) 802.11n wireless local area network (WLAN) standardization activities aiming to support data rates up to 540 Mb/s. As a background and basis of reference, we sketch in Section 1.3 the historic development of WLAN technology and the state of the art in channel estimation techniques for WLAN systems.

1.1.1 Challenges in TwoDOS Systems

In response to the increasing demand for storage capacity, three successive generations of optical storage systems have been developed, *viz.* 1) compact disc (CD), 2) digital versatile disc (DVD), and 3) blu-ray disc (BD) and high density DVD (HD DVD), currently competing with each other for wide adoption as the preferred third generation optical storage standard [43, 91, 143, 151]. Even though each new generation offers significant improvement in storage capacity, the growth rate of storage density in optical storage systems lags behind that of magnetic storage systems, largely due to the comparatively slow pace at which the wavelength of laser diodes and numerical aperture of laser lenses have improved. The relatively slow pace of physical improvements in optical storage systems motivates the use of advanced signal processing techniques to achieve further increase in recording density. One promising example is the recently introduced TwoDOS system [44]. Compared with conventional one-dimensional (1D) optical storage systems, the track pitch in TwoDOS is noticeably reduced and this makes it possible to record at much higher track density. This higher track density is realized by grouping a number of adjacent tracks together with no intertrack spacing, and by using a guard-band as a boundary between groups of tracks as shown in Fig. 1.2. Further, the capacity of the disc is maximized by adopting hexagonal bit-cells instead of traditional square/rectangular bit-cells (which were used in [76, 186]). Moreover, a higher data rate can be achieved by the use of parallel read-out. Therefore, TwoDOS is a good example of a system with multiple-input multiple-output technology. Even though the presence of a guard-band between groups of tracks prevents interferences from adjacent groups, the elimination of spacing between the tracks within the group results in severe intertrack interference (ITI) from bits in the neighboring tracks. In the TwoDOS system, this ITI is even more significant than the intersymbol interference (ISI), which is the main interference in conventional 1D optical storage systems. For example, in Fig. 1.2, the nearest six bit-cells and second nearest six bit-cells of the bit-cell 0 are marked as bit-cell 1 and bit-cell 2, respectively. As the magnitudes of ISI and ITI are determined by the distances relative to the bit-cell 0, the interferences from each bit-cell 1 (or bit-cell 2) are equal. Further, compared to bit-cells 2, bit-cells 1 are closer to the bit-cell 0 and thus cause larger ISI/ITI. As shown, among all the twelve interferences from bit-cells 1 and 2, only two are due to ISI and all the remaining ten are due to ITI. Thus, the overall impact due to ITI is much more on bit-cell 0 compared to that due to ISI. Therefore, it becomes very important to develop powerful two-dimensional (2D) signal processing techniques instead of 1D signal processing techniques to deal with ITI as well as ISI. However, because of the 2D nature of the system, the intrinsic complexity of high-quality receivers tends to be too high to permit cost-effective implementation at high data rates. Therefore, the scope of our research work in the TwoDOS system is the development of reduced-complexity 2D signal processing techniques to facilitate high-quality data reception at high data rates.

Early work on 2D detectors for data storage systems focused on the 2D decision feedback equalizer (DFE) [77, 186] due to its simple implementation. The 2D DFE uses the past decisions to remove ISI and ITI and thus increases the signal margin against noise. A similar detector called the pseudodecision-feedback equalizer [114] uses an iterative procedure, and uses the estimated neighboring bits from past iterations to remove interferences. Compared with the 2D Viterbi detector (VD), which is the optimal 2D detector in the presence of additive white Gaussian noise, these detectors only lead to small detection performance losses for systems with relatively small ITI. Note that in the

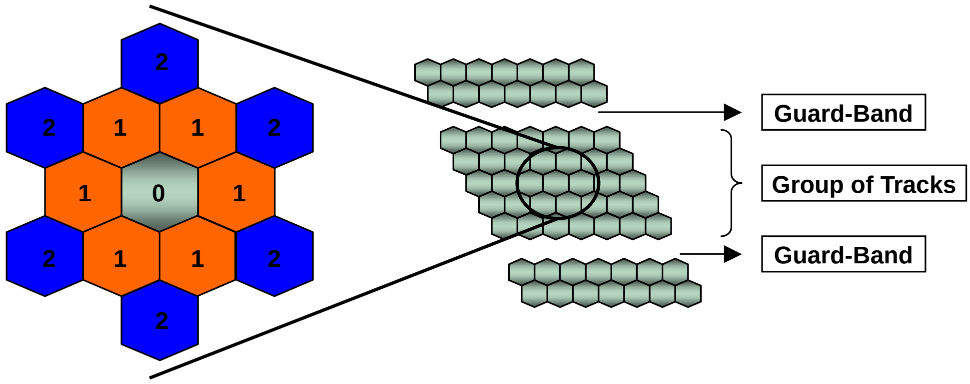

Figure 1.2: The TwoDOS format. The nearest six bit-cells and second nearest six bit-cells of the bit-cell 0 are indexed as 1 and 2, respectively.

decision process, unlike the 2D VD, the 2D DFE and the 2D pseudodecision-feedback equalizer ignore the signal energy available in ISI and ITI. For this reason, they suffer significant performance loss relative to the 2D VD for storage systems that exhibit severe ISI and ITI such as the TwoDOS system. For this reason, a great deal of attention has been paid to 2D Viterbi-like detectors due to their good detection performance even in the presence of severe ISI and ITI [93, 108, 171]. However, since the complexity of a full-fledged 2D VD grows exponentially with the channel memory and the number of tracks per group, the implementation of the full-fledged 2D VD is impractical for systems with large number of tracks per group. Some techniques [44,84,116] have been proposed to reduce the complexity of the full-fledged 2D VD. Nevertheless, there is still great potential to further reduce the complexity without incurring considerable performance degradation. Therefore, most techniques proposed in Part I of the thesis aim at the development of simple detectors with good detection performance. We also note that some work investigated the iterative detection techniques that stem from Turbo or LDPC coding [33, 169, 192] to achieve additional performance gains. However, in view of their even higher computational and storage complexity, these techniques are not considered in this thesis.

1.1.2 Challenges in MIMO-OFDM Systems

OFDM technology can deal with ISI caused by severe multipath effects and achieve a high spectral efficiency. It is adopted in current wireless local area network products, which have achieved great commercial success. The continuous demand for even higher data rates motivates the research into high-data-rate extensions for wireless local area networks. In January 2004, the IEEE set up a new task group aiming to develop the IEEE 802.11n standard. This standard is expected to support a data rate of 540 Mb/s, which is 10 times higher than that in current wireless local area network systems. Multiple-input multiple-out technology is commonly referred to as MIMO technology in wireless communications. As MIMO technology can be used to achieve two objectives: spatial diversity and space-division multiplexing, the IEEE 802.11n standard adopts MIMO-OFDM technology in order to achieve complementary benefits from both the MIMO and OFDM technologies. Note that coherent demodulation, which requires and utilizes the knowledge of channel coefficients, can achieve a 3 dB performance gain compared with differential demodulation [154]. Coherent demodulation is quite commonly used in MIMO-OFDM systems. Therefore, accurate and robust channel estimation that permits the coherent demodulation is very important in order to ensure reliable data recovery.

Early MIMO-OFDM channel estimation techniques treated channels as spatially uncorrelated (*e.g.* [17,168,174]) possibly due to the fact that early MIMO studies assumed the channels to be spatially uncorrelated (*e.g.* [68, 175]). However, in many realistic scenarios, the MIMO-OFDM channel tends to be spatially correlated, for example, due to antenna spacing constraints and limited scattering [132,166,167]. Prior knowledge of this spatial correlation in addition to frequency correlation can be exploited by using the linear minimum mean square error technique [59,133,200]. However, the complexity of the 2D linear minimum mean square error technique, which fully utilizes prior knowledge of both the channel spatial and frequency correlation, is quite high. Further, prior knowledge of the channel spatial and frequency correlation is not always available to the receiver. The least squares technique circumvents this problem but provides much poorer performance. Therefore, it is important to develop reduced-complexity, approximate linear minimum mean square error channel estimation techniques that allow good trade-off among performance, complexity, and availability of channel stochastic information (*e.g.* channel correlation or power) for MIMO-OFDM systems.

1.2 Optical Storage Systems

1.2.1 Historical Overview

Early Optical Discs

Optical recording dates back to the late 1920s. In 1927, J. L. Baird demonstrated a Phonevision system using a wax disc and an optical scanner [27]. In 1961, M. Minsky introduced a scanning microscope with a resolving power that is comparable to the standard microscope [134]. In the early 1960s, the first video-disc recorder that records video information on a standard audio long play disc was developed for the 3M-company at the Stanford Research Institute [157]. However, after a few years, this research work was abandoned largely due to the bad signal quality. In August 1973, the video long play (VLP) disc system [49] was demonstrated by Philips. In 1975, several companies (Philips, Thomson, Music Corporation of America), later joined by Pioneer, united to establish a standard for the VLP system. Further research on the optical video disc (which was later called the “laser disc”) system focused on increasing recording density [29].

CD System

The first big commercial breakthrough of optical storage systems was the CD system, which has a storage capacity of 650 MB and turns out to be simpler and more robust than its video signal predecessor [151]. The CD standard was established in 1980 by Philips and Sony. For a long period, the application of the CD system was restricted to the digital audio domain. Though a read-only digital data storage system was defined in 1985, the lack of an installed base of drives and the lack of sophisticated software to retrieve the information from the disc in a meaningful and convenient fashion deferred its application as a useful computer peripheral, until about 8 to 10 years after the introduction of the audio CD.

DVD System

The DVD system [91] has a storage capacity of 4.7 GB and was introduced in 1994. Initially, there existed two competing formats: super disc and multimedia CD. For a time, it seemed that the battle between these two formats would completely halt the development of the DVD system. But the quickly developing market of video CD sped up the cessation of the battle, and a single standard was finally agreed upon by both camps. The DVD system offers 7 times higher capacity than the CD system. As a result, MPEG2 video information and super-audio signals can be recorded on this medium. Because of the high-quality image and simple interactive functions, DVD-Video gained very quick acceptance from customers.

BD and HD-DVD Systems

BD [43, 143] is the name of the optical storage system jointly developed by 13 leading companies. The BD system has a storage capacity of 25 GB and was developed to enable recording, rewriting and playback of high-definition television. BD is also expected to create an expanded interactive environment as well as broadband content service functions [117]. Compared to the DVD system, the main physical improvement in the BD system lies in the use of a blue laser with wavelength 405 nm instead of a red laser with wavelength 650 nm, and the use of a lens with high numerical aperture (NA=0.85 instead of NA=0.60) [43]. The primary rival to BD is the high density DVD (HD DVD), which also uses a blue laser with wavelength 405 nm. HD DVD has a lower theoretical storage capacity (15 GB), but currently benefits from lower manufacturing costs. Though BD and HD DVD are currently competing with each other for wide adoption as the preferred third generation optical storage standard, it is clear that the third generation optical storage standard will adopt the blue laser to replace the red laser, which is used in the second generation optical storage standard.

1.2.2 Detection in Optical Storage Systems

Because the interferences and distortions become more severe with increase in optical recording density, more sophisticated and complicated detection techniques have been developed to ensure reliable data recovery. In this subsection, various conventional detection techniques for optical storage systems are reviewed, as a basis of reference for the 2D detection techniques developed in this thesis.

We start by introducing runlength-limited (RLL) codes, which are the most popularly used channel codes in optical storage systems. These codes have played an important role to facilitate the design of detection techniques. RLL codes emerged in the 1960s [20, 71]. They are characterized by two parameters, $(d+1)$ and $(k+1)$, which characterize the minimum and maximum number of channel bits, respectively, between consecutive transitions. The parameter d controls the minimum spacing between transitions on the medium and thus can alleviate the linear and nonlinear interactions among the data bits recorded on the optical medium. It also helps to reduce the complexity in detectors by precluding certain states and transitions [25, 71]. The parameter k limits the maximum transition spacing, which ensures that the control loops (*e.g.* timing, gain and equalization) can update frequently enough so that the loops are maintained in good condition. The parameter k also helps to reduce the path memory requirement as well as to avoid certain catastrophic error events in Viterbi-like detectors [188]. The redundancy introduced by the (d, k) constraints is measured by the code rate $R = p/q$, which specifies that groups of p data bits at the encoder input are translated into groups of q data bits at its output, with $q \geq p$. The basis on which d and k values are chosen depends on various factors such as the physical channel, desired data rate, electronics, and media noise characteristics. In practical recording systems, d is restricted to 0, 1 or 2, and k lies between 2 and 10. Current standards use the EFM code $(R = 8/17, d = 2, k = 10)$, EFMPlus code $(R = 8/16, d = 2, k = 10)$ and 17PP code $(R = 2/3, d = 1, k = 7)$ for

CD, DVD and BD, respectively [101, 102, 143].

Bit Detectors

Bit detectors make bit decisions based on single samples of the received signal. They are commonly used in CD and DVD largely due to their simple implementation and good detection performance when combined with RLL codes [58, 79, 104, 131, 142, 153]. The simplest bit detector is the threshold detector (TD) that only detects whether each sample of received signal crosses a threshold value. Intuitively, the full response equalization technique, which shapes the channel into an equalized channel free of ISI through a suitable filter called the equalizer, would seem adequate for this detector design. However, the full response equalizer causes the noise to be enhanced seriously in the frequency regions where the channel frequency response is small [24]. Since optical recording channels do not pass high-frequency components, this technique is not commonly used. The noise enhancement problem in the full response equalization technique can be largely avoided by the partial response (PR) equalization technique, which shapes the channel into a known target with controlled ISI [39]. A good match of the channel and target results in minimizing the noise enhancement. For uncoded channel input data, the target will introduce destructive ISI that is undesirable for the TD. But for the RLL coded data input with (d, k) constraints, if the target is symmetric with a length of $2d + 1$ and has monotonically decreasing positive values on either side (sufficient but not necessary condition), all the ISI induced by the target becomes constructive and can be utilized to improve the detection performance of the TD [79]. However, when the target length exceeds $2d+1$, a part of the ISI induced by the target becomes destructive. For this reason, Gopalaswamy *et al.* [79] utilized the TD and proposed a simple post-processing technique to suppress the destructive ISI components for $d = 2$ modulated optical recording channels. For $d = 1$ optical recording channels, a runlength detector (RD) was proposed to detect and correct the dominant error event of the TD, *i.e.* the $d = 1$ violations [58, 142]. Later, the missing-run detector (MRD) was proposed to also detect double bit errors [153]. It is used in cascade to the RD since the double bit-error is the dominant error event of the RD. The latest bit-by-bit detector is the "maximum-likelihood (ML) transition detector" [31]. The main idea behind this detector is using ML detection in the cases where the TD/RD/MRD are likely to be in error, while using threshold detection in the cases where the TD/RD/MRD are likely to be correct. It can be used to replace the RD/MRD, or as an additional post-processor after the RD/MRD to tackle the remaining error events.

Sequence Detectors

Unlike bit detectors whose bit decisions are each based on a single signal sample, sequence detectors make each decision based on a sequence of signal samples. The Viterbi detector (VD) belongs to the category of sequence detectors, and is widely used in magnetic storage systems because of its good detection performance. However, due to its relatively large complexity compared to bit-by-bit detectors, the VD was not a part of optical data storage systems until recently. Since the ISI gets more severe as the recording density increases such that the bit detection is hardly viable, the VD begins to replace bit detectors in the BD system and will be widely used in future optical storage systems [120].

In current magnetic and optical storage systems, the VD is used in combination with the PR equalization technique [39]. As discussed earlier, PR equalization shapes the channel into a known target with controlled ISI, which is left to be handled by a detector. A good match of the channel and the PR target results in minimizing the noise enhancement. After PR equalization, the noiseless input of the VD is $d(n) = \sum_{i=0}^{N_g-1} g_i a(n-i)$, where g_i $(i = 0, 1, \cdots, N_g - 1)$ represent the coefficients of the PR target whose length is N_g, and $a(n)$ is the channel input bit at time index n. The VD is based on a concise and convenient state transition diagram called the trellis diagram for the data detection in the presence of controlled ISI [67]. For example, the trellis diagram corresponding to a 1D system with target length N_g equal to 3 is shown in Fig. 1.3, where the '+' and '−' represent the bits '+1' and '−1', respectively. Here the trellis is assumed to start at the node S_0, and then becomes steady at instant $n = 3$ (*i.e.* $n = N_g$). The nodes in the diagram are called states. The directed transitions between the nodes are called branches. Each branch corresponds to a particular state transition at a particular time. A sequence of branches through the trellis diagram is referred to as a path. Each possible path corresponds to one input sequence and vice versa. Further, the labels (*e.g.* $--, -+, +-, ++$) of the states represent the channel memory that is associated with the paths that pass through these states. At time index n, each state consists of $N_g - 1$ bits (*i.e.* $\{\breve{a}(n-1), \breve{a}(n-2), \cdots, \breve{a}(n-N_g+1)\}$). Thus, at each time index, the trellis contains 2^{N_g-1} states. At time index n, each branch specifies the channel memory associated with the state that the branch originates from and the possible channel input bit $\breve{a}(n)$. Therefore, each branch corresponds to one possible noiseless detector input $d(n) = \sum_{i=0}^{N_g-1} g_i \breve{a}(n-i)$. For the binary channel input bit, each state possesses two incoming and two outgoing branches and thus there are totally 2^{N_g} incoming branches and 2^{N_g} outgoing branches at each time index of the trellis.

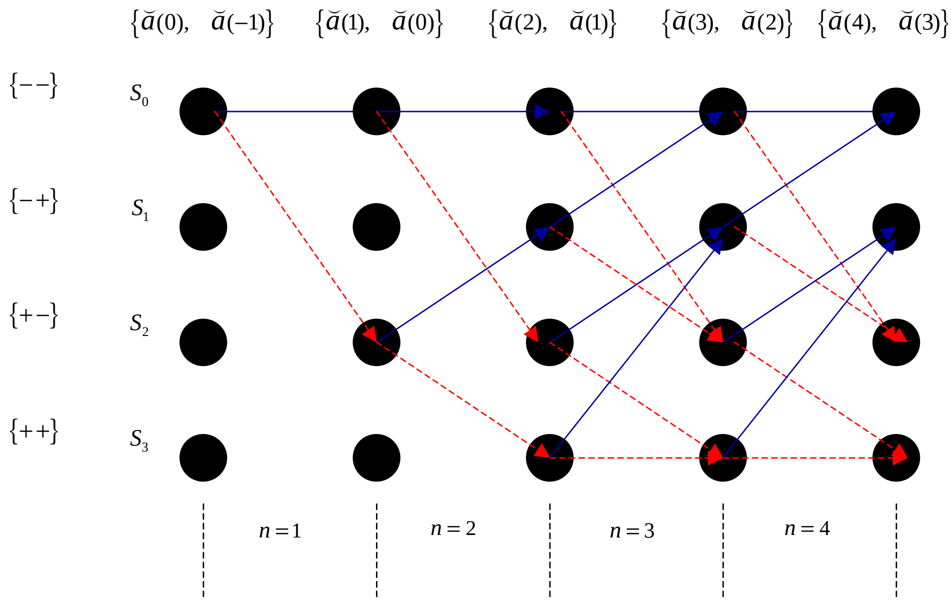

Figure 1.3: Trellis structure for a 1D channel with $N_g = 3$.

As shown in [24], searching the smallest Euclidian distance between the detector input $z(n)$ and the desired noiseless detector input $d(n)$ is optimum in the ML sense when the noise component of the detector input is white and Gaussian. Thus, we define the Euclidian distance $[z(n) - d(n)]^2$ as the branch metric for each branch, and the summation of the branch metrics associated with each path is called the path metric. Since the VD performs ML detection based on a sequence of signal samples, it chooses the path whose path metric is minimum as the most likely transmitted sequence. More specifically, the VD operates recursively as follows [67]:

1. **Initial condition**: At the end of $(n-1)$th step, each state of the trellis retains one surviving data sequence called the survivor path.

2. **Path extension**: The survivor path of each state is extended by the branches emanating from the state. Then, at the nth step, each state possesses at least one incoming branch.

3. **Path selection**: After computing all path metrics for the extended paths, for each state we select the incoming path that has the smallest path metric and discards all the other paths associated with this state. These selected paths serve as surviving paths for the next iteration.

The above recursive procedure continues until the final instant. Then, among the total 2^{N_g} survivor paths, the survivor path that is associated with the minimum survivor metric is chosen as the detected sequence of channel input bits. However, for a long sequence of bits, this technique may result in a prohibitively large path memory. In practice, at any instant n, the survivor path corresponding to each state would converge to a single path for time instants less than or equal to $n - KN_g$ for a sufficiently large positive integer K. Therefore, it is common to modify the VD by making a bit decision at time n for the bit at time $n - KN_g$. More specifically, at a certain time instant n, the "truncated" path memory stores the paths that consist of only the previous KN_g bits. The detector compares all the "truncated" path metrics for these 2^{N_g} "truncated" paths, and chooses the path that has the smallest path metrics. The bit associated with this path at time index $n - KN_g$ is released as the detected bit for the time $n - KN_g$, and will not be stored in the path memory any more. Since the path memory is greatly truncated, this modification is widely used in the practical implementation [119]. Usually, a value of K in the order of 6 to 32 suffices.

For multi-track data storage systems, 2D Viterbi-like detectors are needed to deal with severe ISI and ITI [93, 108, 171]. However, the complexity of these detectors is very high and must be reduced for a practical implementation. As the complexity of Viterbi-like detectors mainly depends on the number of states, merging some states into one superstate is a possible solution to reduce the complexity. This state merging technique has been investigated in 1D Viterbi-like detectors [25,46,116]. However, it may not suffice for 2D Viterbi-like detectors because of the even larger number of states. Therefore, in this thesis, we develop novel techniques that are more effective to reduce the complexity of 2D Viterbi-like detectors.

Some Viterbi-like detectors modify the branch metrics to outperform the traditional VD for the cases when the system has time-dependent and/or correlated input [47, 109, 199]. However, in view of their even higher computational and storage complexity, these techniques are not considered in this thesis.

1.3 Wireless Local Area Network Systems

A wireless local area network (WLAN) uses radio frequency (RF) technology or infrared light to transmit data mainly in indoor environments. The WLAN provides all the features and benefits of traditional wired local areal network (LAN) yet with greatly increased freedom and flexibility. It is becoming more and more popular, especially with the rapid emergence of portable devices such as personal digital assistants (PDAs) and laptops. Currently, there exist three major types of WLAN standards: IEEE 802.11 standards [6], European Telecommunication Standard Institute (ETSI) high performance radio local area network (HIPERLAN) standards [3], and Japanese multimedia mobile access communication (MMAC) standards [9]. Because of the large similarities between these standards, the latter two standards will not be discussed in this thesis.

1.3.1 Historical Overview

900 MHz Band WLAN

Introduced in the early 1990s, the first generation wireless local area network (WLAN) systems operated in the unlicensed 900 MHz to 928 MHz industrial scientific and medical (ISM) band can support data rates up to 500 Kb/s. These systems can achieve greatly increased freedom and flexibility compared to the LAN systems. However, the 900 MHz to 928 MHz ISM band is too crowded with other wireless communication systems. Therefore, the first generation WLAN systems does not perform well because of strong interferences coming from other wireless communication systems.

2.4 GHz Band WLAN

Continuously increasing demand for higher bit rates spurred the development of WLAN systems operating in the 2.40 GHz to 2.483 GHz ISM band. In 1997, the original IEEE 802.11 standard that supports data rates up to 2 Mb/s became available [5]. A weakness of this original standard is that it offers so much flexibility that the compatibility between products from different companies is hard to realize. Thus, this standard was later supplemented into the IEEE 802.11b standard in 1999. This extended standard can achieve data rates up to 11 Mb/s by the use of complementary code keying (CCK), which enables the coded data to be reliably detected even in the presence of strong noise and multipath interferences [7]. It is the first widely accepted WLAN standard that provides data rates comparable to wired LANs. Unlike the above two standards that utilize spread spectrum modulation techniques, the IEEE 802.11g standard utilizes the OFDM technology to achieve further enhanced data rates [8]. The IEEE 802.11g standard can support data rates up to 54 Mb/s and is backward compatible with 802.11b WLAN systems. However, 2.4 GHz band WLAN systems still suffer from many interferences coming from microwave ovens, cordless telephones, Bluetooth devices, and other wireless communication systems. For more information on these standards, we refer to [16] for a general review.

5 GHz Band WLAN

In 1997, the Federal Communications Commission (FCC) allocated unlicensed spectrum in the ISM bands of 5.150 GHz to 5.350 GHz and 5.725 GHz to 5.825 GHz. Unlike the bands employed by previous standards, these newly opened unlicensed national information infrastructure (UNII) bands does not contain many potential unwanted interferences to WLAN systems. The IEEE 802.11a standard with a maximum data rate of 54 Mb/s operates in this band [6]. However, 5 GHz band WLAN systems are more suitable for short-range transmission since signals at this higher carrier frequency band are attenuated more during the transmission compared to the above two lower carrier frequency bands.

1.3.2 Channel Estimation in WLAN Systems

As coherent demodulation, which requires and utilizes the knowledge of channel coefficients, can achieve a 3 dB performance gain compared with differential demodulation [154], it is quite commonly adopted in WLAN systems. Therefore, accurate and robust channel estimation that permits the realization of coherent demodulation is very important in order to ensure reliable data recovery. In this subsection, we will give a brief overview of channel estimation techniques for OFDM systems because the current two most important WLAN standards, *viz.* the IEEE 802.11a and 802.11g standards, are both OFDM-based. The reason for choosing OFDM is its capability to deal with ISI caused by severe multipath effects, and the high spectral efficiency afforded by allowing overlapping subcarriers. We will not cover channel estimation techniques for time-varying scenarios in this thesis as WLAN systems are usually deployed in indoor environments where the channel can be assumed to be time-invariant. The purpose of this subsection is to review various existing channel estimation techniques for OFDM systems and provide a systematic summary of research activities in this area.

Pilot-Aided Channel Estimation

Known transmitted signals that are used for channel estimation are referred to as pilots in this thesis. Channel estimation techniques that utilize these pilots and the corresponding received signals are referred to as pilot-aided channel estimation techniques [48]. Generally speaking, pilot-aided channel estimation is based on either least squares (LS) [19] or linear minimum mean square error (LMMSE) techniques [57, 124]. The essential difference between these two types of techniques is that the channel coefficients are treated as deterministic but unknown constants in the former, and as random variables of a stochastic process in the latter. Compared with LS-based techniques, LMMSE-based techniques yield better performance because they additionally exploit (and hence require) prior knowledge of the channel correlation. However, the channel correlation is sometimes not *a priori* known, which then makes LMMSE-based techniques infeasible. Further, the complexity of LMMSE-based techniques is normally higher than that of LS-based techniques.

In OFDM systems, we can divide the pilot-aided channel estimation techniques into three categories: frequency-domain techniques, time-domain techniques, and discrete Fourier transform (DFT)-based techniques.

- Frequency-domain techniques treat the frequency-domain channel coefficients as the parameters to be estimated. They are the most straightforward techniques because the knowledge of frequency-domain channel coefficients is ultimately required to permit coherent demodulation. However, the frequency-domain LS technique gives the poorest performance, and the frequency-domain LMMSE technique has the highest complexity, among all the pilot-aided techniques discussed here.

- Time-domain techniques treat the time-domain channel coefficients as the parameters to be estimated and the estimated time-domain channel coefficients are finally transformed into the frequency-domain ones.

  – The time-domain LS-based technique is commonly referred to as the time-domain maximum likelihood (ML) technique [52] and it always performs better than the frequency-domain LS technique. This is because a portion of noise is implicitly ignored, resulting from the fact that fewer channel coefficients are required to be estimated in the time domain.

- The time-domain LMMSE technique usually has much lower complexity compared to the frequency-domain LMMSE technique because the number of time-domain channel coefficients is usually much smaller than that of the frequency-domain. Further, the time-domain LMMSE technique can achieve the same performance as the frequency-domain LMMSE technique because the channel correlations in the time domain and frequency domain are interchangeable.

A disadvantage of both the time-domain ML and LMMSE techniques is the requirement of prior knowledge of the channel length.

- DFT-based techniques treat the frequency-domain channel coefficients as the parameters to be estimated. This is their main difference from time-domain techniques. Unlike frequency-domain techniques that also treat the frequency-domain channel coefficients as the parameters to be estimated, DFT-based techniques transform the estimated channel coefficients from the frequency domain to the time domain, where the noise filtering process is performed, and finally back to the frequency domain by the use of inverse DFT (IDFT) and DFT operations, respectively. By assuming that the temporal span of the channel is concentrated over a small number of coefficients, the noise in the coefficients beyond the channel length is removed in the time domain and this results in a performance improvement. DFT-based techniques do not require prior knowledge of the channel length (but estimate the channel length as a part of the estimation procedure), and have been widely used in OFDM systems because of the good trade-off between performance and complexity [19, 56].

Decision-Directed Channel Estimation

A possible way to solve the spectral efficiency loss problem in pilot-aided techniques involves the use of decision directed (DD) techniques [69]. The principle of these techniques is to utilize the already detected channel input bits from a coarsely estimated channel to update the channel estimation. The main drawback of DD techniques is the widely known error prorogation problem whose effect will increase with the size of the signal constellation. This problem can be alleviated by the use of detected data symbols that are already corrected with the use of the error correction codes (ECC), at the cost of introducing some delays in the updating process. In fact, the principles of DD techniques and pilot-aided techniques are not distinctive since they both utilize the known (or already detected) channel input bits to assist the channel estimation. In this sense, we will not treat them separately in this thesis and both techniques are referred to as pilot-aided channel estimation techniques.

Blind Channel Estimation

Blind channel estimation techniques [53,83] utilize the received signals and the stochastic information (*e.g.* second order statistics) of transmitted and received signals, to estimate the channel coefficients. Compared with pilot-aided techniques, blind techniques save on the use of pilots and can thus increase the spectral efficiency. However, blind techniques require prior knowledge of stochastic information of the transmitted and received signals. Further, they always result in poorer performance compared with pilot-aided techniques.

The concept of "blind" estimation/equalization techniques was first introduced in the seminal work of Sato [161] for the linear adaptive equalizer. Ever since, the blind equalization problem has received great attention [22,23,75,107,152,181]. Initially, blind channel estimation techniques were based on higher order statistics [54, 70, 74, 146, 165, 184] to estimate single-input single-output (SISO) channels. These techniques require a large number of data samples to estimate the higher order statistics and thus result in high computational complexity. The recognition that phase information can be extracted based on the cyclostationary properties of single-input multiple-output (SIMO) [178] or oversampled systems [179] motivates blind channel estimation techniques based on second order statistics [127, 180]. In OFDM systems, the cyclostationarity introduced by the cyclic prefix, which is a repeat of the end of the OFDM symbol at the beginning of each symbol, can also be exploited to replace the cyclostationarity in SIMO or oversampled systems [85]. Note that blind techniques always identify the channel up to a scalar ambiguity. Therefore, some pilots are exploited to remove this ambiguity. Moreover, pilots can be used to increase the convergence rate of blind channel estimation techniques. The corresponding techniques are referred to as semi-blind techniques and were originally proposed in [149]. In general, blind techniques in OFDM systems that utilize the cyclostationarity either from SIMO/oversampled outputs or the cyclic prefix can be divided into the following four categories.

- Noise Subspace Techniques: These techniques exploit the low-rank structure of the autocorrelation of received signals by dividing the column space of the received vectors into signal and noise subspaces [139]. Due to the orthogonality of signal and noise subspaces, the channel coefficients can be estimated in a closed form by minimizing a quadratic cost function with certain constraints on the channel. Noise subspace techniques that utilize the cyclostationarity from SIMO/oversampled outputs can be found in [14, 159], and those that utilize the cyclic prefix can be found in [30,85,98,141]. The main advantage of the noise subspace techniques is the closed form solution for channel estimation. However, they are relatively computationally complex because they make use of eigenvalue decomposition.

- Cross-Relation based Techniques: Unlike noise subspace techniques, the cross-relation based techniques do not require the stochastic information of channel input signals and thus impose relaxed requirements on the channel input signals (*e.g.* short input data sequences). They are based on the fact that for a noiseless SIMO system, all the subchannel outputs can be matched pairwise so that $N_r(N_r-1)/2$ zero outputs are obtained, where N_r is the number of receive antennas. Note that these techniques cannot utilize the cyclostationarity introduced by the cyclic prefix and are thus only applicable in SIMO/oversampled OFDM systems [189]. However the assumption of a noiseless SIMO system limits the application. Moreover, the use of receiver diversity (or oversampling) increases the complexity of the receiver.

- Autocorrelation based Techniques: These techniques are based only on the autocorrelation matrix of the received OFDM symbols [86, 140]. The "direct based technique" that directly utilizes part of the first column of estimated autocorrelation matrix is the simplest technique in this category [140]. Its performance can be improved by the "Cholesky decomposition based technique" that exploits a larger portion (*i.e.* one submatrix) of the estimated autocorrelation matrix [140]. To further improve the performance, a "switch based technique" that performs average processing in the submatrices of the estimated autocorrelation matrix is proposed in [86]. Note that the AWGN only affects diagonal elements of the autocorrelation matrix. As these diagonal elements are not utilized for estimation, the autocorrelation based techniques are robust to AWGN. However, the performance of these techniques is highly dependent on the accuracy of the estimated autocorrelation matrix. Further, the frequency-domain channel input signals are assumed to be independent and identically distributed. Therefore, application of these autocorrelation based techniques is quite limited.

- Maximum Likelihood (ML) based Techniques: These techniques jointly detect the channel input signals and estimate the channel [144]. The main advantage is the ability to estimate the channel from only a single received symbol in the SISO-OFDM system. However, the main problem of these techniques is the large computational complexity required for the identification of both the channel input signals and channel coefficients.

1.3.3 Angle-Domain MIMO Channels

The previous subsection has introduced existing channel estimation techniques for OFDM systems. In this subsection, we introduce the angle domain, which motives the development of our novel channel estimation techniques to be discussed in Part II of the thesis. A typical MIMO channel is conceived as the unique link between the transmitted and

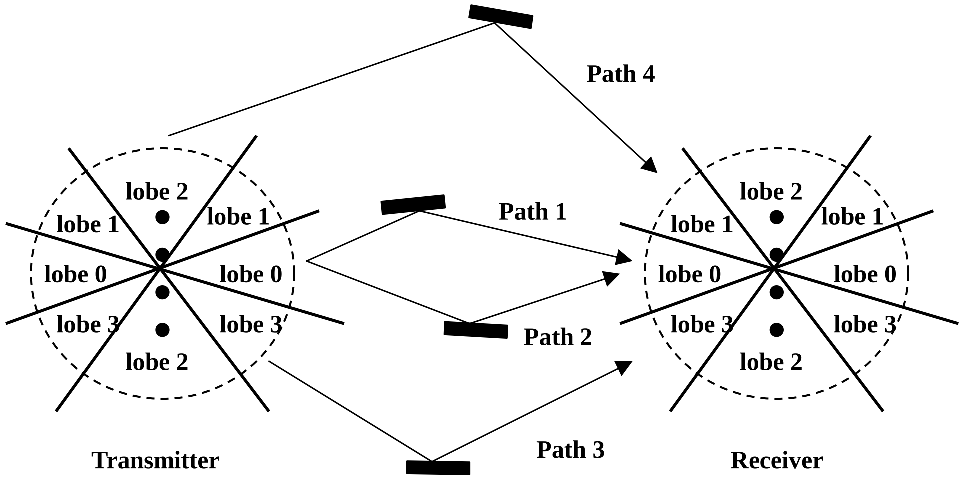

Figure 1.4: A schematic angle-domain representation of MIMO channel with 4 transmit and 4 receive antennas.

noiseless received signals, and referred to as the *array-domain* channel. Almost all the previous channel estimation techniques are developed for estimating this array-domain channel (*e.g.* [17, 168, 173]). However, the channel matrices defined in the array domain may not be the right level of abstraction from the view of design and analysis of spatially correlated MIMO systems. For example, the capacity of MIMO systems is assessed by looking first at the rank, and then the condition number of channel matrices in the array domain. However, this capacity assessment is not straightforward and one may want to abstract the channel matrices into a high level in terms of spatially resolvable paths [182]. The angle-domain representation of MIMO channels is based on the this idea and uses beamforming patterns with different main lobes to characterize the physical propagation environment [162, 182]. For a MIMO system with N_t transmit and N_r receive antennas, the beamforming patterns have N_t transmit lobes and N_r receive lobes. A pair of transmit and receive lobes forms one angle-domain bin and thus the angle domain is partitioned into $(N_t \times N_r)$ angle-domain bins. For example, as shown in Fig. 1.4, the transmit lobe 0 together with receive lobe 0 corresponds to the angle-domain bin $(0, 0)$. Then, multiple unresolvable physical paths (*e.g.* path 1 and 2) that occur in the angle-domain bin $(0, 0)$ can be approximately aggregated into one resolvable path, and the paths from other directions (*e.g.* path 3 and path 4) will have little effect on this resolvable path because they originate or end at other lobes. Consequently, different physical paths approximately contribute to different angle-domain bins, and the channel coefficients in different angle-domain bins can be assumed to be approximately spatially uncorrelated. Further, when some angle-domain bins contain few physical paths due to limited scattering, the corresponding channel coefficients should approach zero. Based on these two special properties for the angle-domain channel coefficients, we focus on developing several novel angle-domain channel estimation techniques for MIMO-OFDM systems. To the best of our knowledge, this is the first work to systematically investigate angle-domain channel estimation techniques.

1.4 Organization of the Thesis

This thesis is divided into two parts. In Part I, from Chapter 2 to Chapter 5, we concentrate on the TwoDOS system where the reduced-complexity detector design is the main concern. In Part II, from Chapter 6 to Chapter 9, we focus on MIMO-OFDM systems where the reduced-complexity channel estimator design is of particular interest. The relationships within the two parts are indicated by filled arrows in Fig. 1.5. Though the treatments in these two parts are self-contained, the parts are related, as indicated by the hollow arrows. First, the TwoDOS system and MIMO-OFDM system share the same block diagram as shown in Fig. 1.1. The source coder/decoder, channel coder/decoder, equalizer, and detector blocks have similar functions in both systems. Thus, most of techniques employed in one system are in principle applicable in the other system. Second, between the detection and estimation techniques, there exists a symbiotic relationship as discussed earlier. Further, as both the detection and estimation techniques belong to the category of statistical signal processing techniques, they have many points in common. For example, the mean square error is one of the most important performance measures in both techniques.

Chapter 2 describes the modeling of the TwoDOS system. It develops an easily computable 2D symbol response model by using the 1D Hankel transform technique in the

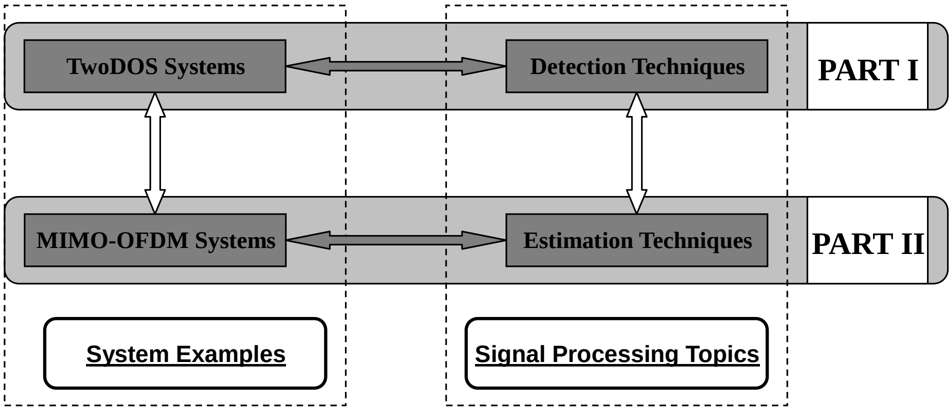

Figure 1.5: Structure of the thesis.

presence of domain bloom and/or transition jitter. Finally, an approximate model for the domain bloom and transition jitter is developed.

Chapter 3 presents prefiltering techniques to reduce the complexity of the 2D VD. First, for a given target, we design a general 2D minimum mean square error (MMSE) equalizer. Next, we jointly design the equalizer and target based on the MMSE technique. Then, we propose a novel technique which converts the 2D target design problem into a 1D problem. Also, a computationally efficient analytical technique is developed to evaluate the detection performance for different targets.

Chapter 4 introduces a new target that is constrained to have the causal ITI. Based on this target, we develop a quasi-1D VD that can noticeably reduce the complexity of full-fledged 2D VD. The performance of this newly proposed detector is simulated and the factors that degrade its performance are also investigated.

Chapter 5 develops a new detector called FDTS/DF-VD. The full-fledged 2D VD and quasi-1D VD are its special cases. Then, the reduced-complexity FDTS/DF-VD (RFDTS/DF-VD) with negligible performance loss is presented. In addition, several new targets suitable for the FDTS/DF-VD and RFDTS/DF-VD are proposed.

Chapter 6 provides technical background on pilot-aided channel estimation techniques for OFDM systems. It introduces the basic concepts of OFDM systems and gives an overview of existing channel estimation techniques.

Chapter 7 represents MIMO-OFDM systems in the angle domain, which inspires our work for the development of MIMO-OFDM channel estimation techniques. Additionally, we address pilot design techniques that facilitate direct implementation of angle-domain channel estimation techniques.

Chapter 8 proposes channel instantaneous power based angle-domain channel estimation techniques that are applicable when the channel stochastic information is not available to the receiver. Two types of techniques, *viz.* the most significant taps selection and channel instantaneous power based approximated LMMSE techniques, are described. A unified approach is also developed to analyze the performance of these techniques.

Chapter 9 develops several reduced-complexity, suboptimal, approximated LMMSE-based channel estimation techniques in the angle domain. Their performances and complexity are also analyzed in this chapter.

Chapter 10 draws conclusions and points out relevant areas for future research.

1.5 Major Contributions of the Thesis

We explore reduced-complexity signal processing techniques for two different multiple-input multiple-output systems in this thesis. In Part I of the thesis, we concentrate on the development of reduced-complexity detection techniques for the TwoDOS system. Though this work is of special interest to the TwoDOS system whose bit-cells are arranged in a hexagonal structure, most of our proposals can also be applied to data storage systems with square or rectangular bit-cells. In Part II of the thesis, we focus on the design of reduced-complexity channel estimation techniques for MIMO-OFDM systems. These techniques perform noise filtering in the angle domain, where the channel model lends itself to a physical interpretation. The proposed angle-domain estimation techniques are not restricted to MIMO-OFDM systems and are applicable to other MIMO wireless communication systems as well.

In Part I of the thesis, we develop several novel techniques to ensure the high-speed implementation of the TwoDOS system. The main contributions are as follows.

- We present a linear channel model which incorporates the linear pulse modulator for the TwoDOS system. Though this linear channel model is constructed solely from a signal processing point of view, it reflects physical reality since we show that it is consistent with the linear part of the physical channel model constructed by Coene [45] based on the scalar diffraction theory. Further, we modify the channel model to include domain bloom and transition jitter.

- We use the 1D Hankel transform to develop a computationally simple technique for generating received signals in the presence of additive noise, domain bloom and transition jitter. The computational advantage resulting from the use of the Hankel transform instead of the 2D Fourier transform is shown to be quite significant.

- We develop an approximated model for domain bloom and transition jitter to simplify the received signal generation process for the TwoDOS system. This is very important for doing simulation studies over a large number of data bits since the channel matrix recomputation, which has a large computational complexity at each time index, is avoided during the received signal generation process.

- We design a general 2D MMSE equalizer, for a given 2D target, for the TwoDOS system. This equalizer is much more widely applicable than the one of [156] in that it can deal with correlated data, colored additive noise, domain bloom, and transition jitter.

- We propose a novel technique that can be used for any 2D target design by converting the 2D target design problem into a 1D problem. In particular, this technique can significantly reduce the computational complexity for the design of symmetric constrained targets.

- We present a theoretical framework for the analysis of 2D equalizer, target and VD for the TwoDOS system. Moreover, based on a reasonable approximation, we noticeably reduce the computational complexity of the performance analysis.

- We develop a novel quasi-1D VD by constraining the target to have causal ITI. This quasi-1D VD uses a computationally efficient technique whose complexity grows only linearly with the number of tracks. We also analyze factors that degrade the performance of this detector.

- We propose a generalized 2D VD called the FDTS/DF-VD. The conventional 2D full-fledged VD, QR detector, and our proposed quasi-1D VD are all special cases of this detector. The reduced-complexity FDTS/DF-VD, which results in negligible performance loss relative to the full-complexity FDTS/DF-VD, is also developed. Additionally, several novel targets that are specific for this type of detector are addressed. Our simulation results indicate that by judiciously choosing the target and number of tracks under consideration in the FDTS/DF-VD, we can develop a reduced-complexity 2D Viterbi-like detector that facilitates the high-speed TwoDOS implementation without paying a large penalty in detection performance.

In Part II of the thesis, to the best of our knowledge, we are the first to systematically investigate channel estimation techniques in the angle domain. The proposed angle-domain techniques can achieve significant complexity reduction compared to the 2D LMMSE technique, which is seen as the optimum estimation technique in MIMO-OFDM systems, while maintaining good estimation performance. The main contributions are as follows:

- We develop channel instantaneous power based angle-domain channel estimation techniques for MIMO-OFDM systems. In particular, we classify the angle domain in MIMO-OFDM systems as the angle-time and angle-frequency domain. We investigate estimation techniques for both these domains and find that the proposed techniques perform especially well in the angle-time domain.

- We propose a unified approach to analyze the performance of channel instantaneous power based angle-domain channel estimation techniques in terms of mean square error (MSE). Based on this approach, we develop a simple way to compare the performances of different channel instantaneous power based angle-domain estimation techniques with the help of the first derivative test [11]. We show that setting the threshold to be two times the noise variance suffices for the channel instantaneous power based angle-domain channel estimation techniques to perform better than the conventional LS technique at each signal-to-noise ratio (SNR) for various IEEE 802.11 TGn channel models.

- We develop reduced-complexity LMMSE-based channel estimation techniques in both the angle-time and angle-frequency domains for MIMO-OFDM systems. The choice of LMMSE-based techniques is largely dependent on the amount of channel stochastic information (*e.g.* channel correlation or power) available to the receiver.

- We analyze the performance and complexity of various channel estimation techniques for MIMO-OFDM systems. Our simulation results show that the proposed quasi one-dimensional (Q1D) LMMSE technique based on the angle-frequency domain correlation can achieve performance similar to the 2D LMMSE technique for typical MIMO-OFDM channel models, yet has significantly lower complexity.

- We tailor pilots to facilitate direct implementation of angle-domain channel estimation techniques. This makes all our proposed angle-domain channel estimation techniques flexible in implementation. These estimation technique can either use conventional array-domain estimators as the coarse estimators and perform post-processing in the angle domain, or use the specifically designed pilots for direct implementation.

The above contributions have resulted in 9 publications, which are listed on Author's Publications.

Part I

TwoDOS system

In Part I of the thesis, we concentrate on developing detection techniques for multiple-input multiple-output systems. As received signals of multiple-input multiple-output systems may contain much more severe interferences compared to single-input single-output systems, it becomes very important to develop advanced detection techniques for reliable data recovery. On the other hand, the complexity of these detection techniques should be kept as low as possible to permit high-speed implementation. In particular, we focus on coherent detection techniques, which require and utilize the prior knowledge of channel coefficients. Note that prior knowledge of these coefficients is assumed to be available to the receiver in this part of the thesis. This assumption is realistic to optical storage systems since channel characteristics in these systems are mainly defined by the optical light path and the employed storage media, which are both manufactured within tight tolerances.

The application we consider is the two-dimensional optical storage (TwoDOS) system, which takes the form of a multiple-input multiple-output system and is expected to increase the storage density with a factor of 2 and data rate with a factor of 10 [44] compared with the current blu-ray disc based one-dimensional (1D) optical storage systems. The higher storage density mainly results from the use of smaller track pitch, which makes it possible to record at much higher track density in the TwoDOS system. This small track pitch is realized by grouping a number of adjacent tracks together with no intertrack spacing, and by using a guard-band as a boundary between groups of tracks. Due to the absence of intertrack spacing within one group in the TwoDOS system, both the intersymbol interferences (ISI) and intertrack interferences (ITI) are quite severe in the received signals. The 2D Viterbi detector (VD) seems ideally suited for the TwoDOS system due to its good detection performance [93, 108, 171]. However, since its complexity grows exponentially with the channel memory and number of tracks per group, its implementation is impractical for systems with large numbers of tracks per group. Therefore, the scope of this part of the thesis covers the development of reduced-complexity 2D Viterbi-like detection techniques.

The organization of this part is as follows. Chapter 2 introduces the TwoDOS system with a detailed channel model in the presence of additive noise and nonlinear channel distortions. This chapter states the main definitions and assumptions in Part I. The channel model developed in this chapter serves as a platform to design and evaluate various reduced-complexity detection techniques to be discussed in the subsequent chapters in Part I. In general, the reduction of complexity of the 2D VD is realized in two directions. The first is to reduce the channel memory, which will be discussed in Chapter 3. The other is to divide the full-fledged 2D VD into a set of sub-2D VDs, each dealing with a smaller number of tracks. More specifically, Chapter 3 presents equalization and target design techniques to reduce the channel memory. The results show that the newly proposed "2D monic constraint" target can be used for the TwoDOS system to achieve reliable detection performance. Then, another new target that is constrained to have causal ITI is introduced in Chapter 4. With this new target, a quasi-1D VD is developed to significantly reduce the complexity of the full-fledged 2D VD. A more general 2D detector is proposed in Chapter 5. The full-fledged 2D VD and quasi-1D VD are its special cases. This general detector has the flexibility to design systems with varying performance and complexity. Further, it can also be seen as the generalized version of the QR detector [51], which can be used in multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems that will be discussed in Part II of the thesis.

Chapter 2

TwoDOS Channel Model

2.1 Introduction

We start, in this chapter, with a description of two-dimensional optical storage (TwoDOS) system for the purpose of stating the main definitions and assumptions relevant to this work. The channel model developed in this chapter will be used in the simulations done for evaluating the various techniques to be discussed in the remaining chapters of Part I of the thesis. The linear two-dimensional (2D) channel model for TwoDOS was first introduced in [135] by taking the 2D inverse Fourier transform of the 2D modulation transfer function (MTF), resulting in 2D impulse response of the channel. However, this channel model may not be accurate enough unless the recording density is sufficiently high since this model does not account for the presence of the linear pulse modulator in the write circuit. In this chapter, we present the linear channel model which incorporates the 2D impulse response of the channel and the linear pulse modulator. The resulting response of the channel is referred to as the 2D symbol response. More importantly, we show that by exploiting the radial symmetry property of the 2D symbol response, the 2D symbol response can be efficiently calculated by using a one-dimensional (1D) Hankel transform approach [150]. We also develop the 2D symbol response model in the presence of domain bloom and transition jitter, and show that the Hankel transform can be used to develop a computationally simple approach for generating received signals in the presence of domain bloom and/or transition jitter. Finally, an approximate model for domain bloom and transition jitter is constructed to further simplify the signal generation process for the TwoDOS system with such domain bloom and transition jitter. This approximation is valid when the domain bloom and transition jitter are sufficiently small and the recording (linear) density is sufficiently high.

This chapter is organized as follows. Section 2.2 introduces the linear channel model for TwoDOS in the presence of additive white Gaussian noise. The linear channel model is extended to include the nonlinear distortions, *i.e.* domain bloom and transition jitter, in Section 2.3. An approximated model that can significantly reduce the complexity of the received signal generation process is developed. Finally, Section 2.4 concludes this chapter.

2.2 Linear Channel Model

2.2.1 Symbol Response for A Single Spot

In the TwoDOS system, we classify the signals and channels in two domains: the spatial-temporal domain and spatial-frequency domain. Each domain contains two dimensions, *i.e.* the spatial and temporal dimensions in the spatial-temporal domain and the spatial and frequency dimensions in the spatial-frequency domain. These two domains are related by the 2D Fourier transform. In general, a multi-spot light beam is used for parallel read-out in the TwoDOS system. For a linear read-out, each spot can be characterized by the 2D modulation transfer function (MTF) given by the Braat-Hopkins formula [28]:

$$\mathbf{F}_{sf}(\phi,\rho) = \begin{cases} \frac{2}{\pi}\left[\arccos(\frac{\rho}{\rho_c}) - (\frac{\rho}{\rho_c})\sqrt{1-(\frac{\rho}{\rho_c})^2}\right], & \text{when } \rho \le \rho_c, \\ 0, & \text{else,} \end{cases} \tag{2.1}$$

represented in the 2D spatial-frequency domain, where ρ is the spatial angular frequency, $\rho_c = 2NA/\lambda$ is the angular cut-off frequency, λ is the wave length of laser diodes, NA is the numerical aperture of lens, and ϕ is the azimuth angle in the 2D spatial-frequency plane. The 2D linear pulse modulator is defined as

$$\mathbf{C}_s(\theta,r) = \begin{cases} 1, & \text{when } r \le R, \\ 0, & \text{else,} \end{cases} \tag{2.2}$$

in the 2D spatial-temporal domain, where r and θ are radial and angular coordinates, respectively, in the 2D spatial-temporal plane and R is the radius of the pit hole.

We use polar coordinates in deriving the 2D channel symbol response in view that the MTF and linear pulse modulator are both radially symmetric. Let $\mathbf{C}_{sf}(\phi,\rho)$ represent the linear pulse modulator in spatial frequency domain. By taking 2D Fourier transform on both sides of (2.2), we obtain

$$\mathbf{C}_{sf}(\phi,\rho) = 2\pi R\frac{J_1(\rho R)}{\rho} \tag{2.3}$$

where $J_1(x)$ is the Bessel function of the first kind and first order. Then, the 2D symbol response in the spatial-frequency domain is given by

$$\mathbf{H}_{sf}(\phi,\rho) = \mathbf{C}_{sf}(\phi,\rho)\mathbf{F}_{sf}(\phi,\rho). \tag{2.4}$$

Consequently, the 2D symbol response in the spatial-temporal domain, $\mathbf{H}_s(\theta,r)$, can be obtained by taking the 2D inverse Fourier transform of $\mathbf{H}_{sf}(\phi,\rho)$ as

$$\mathbf{H}_s(\theta,r) = \frac{1}{(2\pi)^2}\int_0^{\rho_c}\int_0^{2\pi} \rho\mathbf{H}_{sf}(\phi,\rho)\exp[j\rho r\cos(\theta-\phi)]d\phi d\rho. \tag{2.5}$$

It should be noted that the above symbol response considers only linear intersymbol interference (ISI) and intertrack interference (ITI) and it is developed purely from a signal processing point of view. In the next section, we will modify this to include nonlinear distortions also. A rigorous derivation of the channel model based on physical generation of the signal, including linear and nonlinear ISI and ITI, can be found in [45]. However, as we show below, our model in (2.4) is consistent with the model in [45] when only linear ISI is considered. Based on the general formulation in [45], we can write the linear symbol response $H_{\boldsymbol{x}}(x,y)$ in spatial-temporal domain as ((x,y) being Cartesian coordinates)

$$H_{\boldsymbol{x}}(x,y) = \iint |p(\vec{x},\vec{y})|^2\, C_s(\vec{x}-x,\vec{y}-y)d\vec{x}d\vec{y} \tag{2.6}$$

where $p(\vec{x},\vec{y})$ is the spot function and $C_s(\vec{x}-x,\vec{y}-y)$ is the window function centered at (x,y). Because $C_s(x,y)$ is circularly symmetric in TwoDOS, $C_s(\vec{x}-x,\vec{y}-y) =$

$C_s(x-\vec{x}, y-\vec{y})$. Therefore, (2.6) implies that $H_{sc}(x, y)$ is the 2D convolution between $|p(x, y)|^2$ and $C_s(x, y)$. Since $\mathbf{F}_{sf}(\phi, \rho)$ is the 2D Fourier transform of $|p(x, y)|^2$, the signal processing based model given in (2.4) is consistent with the physical model given in (2.6).

2.2.2 1D Hankel Transform Approach

While studying TwoDOS channels with time-varying characteristics (*e.g.* media noise dominated channels), it may be necessary to recompute the channel symbol response very often. In such situations, directly using the 2D Fourier transform approach may not be preferable because of its computational complexity. We now show that the 2D Fourier transform can be replaced by the 1D Hankel transform in TwoDOS, thus resulting in significant computational savings in the calculation of $\mathbf{H}_s(\theta, r)$.

From (2.1) and (2.3) we find that since $\mathbf{F}_{sf}(\phi, \rho)$ and $\mathbf{C}_{sf}(\phi, \rho)$ are both radially symmetric, the symbol response $\mathbf{H}_{sf}(\phi, \rho)$ in the spatial-frequency domain is also radially symmetric, *i.e.* $\mathbf{H}_{sf}(\phi, \rho) = \tilde{\mathbf{H}}_{sf}(\rho)$. Then, (2.5) can be simplified as

$$\mathbf{H}_s(\theta, r) = \frac{1}{2\pi} \int_0^{\rho_c} \rho \tilde{\mathbf{H}}_{sf}(\rho) J_0(\rho r) d\rho \tag{2.7}$$

where $J_0(x)$ is the Bessel function of the first kind and zero order. Observe that $\mathbf{H}_s(\theta, r)$ is also radially symmetric, *i.e.* $\mathbf{H}_s(\theta, r) = \tilde{\mathbf{H}}_s(r)$. Equation (2.7) shows that $\mathbf{H}_s(\theta, r)$ is the inverse Hankel transform of $\tilde{\mathbf{H}}_{sf}(\rho)$ [150]. Thus, by taking advantage of the circular symmetry of the TwoDOS channel, the 2D Fourier transform can be reduced to a 1D Hankel transform. This is important since, as we show below, the Hankel transform approach is a computationally efficient means to compute $\mathbf{H}_s(\theta, r)$, compared with 2D Fourier transform. The fast Hankel transform can be computed with a complexity of about $O(N_s \log_2 N_s)$, where N_s denotes the number of data points in each dimension [82, 115]. Compared to this, the complexities of traditional 2D discrete Fourier transform and 2D fast Fourier transform are $O(N_s^4)$ and $O(N_s^2 \log_2 N_s)$, respectively [55]. Thus, the fast Hankel means of computing the channel symbol response is indeed a computationally efficient approach.

Note that the use of the Hankel transform to compute the symbol response for a single spot requires radial symmetry in both the MTF and the linear pulse modulator. However, some radial asymmetry might arise in practice due, for example, to radial wobble resulting from the unbalance of the disc. In such cases, the radially asymmetric MTF may be factorized into a radially symmetric component and a simple 2D transformation matrix. A radially asymmetric linear pulse modulator can be factorized similarly. Then, the Hankel transform is still applicable here to deal with these radially symmetric factors of the MTF and the linear pulse modulator, and considerable complexity savings can still be expected.

2.2.3 Discrete-Time Linear Channel Model

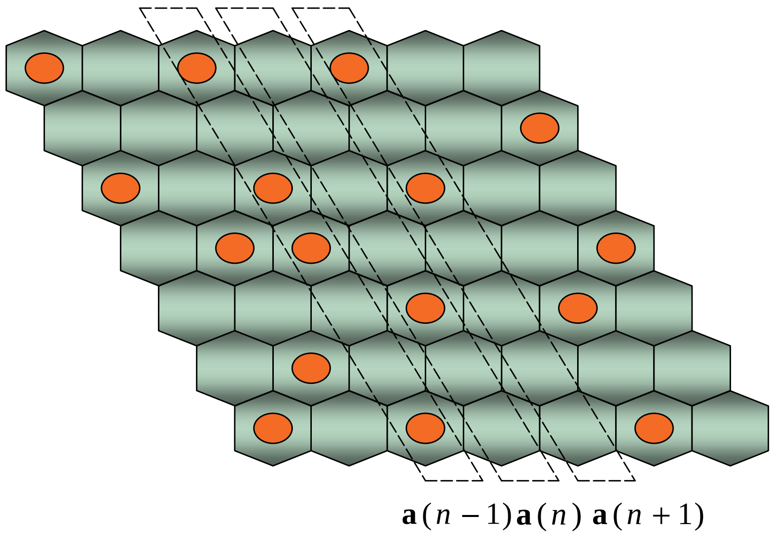

Figure 2.1: Arrangement of bit-cells in the TwoDOS system. Bit-cells with 'circles' inside indicate '+1' (*i.e.* pits) and the bit-cells without circles indicate '−1' data bits.

Due to the hexagonal structure of bit-cells, it was shown in [44] that if the sampling rate is larger than $\sqrt{3}\,(2\text{NA}/\lambda)$, the discrete-time sampled received signals in the optical read-out provide sufficient statistics for optimum detection. Therefore, these discrete-time received signals can be thought of as resulting from the recorded bits through a discrete-time channel and noise. Such a discrete-time channel is simpler than the continuous-time channel discussed in the previous two subsections, and thus facilitates mathematical analysis. Note that $\tilde{\mathbf{H}}_s(r)$ characterizes a single spot only. The discrete-time channel symbol response of the 2D system is given by $\underline{\mathbf{H}} = [\mathbf{H}_0,\ \mathbf{H}_1,\ \cdots,\ \mathbf{H}_{N_h-1}]^T$, where N_h is the temporal span of the channel, $\mathbf{H}_k$ is an $N_r \times N_r$ matrix, and N_r denotes the number of tracks forming one group. Fig. 2.1 shows the arrangement of bit-cells in the TwoDOS system. The dashed slanted rectangular boxes indicate the array of bit-cells onto which the parallel read-out laser beam is focused at each time instant. Considering the hexagonal structure of each bit-cell, the (i, j)th element of $\mathbf{H}_k$ can be obtained as

$$H_k(i,j) = \tilde{\mathbf{H}}_s\left(\sqrt{x_{ij,k}^2 + y_{ij}^2}\, T\right), i,j = 1,2,\cdots,N_r \tag{2.8}$$

where $x_{ij,k} = (i-j)\cos 60° + t_k$, $y_{ij} = (j-i)\sin 60°$, $t_k = (k - \frac{N_h+1}{2})T$ is the spatial difference along the track (for odd N_h) and T is the center-to-center distance between adjacent bits. In fact, $H_k(i,j)$ can be considered as the interference of jth track to ith track. The received signal vector resulting from the parallel read-out at time index 'n' is given by

$$\mathbf{z}(n) = \sum_{k=0}^{N_h-1} \mathbf{H}_k \mathbf{a}(n-k) + \boldsymbol{\theta}(n) \tag{2.9}$$

where $\mathbf{z}(n) = [z_1(n),\ z_2(n),\ \cdots,\ z_{N_r}(n)]^T$, $\mathbf{a}(n) = [a_1(n),\ a_2(n),\ \cdots,\ a_{N_r}(n)]^T$, $\boldsymbol{\theta}(n) = [\theta_1(n),\ \theta_2(n),\ \cdots,\ \theta_{N_r}(n)]^T$, $z_i(n)$ denotes the received signal component from the ith track, $a_i(n) \in \{-1,\ 1\}$ denotes the channel input bit written on the ith track, and $\theta_i(n)$ denotes the noise picked up from the ith track, for $i = 1, 2, \cdots, N_r$.

Fig. 2.2 shows the discrete-time channel model of the TwoDOS system with the partial response (PR) equalizer and Viterbi detector (VD). In the figure, $\mathbf{G}_k$ $(k = 0, 1, \cdots, N_g-1)$ and $\mathbf{W}_k$ $(k = 0, 1, \cdots, N_w - 1)$ represent $N_r \times N_r$ coefficient matrices of the PR target and equalizer, respectively, and m_0 and D denote the delay (in number of bits) from the channel input to the equalizer output and that from the detector input to the detector output, respectively.

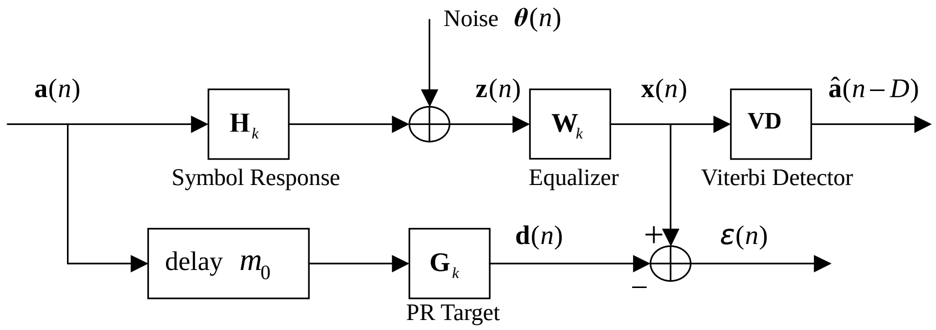

Figure 2.2: Discrete-time channel model of the TwoDOS system with the PR equalization and VD.

2.3 Channel Model with Nonlinear Distortions

2.3.1 Effect of Domain Bloom

It has been shown that choosing the radius of the pit hole $R\ \frac{1}{2}\sqrt{\frac{\sqrt{3}}{\pi}}T$ is well-suited for differentiating received signals picked up from parallel read-out in the TwoDOS system [103]. In practice, however, it is difficult to accurately control the size of the recorded domain. For example, domain bloom, which is due to under- or over-etching during the mastering process, causes pits to be systematically larger or smaller and it is one of the major write imperfections. In this case, (2.3) becomes

$$\mathbf{C}_{sf}(\phi,\rho) = 2\pi(1+\Delta_b)R\frac{J_1(\rho(1+\Delta_b)R)}{\rho} \tag{2.10}$$

where Δ_b reflects the degree of bloom. A positive value of Δ_b means that the recorded pits are systematically larger than the nominal pits and vice versa when Δ_b is negative. The channel symbol response $\mathbf{H}_k$ $(k = 0, 1, \cdots, N_h - 1)$ should be computed by taking this into account. Unlike the 1D system, where domain bloom manifests only when pit-to-land or land-to-pit transitions occur, the transitions in TwoDOS occur whenever bits corresponding to '+1' are written. This is because every '+1' is recorded as a pit of radius R smaller than $T/2$ and thus there are two transition edges between consecutive pits, instead of none in the 1D system.

2.3.2 Effect of Transition Jitter

In this section, it has been implicitly assumed so far that the recorded domains have infinitely steep vertical edges. In reality, however, the edges of recorded domains are not always sharply defined and tend to exhibit random displacements from their ideal positions. In read-only systems, this kind of transition jitter is typically the dominant type of noise. Accounting for this jitter in (2.10), we get

$$\mathbf{C}_{sf}(\phi, \rho) = 2\pi(1 + \Delta_b + \Delta_t)R\frac{J_1(\rho(1 + \Delta_b + \Delta_t)R)}{\rho} \tag{2.11}$$

where the normalized jitter Δ_t is a zero-mean random variable that is statistically independent of the channel input vector $\mathbf{a}(n)$. Due to the same reason as explained above, transition jitter occurs whenever the channel input bit is '+1'. The difference between the domain bloom and transition jitter is that the bloom parameter Δ_b can be considered to remain constant for all the bits in a given recording process, whereas the jitter Δ_t varies randomly for each bit.

2.3.3 Discrete-Time Channel Model with Nonlinear Distortions

In the presence of nonlinear distortions, *i.e.* domain bloom and transition jitter, it is easy to see that the channel becomes time-varying since the symbol response tends to differ from one bit to another. As a result, the received signal takes the form (see Fig. 2.2)

$$\mathbf{z}(n) = \sum_{k=0}^{N_h-1} \mathbf{H}_{k,n}\mathbf{a}(n-k) + \boldsymbol{\theta}(n) \tag{2.12}$$

where the channel matrices $\mathbf{H}_{k,n}$ are now also dependent on the time index 'n'. The elements $H_{k,n}(i,j)$ $(i,j = 1, 2, \cdots, N_r)$ of $\mathbf{H}_{k,n}$ can be computed using (2.8), (2.7) and (2.4) after replacing the $\mathbf{C}_{sf}(\phi, \rho)$ in (2.4) with the time-varying form given in (2.10) or (2.11). Thus, to compute $\mathbf{z}(n)$ for each 'n', we need to recompute $\mathbf{H}_{k,n}$ for all k, which is a very computationally demanding task. A simple table look-up can be used to save computational complexity by storing pre-computed channel responses for quantized values of pit sizes. On the other hand, since the time-varying symbol response still satisfies the radial symmetry condition in spatial frequency domain, we can use the fast Hankel transform approach to compute the individual symbol responses. This approach results in significant savings in computational complexity and memory requirements compared with the 2D Fourier transform approach and the table look-up approach, respectively. This is very important for doing simulation studies over very a large number of data bits.

The Fourier transform of the channel symbol response in the presence of domain bloom and transition jitter is given by (using (2.11) and (2.4))

$$\mathbf{H}_{sf,\,t}(\phi,\rho) = 2\pi(1+\Delta_b+\Delta_t)R\frac{J_1(\rho(1+\Delta_b+\Delta_t)R)}{\rho}\mathbf{F}_{sf}(\phi,\rho) \tag{2.13}$$

where Δ_b and Δ_t are the normalized domain bloom and random jitter, respectively, with respect to R for the bit under consideration. If Δ_b and Δ_t are sufficiently small and the recording density (linear) is sufficiently high, the above transfer function can be approximated as [122]

$$\mathbf{H}_{sf,\,t}(\phi,\rho) \approx 2\pi(1+\Delta_b+\Delta_t)R\frac{J_1(\rho R)}{\rho}\mathbf{F}_{sf}(\phi,\rho) = (1+\Delta_b+\Delta_t)\mathbf{H}_{sf}(\phi,\rho) \tag{2.14}$$

where $\mathbf{H}_{sf}(\phi,\rho)$ is the time-invariant symbol response given in (2.4). Using this approximation, we can write (2.12) as

$$\mathbf{z}(n) = \sum_{k=0}^{N_h-1}\mathbf{H}_k\tilde{\mathbf{a}}(n-k) + \boldsymbol{\theta}(n) \tag{2.15}$$

where $\mathbf{H}_k$ $(k = 0, 1, \cdots, N_h - 1)$ corresponds to the time-invariant symbol response matrices, $\tilde{\mathbf{a}}(n) = [\tilde{a}_1(n),\, \tilde{a}_2(n),\, \cdots,\, \tilde{a}_{N_r}(n)]^T$ is the modified channel input vector,

$$\tilde{a}_i(n) = \begin{cases} (1+\Delta_i(n))a_i(n), & \text{if } \ a_i(n) = +1, \\ a_i(n), & \text{if } \ a_i(n) = -1, \end{cases} \tag{2.16}$$

for $i = 1,\, 2,\, \cdots,\, N_r$, and $\Delta_i(n)$ corresponds to the sum of domain bloom and random jitter experienced by the bit $a_i(n)$. Compared with (2.15) and (2.12), it is clear that during the received signal generation process, the time-dependent component, originally coming from the channel matrix as shown in (2.12), is now merged with the input bit vector as shown in (2.15). This is indeed important for doing simulation studies over a very large number of data bits since the channel matrix recomputation, which requires much more computational complexity compared to the simple channel input vector modification, is

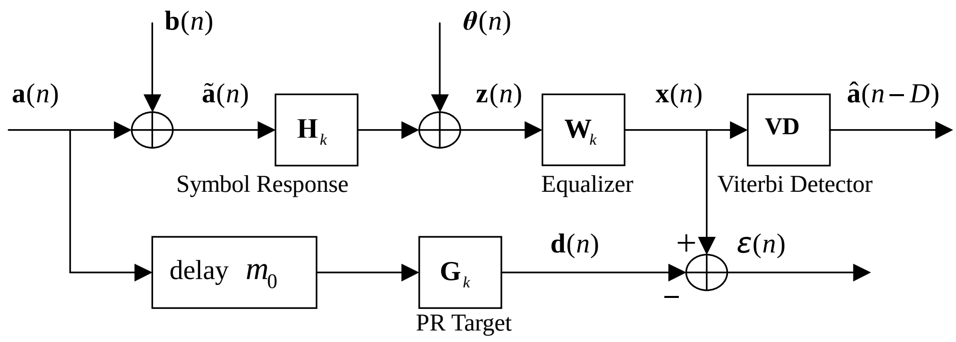

Figure 2.3: Approximated additive discrete-time channel model of TwoDOS for the domain bloom and transition jitter.

avoided during the received signal generation process.

The resulting approximated additive discrete-time channel model of the TwoDOS system for domain bloom and transition jitter is shown in Fig. 2.3. In this figure, the additive disturbance $\mathbf{b}(n)$ that accounts for domain bloom and transition jitter is given by $\mathbf{b}(n) = [b_1(n), b_2(n), \cdots, b_{N_r}(n)]^T$ where

$$b_i(n) = \begin{cases} \Delta_i(n), & \text{if } a_i(n) = +1, \\ 0, & \text{if } a_i(n) = -1, \end{cases}$$

for $i = 1, 2, \cdots, N_r$. Then, the modified channel input vector is given by

$$\tilde{\mathbf{a}}(n) = \mathbf{a}(n) + \mathbf{b}(n). \tag{2.17}$$

2.4 Conclusions

In this chapter, we have first presented a linear channel model that incorporates the linear pulse modulator. More importantly, we have shown that by exploiting the radial symmetry property of the symbol response, the 2D symbol response can be efficiently calculated by using the 1D Hankel transform. We have also indicated that the 1D Hankel transform may still be applicable in cases with modest radial asymmetry. In addition, We have extended this 2D symbol response model to include the domain bloom and transition jitter. Finally, we have introduced a discrete-time additive model to approximate the domain bloom and transition jitter. This additive model further significantly simplifies the received signal generation process in the presence of the domain bloom and transition jitter for the TwoDOS system.

The channel models developed in this chapter are used to design and evaluate different signal processing techniques that ensure reliable data recovery for the received signals, which suffer severe ISI and ITI, from the parallel read-out of the TwoDOS system. These techniques are explored in the remaining chapters of Part I.

Chapter 3

2D Equalization and Target Design

3.1 Introduction

Because of the presence of serious intersymbol interference (ISI) and intertrack interference (ITI) in the received signals of the two-dimensional optical storage (TwoDOS) system, the two-dimensional (2D) Viterbi detector (VD) seems ideal for bit detection. However, as the complexity of the 2D VD grows exponentially with both the channel memory and number of tracks per group, for the TwoDOS channel that has very large channel memory, this complexity becomes unmanageable. Therefore, this chapter describes two prefiltering techniques to shorten the channel memory and thus reduce the complexity of 2D VD. In the first technique, for a given target, we design a general 2D minimum mean square error (MMSE) equalizer, which can be used to deal with correlated data, colored additive noise, domain bloom and jitter noise. In the second technique, we jointly design the equalizer and target based on minimizing the mean square error (MSE) to improve the performance of 2D VD. However, to avoid the trivial solution, we need to impose a suitable constraint on the 2D partial response (PR) target. Since doing constrained minimization in 2D is not easy, we propose a novel technique which converts the 2D target design problem into a one-dimensional (1D) problem. To evaluate different targets, a theoretical platform is constructed to analyze the performance of 2D VD with different targets. Since the complexity of exhaustive search over all possible channel input data patterns to compute the theoretical performance is prohibitively high, we make a reasonable approximation which results in significant savings in computational complexity with an affordable performance loss. We also develop an appropriate target constraint that results in good bit detection performance for the 2D VD.

This chapter is organized as follows. Section 3.2 presents the design of a general 2D MMSE equalizer that can deal with correlated data, colored additive noise, domain bloom and jitter noise. Section 3.3 investigates the theoretical performance of 2D VD in the additive noise, domain bloom and jitter noise. This analysis suggests the criterion to jointly optimize the target and equalizer. Based on this criterion, a novel 2D target design technique that converts the 2D target design problem into a 1D problem is also proposed in this section. The investigation of different target constraints along with performance comparison through bit error rate simulations are presented in Section 3.4. Finally, Section 3.5 concludes the chapter.

3.2 2D MMSE Equalizer Design

3.2.1 Generalized 2D MMSE Equalizer

The 2D MMSE equalizer design for TwoDOS was first proposed in [156]. However, the derivation in [156] assumes the channel input vector $\mathbf{a}(n)$ as uncoded and the channel noise $\boldsymbol{\theta}(n)$ as 2D white. In this chapter, we present a generalized equalizer, which is applicable even when the channel contains domain bloom and jitter noise.

In this section, we assume the channel symbol response $\mathbf{H}_k$ $(k = 0, 1, \cdots, N_h - 1)$ is linear and time-invariant, and the PR target $\mathbf{G}_k$ is known. From Fig. 2.2, we get the error vector $\boldsymbol{\varepsilon}(n)$ as

$$\begin{aligned}\boldsymbol{\varepsilon}(n) &= \mathbf{x}(n) - \mathbf{d}(n) = \sum_{k=0}^{N_w-1} \mathbf{W}_k \mathbf{z}(n-k) - \sum_{k=0}^{N_g-1} \mathbf{G}_k \mathbf{a}(n-k-m_0) \\ &= \underline{\mathbf{W}}^T \underline{\mathbf{z}}(n) - \underline{\mathbf{G}}^T \underline{\mathbf{a}}(n-m_0) \end{aligned} \tag{3.1}$$

where $\underline{\mathbf{G}} = [\mathbf{G}_0, \mathbf{G}_1, \cdots, \mathbf{G}_{N_g-1}]^T$, $\underline{\mathbf{a}}(n) = [\mathbf{a}^T(n), \mathbf{a}^T(n-1), \cdots, \mathbf{a}^T(n-N_g+1)]^T$, $\underline{\mathbf{W}} = [\mathbf{W}_0, \mathbf{W}_1, \cdots, \mathbf{W}_{N_w-1}]^T$ and $\underline{\mathbf{z}}(n) = [\mathbf{z}^T(n), \mathbf{z}^T(n-1), \cdots, \mathbf{z}^T(n-N_w+1)]^T$. Then, the total MSE at the equalizer output can be expressed as

$$J(\underline{\mathbf{W}}) = E[\boldsymbol{\varepsilon}^T(n)\boldsymbol{\varepsilon}(n)] = \text{trace}\left(\underline{\mathbf{W}}^T\underline{\mathbf{R}}_z\underline{\mathbf{W}} - 2\underline{\mathbf{W}}^T\underline{\mathbf{R}}_{za}\underline{\mathbf{G}} + \underline{\mathbf{G}}^T\underline{\mathbf{R}}_{aa}\underline{\mathbf{G}}\right) \tag{3.2}$$

where $\underline{\mathbf{R}}_{aa} = E[\underline{\mathbf{a}}(n-m_0)\underline{\mathbf{a}}^T(n-m_0)]$, $\underline{\mathbf{R}}_{za} = E[\underline{\mathbf{z}}(n)\underline{\mathbf{a}}^T(n-m_0)]$ and $\underline{\mathbf{R}}_z = E[\underline{\mathbf{z}}(n)\underline{\mathbf{z}}^T(n)]$. By minimizing the MSE, the optimum equalizer matrices are obtained as

$$\underline{\mathbf{W}}_{opt} = \underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}_{za}\underline{\mathbf{G}} \tag{3.3}$$

and the MMSE is given by

$$J_{min}(\underline{\mathbf{W}}_{opt}) = \text{trace}\left[\underline{\mathbf{G}}^T\underline{\mathbf{R}}_{aa}\underline{\mathbf{G}} - (\underline{\mathbf{R}}_{za}\underline{\mathbf{G}})^T\underline{\mathbf{R}}_z^{-T}(\underline{\mathbf{R}}_{za}\underline{\mathbf{G}})\right]. \tag{3.4}$$

We now derive the expressions for $\underline{\mathbf{R}}_z$, $\underline{\mathbf{R}}_{za}$ and $\underline{\mathbf{R}}_{aa}$. Here, $\underline{\mathbf{R}}_z$ is an $N_wN_r \times N_wN_r$ autocorrelation matrix of the channel output vector and its (k,l)th sub-matrix $(k,l = 0,1,\cdots,N_w-1)$ is given by $\underline{\mathbf{R}}^z_{kl} = E[\mathbf{z}(n-k)\mathbf{z}^T(n-l)]$. Since $\boldsymbol{\theta}(n)$ is of zero mean and uncorrelated with $\mathbf{a}(n)$, using (2.9) we get

$$\underline{\mathbf{R}}^z_{kl} = \underline{\mathbf{H}}^T\underline{\mathbf{R}}^a_{kl}\underline{\mathbf{H}} + \mathbf{R}^\theta_{kl} \tag{3.5}$$

where $\underline{\mathbf{R}}^a_{kl} = E[\breve{\underline{\mathbf{a}}}(n-k)\breve{\underline{\mathbf{a}}}^T(n-l)]$ is an $N_hN_r \times N_hN_r$ autocorrelation matrix of the channel input data, $\breve{\underline{\mathbf{a}}}(n) = [\mathbf{a}^T(n), \mathbf{a}^T(n-1), \cdots, \mathbf{a}^T(n-N_h+1)]^T$, and $\mathbf{R}^\theta_{kl} = E\left[\boldsymbol{\theta}(n-k)\boldsymbol{\theta}^T(n-l)\right]$ is an $N_r \times N_r$ autocorrelation matrix of the noise. The (k',l')th sub-matrix of $\underline{\mathbf{R}}^a_{kl}$ is given by

$$\underline{\mathbf{R}}^a_{kl,k'l'} = E\left[\mathbf{a}(n-k-k')\mathbf{a}^T(n-l-l')\right], \quad k',l' = 0,1,\cdots,N_h-1 \tag{3.6}$$

and the (i,j)th elements of $\underline{\mathbf{R}}^a_{kl,k'l'}$ and $\mathbf{R}^\theta_{kl}$ are given by

$$\underline{\mathbf{R}}^a_{kl,k'l'}(i,j) = E\left[a_i(n-k-k')a_j(n-l-l')\right], \tag{3.7}$$

$$\mathbf{R}^\theta_{kl}(i,j) = E\left[\theta_i(n-k)\theta_j(n-l)\right], \tag{3.8}$$

respectively, for $i,j = 1,2,\cdots,N_r$.

The matrix $\underline{\mathbf{R}}_{za}$ is an $N_wN_r \times N_gN_r$ cross-correlation matrix between the channel output and input vectors, and its (k,l)th ($k = 0,1,\cdots,N_w-1$; $l = 0,1,\cdots,N_h-1$) sub-matrix is given by $\underline{\mathbf{R}}_{kl}^{za} = E\left[\underline{\mathbf{z}}(n-k)\mathbf{a}^T(n-m_0-l)\right]$. Then, as before, we get

$$\underline{\mathbf{R}}_{kl}^{za} = \underline{\mathbf{H}}^T\tilde{\underline{\mathbf{R}}}_{kl}^{a} \tag{3.9}$$

where $\tilde{\underline{\mathbf{R}}}_{kl}^{a} = E[\underline{\breve{\mathbf{a}}}(n-k)\mathbf{a}^T(n-m_0-l)]$ is an $N_hN_r \times N_r$ autocorrelation matrix of the channel input data. Its (k')th $N_r \times N_r$ sub-matrix is given by

$$\tilde{\underline{\mathbf{R}}}_{kl,k'}^{a} = E\left[\mathbf{a}(n-k-k')\mathbf{a}^T(n-m_0-l)\right], \quad k' = 0,1,\cdots,N_h-1 \tag{3.10}$$

and the (i,j)th element of $\tilde{\underline{\mathbf{R}}}_{kl,k'}^{a}$ is given by

$$\tilde{\underline{\mathbf{R}}}_{kl,k'}^{a}(i,j) = E\left[a_i(n-k-k')a_j(n-m_0-l)\right], \quad i,j = 1,2,\cdots,N_r. \tag{3.11}$$

Finally, $\underline{\mathbf{R}}_{aa}$ is an $N_gN_r \times N_gN_r$ autocorrelation matrix of channel input vectors. As before, its (k,l)th $(k,l = 0,1,\cdots,N_g-1)$ sub-matrix is given by $\underline{\mathbf{R}}_{kl}^{aa} = E[\mathbf{a}(n-m_0-k)\mathbf{a}(n-m_0-l)]$, and the (i,j)th element of $\underline{\mathbf{R}}_{kl}^{aa}$ is given by

$$\underline{\mathbf{R}}_{kl}^{aa}(i,j) = E\left[a_i(n-m_0-k)a_j(n-m_0-l)\right], \quad i,j = 1,2,\cdots,N_r. \tag{3.12}$$

3.2.2 Special Cases

(1) Channel input data and noise are 2D wide-sense stationary.

In this case, (3.7), (3.8), (3.11) and (3.12) can be simplified as

$$\begin{aligned}
\underline{\mathbf{R}}_{kl,k'l'}^{a}(i,j) &= r_a(l-k+l'-k')(i-j), &(3.13)\\
\mathbf{R}_{kl}^{\theta}(i,j) &= r_\theta(l-k)(i-j), &(3.14)\\
\tilde{\underline{\mathbf{R}}}_{kl,k'}^{a}(i,j) &= r_a(l-k-k'+m_0)(i-j), &(3.15)\\
\underline{\mathbf{R}}_{kl}^{aa}(i,j) &= r_a(l-k)(i-j), &(3.16)
\end{aligned}$$

respectively, where $r_a(k)(i)$ and $r_\theta(k)(i)$ are defined as $E[a_{i+j}(n+k)a_j(n)]$ and $E[\theta_{i+j}(n+k)\theta_j(n)]$, respectively, for any n and j.

(2) Noise is 2D white with variance σ^2, and channel input bits are uncoded.

This is the simplest case. In this case, (3.7), (3.8), (3.11) and (3.12) can be simplified as

$$\begin{aligned}
\underline{\mathbf{R}}^a_{kl,k'l'}(i,j) &= \delta_{l-k+l'-k'}\ \delta_{i-j}, && (3.17)\\
\mathbf{R}^\theta_{kl}(i,j) &= \sigma^2\delta_{l-k}\ \delta_{i-j}, && (3.18)\\
\underline{\tilde{\mathbf{R}}}^a_{kl,k'}(i,j) &= \delta_{l-k-k'+m_0}\ \delta_{i-j}, && (3.19)\\
\underline{\mathbf{R}}^{aa}_{kl}(i,j) &= \delta_{l-k}\ \delta_{i-j}, && (3.20)
\end{aligned}$$

respectively, where $\delta_k = 1$ when $k = 0$ and $\delta_k = 0$ when $k \neq 0$. Substituting (3.17) and (3.18) in (3.5), we get

$$\underline{\mathbf{R}}^z_{kl} = \begin{cases} (\mathbf{H} * \mathbf{H}^T_-)_{l-k} + \sigma^2 \mathbf{I}_{N_r} & \text{if } k = l \\ (\mathbf{H} * \mathbf{H}^T_-)_{l-k} & \text{else} \end{cases} \tag{3.21}$$

where $\mathbf{H}_-$ is the mirror filter of $\mathbf{H}$, *i.e.* $(\mathbf{H}_-)_k = \mathbf{H}_{-k}$, $\mathbf{I}_{N_r}$ is the $N_r \times N_r$ identity matrix, and '$*$' refers to 2D convolution given by $(\mathbf{A} * \mathbf{B})_k = \sum_n \mathbf{A}_n \mathbf{B}_{k-n}$. Using (3.9), we get the kth sub-matrix of $\underline{\mathbf{R}}_{za}\underline{\mathbf{G}}$ as

$$(\underline{\mathbf{R}}_{za}\underline{\mathbf{G}})_k = \sum_{k'=0}^{N_h-1} \sum_{l=0}^{N_g-1} \mathbf{H}_{k'} \underline{\tilde{\mathbf{R}}}^a_{kl,k'} \mathbf{G}^T_l. \tag{3.22}$$

Substituting (3.19) in (3.22) and letting $l = k + k' - m_0$, we obtain

$$\underline{\mathbf{R}}_{za}\underline{\mathbf{G}} = \sum_{k'=0}^{N_h-1} \begin{bmatrix} \mathbf{H}_{k'}\mathbf{G}^T_{k'-m_0} \\ \mathbf{H}_{k'}\mathbf{G}^T_{k'-m_0+1} \\ \vdots \\ \mathbf{H}_{k'}\mathbf{G}^T_{k'-m_0+N_w-1} \end{bmatrix} = \begin{bmatrix} (\mathbf{H}_- * \mathbf{G}^T)_{-m_0} \\ (\mathbf{H}_- * \mathbf{G}^T)_{1-m_0} \\ \vdots \\ (\mathbf{H}_- * \mathbf{G}^T)_{N_w-1-m_0} \end{bmatrix}. \tag{3.23}$$

We can compute the optimum equalizer matrices by using (3.21) and (3.23) in (3.3). Since $\underline{\mathbf{R}}_{aa}$ is an $N_gN_r \times N_gN_r$ identity matrix in this simplest case, by using (3.4), we get the corresponding MMSE as

$$J_{min}(\underline{\mathbf{W}}) = \text{trace}\left[\underline{\mathbf{G}}^T\underline{\mathbf{G}} - (\underline{\mathbf{R}}_{za}\underline{\mathbf{G}})^T \underline{\mathbf{R}}_z^{-T} (\underline{\mathbf{R}}_{za}\underline{\mathbf{G}})\right]. \tag{3.24}$$

The expressions (3.21)-(3.24) corresponding to the simplest case can be found in [156].

(3) Channel contains domain bloom and transition jitter.

From (2.17) and Fig. 2.3, it is easy to see that Case (1) discussed above (*i.e.* wide-sense stationary data and noise) can be used to design the optimum equalizer for the case when the channel contains domain bloom and transition jitter since the modified data $\tilde{a}_k(n)$ and white additive noise $\theta_k(n)$ are mutually uncorrelated and wide-sense stationary. In this case, (3.7), (3.8), (3.11) and (3.12) can be simplified as

$$\underline{\mathbf{R}}^a_{kl,k'l'}(i,j) = r_{\tilde{a}}(l-k+l'-k')(i-j), \tag{3.25}$$

$$\mathbf{R}^\theta_{kl}(i,j) = r_\theta(l-k)(i-j), \tag{3.26}$$

$$\underline{\tilde{\mathbf{R}}}^a_{kl,k'}(i,j) = r_{\tilde{a}}(l-k-k'+m_0)(i-j), \tag{3.27}$$

$$\underline{\mathbf{R}}^{aa}_{kl}(i,j) = r_{\tilde{a}}(l-k)(i-j), \tag{3.28}$$

respectively, where $r_{\tilde{a}}(k)(i)$ and $r_\theta(k)(i)$ are defined as $E[\tilde{a}_{i+j}(n+k)\tilde{a}_j(n)]$ and $E[\theta_{i+j}(n+k)\theta_j(n)]$, respectively, for any n and j. Let $\Delta^t_j(n)$ represent the random jitter experienced by the bit $a_j(n)$. Noting that $\Delta^t_j(n)$ is statistically independent of $a_j(n)$ and $\mathbf{a}(n)$ is of zero mean, it can be shown that

$$r_{\tilde{a}}(k)(i) = (1+\Delta_b)\, r_a(k)(i) + \frac{[1+r_a(k)(i)]\left[\Delta_b^2 + r_{\Delta_t}(k)(i)\right]}{4} \tag{3.29}$$

where $r_a(k)(i)$ and $r_{\Delta_t}(k)(i)$ are defined as $E[a_{i+j}(n+k)a_j(n)]$ and $E[\Delta^t_{i+j}(n+k)\Delta^t_j(n)]$, respectively, for any n and j. If $\Delta_b = 0$, the noise is 2D white with variance σ^2 and the channel input bits are uncoded, then the expressions of correlations given in (3.7), (3.8), (3.11) and (3.12) can be simplified for this case as

$$\underline{\mathbf{R}}^a_{kl,k'l'}(i,j) = \left(1+\sigma_t^2/2\right)\delta_{l-k+l'-k'}\;\delta_{i-j}, \tag{3.30}$$

$$\mathbf{R}^\theta_{kl}(i,j) = \sigma^2\delta_{l-k}\;\delta_{i-j}, \tag{3.31}$$

$$\underline{\tilde{\mathbf{R}}}^a_{kl,k'}(i,j) = \left(1+\sigma_t^2/2\right)\delta_{l-k-k'+m_0}\;\delta_{i-j}, \tag{3.32}$$

$$\underline{\mathbf{R}}^{aa}_{kl}(i,j) = \left(1+\sigma_t^2/2\right)\delta_{l-k}\;\delta_{i-j}, \tag{3.33}$$

respectively, where σ_t^2 is the variance of Δ_t. The factor 1/2 accounts for the probability of pit bit '+1' where the transition jitter occurs.

3.3 Target Design for TwoDOS

3.3.1 Theoretical Platform for Target Evaluation

To deal with serious ISI and ITI present in the TwoDOS received signals, powerful detectors such as the 2D VD become necessary to ensure reliable data recovery [103]. The bit error rate (BER) performance of 2D VD may be used as a criterion to evaluate different PR targets. Even though a BER expression for the 2D VD has been reported in [156], the variables used in the expression correspond to the 1D VD. In this section, we briefly give the derivation of BER of the 2D VD using the 2D definitions and notations of error event, target *etc.*

As a starting point for determining the BER, we define a 2D error event vector $\underline{\mathbf{e}}$. Let $\hat{\mathbf{a}}(n)$ denote the detected data vector corresponding to $\mathbf{a}(n)$. Further, define $\mathbf{e}(n) = (\mathbf{a}(n) - \hat{\mathbf{a}}(n))/2$, $\underline{\mathbf{a}} = [\mathbf{a}^T(n), \mathbf{a}^T(n+1), \cdots, \mathbf{a}^T(n+N_e-1)]^T$, $\underline{\hat{\mathbf{a}}} = [\hat{\mathbf{a}}^T(n), \hat{\mathbf{a}}^T(n+1), \cdots, \hat{\mathbf{a}}^T(n+N_e-1)]^T$ and $\underline{\mathbf{e}} = [\mathbf{e}^T(n), \mathbf{e}^T(n+1), \cdots, \mathbf{e}^T(n+N_e-1)]^T$. Then, $\underline{\mathbf{e}}$ is a 2D error event of length N_e if

1. $\mathbf{e}(n) \neq 0$ and $\mathbf{e}(n+N_e-1) \neq 0$,
2. the length of strings of zero vectors in $\underline{\mathbf{e}}$ does not exceed $N_g - 1$, and
3. $\hat{\mathbf{a}}(n+i) = \mathbf{a}(n+i)$ for $-N_f \leq i < 0$ and $N_e \leq i \leq N_e + N_f - 1$.

Here N_f is called the "error-free interval" and $N_f \geq N_g - 1$, where N_g is the target length. Let $W(\underline{\mathbf{e}})$ be the number of error bits in the error event $\underline{\mathbf{e}}$. Then, the BER of 2D VD is given by

$$\text{BER} = \sum_{\underline{\mathbf{e}} \in E_s} \sum_{\underline{\mathbf{a}} \in S_e} W(\underline{\mathbf{e}}) \Pr(\underline{\mathbf{a}}) \Pr(error|\underline{\mathbf{a}}) \tag{3.34}$$

where E_s is the set of all possible error events, S_e is the set of all possible data patterns that support the error event $\underline{\mathbf{e}}$, and $\Pr(error|\underline{\mathbf{a}})$ is the conditional probability that the 2D VD detects $\underline{\hat{\mathbf{a}}}$ as the recorded data when the true recorded data is $\underline{\mathbf{a}}$.

According to the principle of 2D VD, $\Pr(error|\underline{\mathbf{a}})$ can be upper-bounded by the probability that the metric corresponding to path $\underline{\hat{\mathbf{a}}}$ is smaller than that corresponding to path $\underline{\mathbf{a}}$. That is

$$\sum_{i=0}^{N_e+N_g-2} \left[\mathbf{x}(n+i) - (\mathbf{G} * \hat{\mathbf{a}})_{n+i}\right]^T \left[\mathbf{x}(n+i) - (\mathbf{G} * \hat{\mathbf{a}})_{n+i}\right]$$
$$< \sum_{i=0}^{N_e+N_g-2} \left[\mathbf{x}(n+i) - (\mathbf{G} * \mathbf{a})_{n+i}\right]^T \left[\mathbf{x}(n+i) - (\mathbf{G} * \mathbf{a})_{n+i}\right] \tag{3.35}$$

where $\mathbf{x}(n)$ is the equalizer output which can be expressed as (see Fig. 2.2)

$$\mathbf{x}(n) = \sum_{i=0}^{N_w-1} \mathbf{W}_i \mathbf{z}(n-i) = \left[(\mathbf{G}+\mathbf{M}) * \mathbf{a}\right]_n + (\mathbf{W} * \boldsymbol{\theta})_n \tag{3.36}$$

where $\mathbf{M}_0$, $\mathbf{M}_1$, $\cdots$, $\mathbf{M}_{N_h+N_w-2}$ represent the matrices of the 2D residual ISI channel, *i.e.* $\mathbf{M}_n = (\mathbf{W} * \mathbf{H})_n - \mathbf{G}_n$, and $\boldsymbol{\theta}(n)$ is a 2D additive white Gaussian noise vector with variance σ^2. Using (3.36), we can express (3.35) as

$$d^2(\underline{\mathbf{e}}) + \sum_{i=0}^{N_e+N_g-2} \tilde{\mathbf{e}}^T(n+i)[(\mathbf{M} * \mathbf{a})_{n+i} + (\mathbf{W} * \boldsymbol{\theta})_{n+i}] < 0 \tag{3.37}$$

where $d^2(\underline{\mathbf{e}}) = \sum_{i=0}^{N_e+N_g-2}[(\mathbf{G} * \mathbf{e})_{n+i}^T(\mathbf{G} * \mathbf{e})_{n+i}]$ and $\tilde{\mathbf{e}}(n) = (\mathbf{G} * \mathbf{e})_n$. For given $\underline{\mathbf{a}}$ and $\underline{\mathbf{e}}$, the left hand side of the inequality in (3.37) is a Gaussian random variable with mean

$$m_u = d^2(\underline{\mathbf{e}}) + \sum_{i=-(N_e+N_g-2)}^{N_e+N_g-2} \left(\mathbf{M}_-^T * \tilde{\mathbf{e}}\right)_{n+i}^T \mathbf{a}(n+i) \tag{3.38}$$

and variance

$$\sigma_u^2 = \sigma^2 \sum_{i=-(N_w-1)}^{N_e+N_g-2} \left(\mathbf{W}_-^T * \tilde{\mathbf{e}}\right)_{n+i}^T \left(\mathbf{W}_-^T * \tilde{\mathbf{e}}\right)_{n+i}. \tag{3.39}$$

Thus, we get

$$\Pr(error|\underline{\mathbf{a}}) \leq Q\left(\frac{m_u}{\sigma_u}\right) \tag{3.40}$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{-x^2/2}dx$ is the tail probability of Gaussian distribution. By substituting (3.40) into (3.34), we get the BER of 2D VD as

$$\text{BER} \leq \sum_{\underline{\mathbf{e}} \in E_s} \sum_{\underline{\mathbf{a}} \in S_e} W(\underline{\mathbf{e}}) \Pr(\underline{\mathbf{a}}) Q\left(\frac{m_u}{\sigma_u}\right). \tag{3.41}$$

Based on the BER expression given above, we may define an effective signal-to-noise ratio (SNR_{eff}) for a given error event $\underline{\mathbf{e}}$ and target $\underline{\mathbf{G}}$ as

$$\text{SNR}_{\text{eff}} = \left(\frac{d^2(\underline{\mathbf{e}}) + \sum_n \left(\mathbf{M}_-^T * \tilde{\mathbf{e}}\right)_n^T \mathbf{a}(n)}{\sigma \sqrt{\sum_n \left(\mathbf{W}_-^T * \tilde{\mathbf{e}}\right)_n^T \left(\mathbf{W}_-^T * \tilde{\mathbf{e}}\right)_n}} \right)^2 . \tag{3.42}$$

The error event $\underline{\mathbf{e}}_d$ that minimizes SNR_{eff} for a given target $\underline{\mathbf{G}}$ is called the dominant error event since this error event has the highest probability of occurrence. Consequently, we could use SNR_{eff} as a criterion to determine the target since it is closely related to BER and can be computed very easily compared to BER. The optimum target should maximize the SNR_{eff} corresponding to the dominant error event. In other words, the design of optimum target is a max-min optimization problem:

$$\underline{\mathbf{G}}_{opt} = \arg \left\{ \max_{\underline{\mathbf{G}}} \ \min_{\underline{\mathbf{e}}} \ \text{SNR}_{\text{eff}} \right\} . \tag{3.43}$$

As seen from (3.42), the effect of noise correlation on the detection performance can be inferred by examining the denominator of SNR_{eff}. Further, the presence of 2D residual ISI makes this optimization data-dependent in nature, and the complexity of exhaustive search of all possible channel input data patterns to compute the theoretical performance is prohibitively large. Therefore, the usual practice is to assume that the 2D residual ISI is very small and to add it as part of the Gaussian noise $(\mathbf{W} * \boldsymbol{\theta})_n$ [137]. In this section, however, we will introduce a practical approach with significantly lower complexity to analyze the systems that suffer from severe 2D residual ISI. As shown in (3.42), each element of $\mathbf{a}(n)$ corresponds to one unique element in $(\mathbf{M}_-^T * \tilde{\mathbf{e}})_n^T$. Consequently, we could only consider the bit positions whose corresponding elements in $(\mathbf{M}_-^T * \tilde{\mathbf{e}})_n^T$ are beyond a prescribed value. The remaining bits are ignored since they have much less effect on the mean m_u (and thus the BER of the 2D VD). Then, we exhaustively search over all the possible data patterns of the bit positions under consideration to approximately compute the BER. Intuitively, the choice of the threshold is a trade-off between the complexity and performance. Our simulation research results show that this technique is able to evaluate different targets with acceptable computational complexity with a suitable threshold as illustrated in Fig. 3.1. For the purpose of assessing the suitability of this approximated theoretical platform, let us assume that the three targets shown in Fig. 3.1 are already known. Details of designing these targets will be explained in the following two subsections. As illustrated, the approximated theoretical BER performance is asymptotic to the simulation BER performance, thus indicating that the approximation described above is suitable to evaluate different targets for the TwoDOS system.

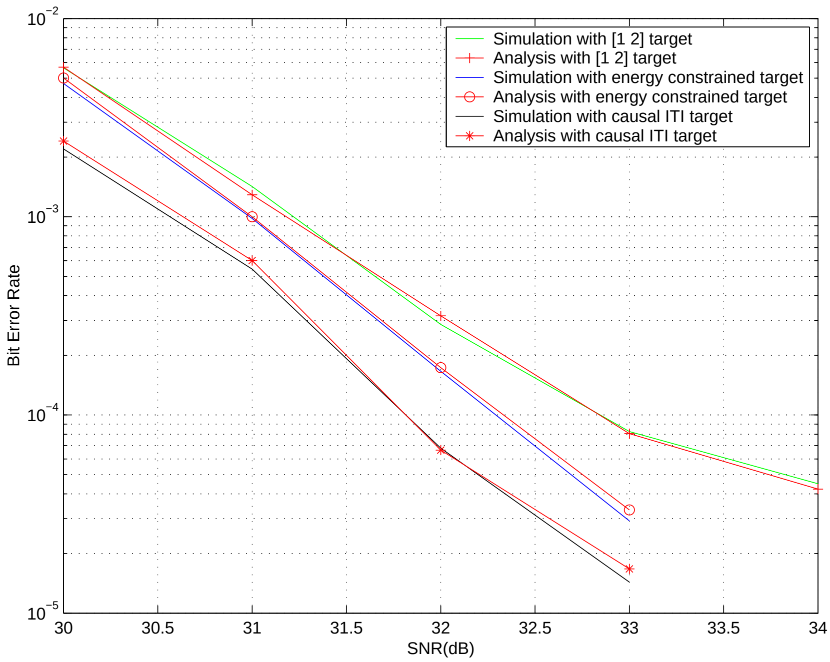

Figure 3.1: Comparison between approximated theoretical and simulation BER performance for different targets.

For the case when the channel contains domain bloom and transition jitter, (3.37) is revised as

$$\begin{aligned}&\sum_{i=0}^{N_e+N_g-2} \tilde{\mathbf{e}}^T(n+i)\left[(\mathbf{M}*\mathbf{a})_{n+i}+(\mathbf{W}*\boldsymbol{\theta})_{n+i}\right]\\ &< -d^2(\underline{\mathbf{e}})-\sum_{i=0}^{N_e+N_g-2}\tilde{\mathbf{e}}^T(n+i)[\tilde{\mathbf{G}}*(\tilde{\mathbf{a}}-\mathbf{a})]_{n+i}\end{aligned} \tag{3.44}$$

where $\tilde{\mathbf{a}}(n)$ is the modified channel input vector defined in (2.16) and $\tilde{\mathbf{G}}_n = \mathbf{G}_n + \mathbf{M}_n$. Under the assumption that the transition jitter and noise $\boldsymbol{\theta}_n$ are uncorrelated and jointly Gaussian distributed, the left hand side of the inequality in (3.44) is still Gaussian with mean

$$m_{ur} = d^2(\underline{\mathbf{e}}) + \sum_i \left(\mathbf{M}_-^T * \tilde{\mathbf{e}}\right)_{n+i}^T \mathbf{a}(n+i) + \sum_i \Delta_b \left(\tilde{\mathbf{G}}_-^T * \tilde{\mathbf{e}}\right)_{n+i}^T \mathbf{u}(n+i) \quad (3.45)$$

and variance

$$\sigma_{ur}^2 = \sigma^2 \sum_i \left(\mathbf{W}_-^T * \tilde{\mathbf{e}}\right)_{n+i}^T \left(\mathbf{W}_-^T * \tilde{\mathbf{e}}\right)_{n+i} + \sigma_t^2 \sum_i \left[\tilde{\mathbf{p}}(n+i)\right]^T \left[\tilde{\mathbf{p}}(n+i))\right] \quad (3.46)$$

where $\tilde{\mathbf{p}}(n) = \left(\tilde{\mathbf{G}}_-^T * \tilde{\mathbf{e}}\right)_n \bullet \mathbf{u}(n)$, Δ_b reflects the degree of domain bloom, σ_t^2 is the variance of the transition jitter, $\mathbf{u}(n) = (1 + \mathbf{a}(n))/2$, and '$\bullet$' indicates element-wise multiplication of vectors. Then, (3.42) becomes

$$\text{SNR}_{\text{eff}} = \left(\frac{m_{ur}}{\sigma_{ur}}\right)^2. \quad (3.47)$$

From (3.47), it can be seen that the presence of transition jitter will degrade the performance since it increases the denominator of (3.47). However, its destructive effect may be reduced if we consider it in the target design that will be covered in the following two subsections.

3.3.2 Novel Target Design Technique

Several techniques have been reported in literature for designing the PR target for 1D data storage systems [64, 118, 137]. For the sake of convenience, the criterion used for the design is to minimize the MSE (at equalizer output) rather than to maximize the SNR_{eff} given in (3.42). The design of equalizer and target by minimizing MSE subject to the monic constraint (*i.e.* the first tap of the target is set to unity) has been reported to result in near-optimum performance in 1D data storage systems [137]. Usually, a constraint is needed for each target to avoid trivial all-zero solutions for the equalizer and target. Further, the choice of target constraints is quite important since the constraint will greatly affect the performance of the 2D VD. Therefore, in this subsection, in addition to developing the 2D target design technique, we also investigate the common constraints that are applicable to different 2D targets. Then, in the next subsection, we will study several specific target constraints. Since the PR target is in the form of a sequence of matrices in the 2D case (as compared to a target vector in 1D case), solving the constrained minimization is quite difficult. Therefore, we introduce a novel technique which transforms the 2D target design problem from 2D into 1D form. Using this technique, targets with different constraints can be designed with relatively less effort as in the 1D case.

Without any loss of generality, we define a vector $\mathbf{g}$ which transforms the target $\underline{\mathbf{G}} = [\mathbf{G}_0,\ \mathbf{G}_1,\ \cdots,\ \mathbf{G}_{N_g-1}]^T$ into a 1D vector target $\mathbf{g} = [g_0,\ g_1,\ \cdots,\ g_{N_v-1}]^T$ where g_i $(i = 0, 1, \cdots, N_v - 1)$ are the distinct non-zero elements in $\underline{\mathbf{G}}$. It should be noted that imposing some symmetry constraints on the coefficients of targets can decrease the number of parameters to be optimized and thus reduce the complexity of target design process. Replacing $\underline{\mathbf{G}}$ with $\mathbf{g}$ in (3.2), the MSE for the 2D case becomes

$$J(\underline{\mathbf{W}}) \quad = \quad \text{trace}\left[\underline{\mathbf{W}}^T \underline{\mathbf{R}}_z \underline{\mathbf{W}} - 2(\underline{\mathbf{W}}^T)^{'}(\underline{\mathbf{R}}_{za})^{'}\mathbf{g} + \mathbf{g}^T(\underline{\mathbf{R}}_{aa})^{'}\mathbf{g}\right] \tag{3.48}$$

where $(\underline{\mathbf{R}}_{za})^{'}$, $(\underline{\mathbf{W}}^T)^{'}$ and $(\underline{\mathbf{R}}_{aa})^{'}$ represent rearranged forms of $\underline{\mathbf{R}}_{za}$, $\underline{\mathbf{W}}^T$ and $\underline{\mathbf{R}}_{aa}$, respectively. More specifically, $(\underline{\mathbf{W}}^T)^{'}$ is a vector given by

$$(\underline{\mathbf{W}}^T)^{'} = \begin{bmatrix} 1^{st} \text{ row of } \underline{\mathbf{W}}^T & \cdots & (N_r)^{th} \text{ row of } \underline{\mathbf{W}}^T \end{bmatrix}. \tag{3.49}$$

Then, we can obtain $(\underline{\mathbf{R}}_{za})^{'}$ as

$$(\underline{\mathbf{R}}_{za})^{'} = \begin{bmatrix} \begin{bmatrix} 1^{st} \text{ column of } \underline{\mathbf{R}}^{'}_{za,g_0} \\ \vdots \\ {N_r}^{th} \text{ column of } \underline{\mathbf{R}}^{'}_{za,g_0} \end{bmatrix} & \cdots & \begin{bmatrix} 1^{st} \text{ column of } \underline{\mathbf{R}}^{'}_{za,g_{N_v-1}} \\ \vdots \\ {N_r}^{th} \text{ column of } \underline{\mathbf{R}}^{'}_{za,g_{N_v-1}} \end{bmatrix} \end{bmatrix} \tag{3.50}$$

where $\text{trace}(\underline{\mathbf{W}}^T\underline{\mathbf{R}}_{za}\underline{\mathbf{G}}) = (\underline{\mathbf{W}}^T)^{'}(\underline{\mathbf{R}}_{za})^{'}\mathbf{g}$ and $\underline{\mathbf{R}}^{'}_{za,g_k}$ is the matrix associated with the coefficient g_k such that

$$\underline{\mathbf{R}}_{za}\underline{\mathbf{G}} = \underline{\mathbf{R}}^{'}_{za,g_0}\, g_0 + \underline{\mathbf{R}}^{'}_{za,g_1}\, g_1 + \cdots + \underline{\mathbf{R}}^{'}_{za,g_{N_v-1}}\, g_{N_v-1}. \tag{3.51}$$

Similarly, $(\underline{\mathbf{R}}_{aa})^{'}$ is an $N_v \times N_v$ matrix that can be obtained such that $\text{trace}(\underline{\mathbf{G}}^T\underline{\mathbf{R}}_{aa}\underline{\mathbf{G}}) = \mathbf{g}^T(\underline{\mathbf{R}}_{aa})^{'}\mathbf{g}$.

To obtain the solutions subject to different constraints, we formulate a cost function using the Lagrange multiplier method [24]:

$$J(\underline{\mathbf{W}}, \mathbf{g}) = \text{trace}\left[\underline{\mathbf{W}}^T \underline{\mathbf{R}}_z \underline{\mathbf{W}} - 2(\underline{\mathbf{W}}^T)^{'}(\underline{\mathbf{R}}_{za})^{'}\mathbf{g} + \mathbf{g}^T(\underline{\mathbf{R}}_{aa})^{'}\mathbf{g}\right] - \lambda c(\mathbf{g}) \tag{3.52}$$

where $c(\mathbf{g})$ represents the constraint function and λ is the Lagrange multiplier. Taking the gradient of $J(\underline{\mathbf{W}}, \mathbf{g})$ with respect to $\mathbf{g}$, we get

$$\nabla_{\mathbf{g}} J(\underline{\mathbf{W}}, \mathbf{g}) = -2[(\underline{\mathbf{R}}_{za})^{'}]^{T}[(\underline{\mathbf{W}}^{T})^{'}]^{T} + 2(\underline{\mathbf{R}}_{aa})^{'}\mathbf{g} - \lambda \nabla_{\mathbf{g}} c(\mathbf{g}). \tag{3.53}$$

Comparing (3.48) and (3.52), it is clear that the expressions for the optimum equalizer should be the same, with or without the constraint. Therefore, from (3.3) and (3.51), the optimum equalizer matrices can be expressed as

$$\underline{\mathbf{W}}_{opt} = \underline{\mathbf{R}}_{z}^{-1}\left(\underline{\mathbf{R}}^{'}_{za,g_0}\, g_0 + \underline{\mathbf{R}}^{'}_{za,g_1}\, g_1 + \cdots + \underline{\mathbf{R}}^{'}_{za,g_{N_v-1}}\, g_{N_v-1}\right). \tag{3.54}$$

Then, the rearranged optimum equalizer matrices can be written as

$$[(\underline{\mathbf{W}}^{T}_{opt})^{'}]^{T} = (\underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za})^{'}\mathbf{g} \tag{3.55}$$

where $(\underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za})^{'}$ can be obtained as

$$(\underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za})^{'} = \left[\begin{bmatrix} 1^{st} \text{ column of } \underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za,g_0} \\ \vdots \\ {N_r}^{th} \text{ column of } \underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za,g_0} \end{bmatrix} \cdots \begin{bmatrix} 1^{st} \text{ column of } \underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za,g_{N_v-1}} \\ \vdots \\ {N_r}^{th} \text{ column of } \underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za,g_{N_v-1}} \end{bmatrix}\right] \tag{3.56}$$

Substituting (3.55) in (3.53), we get

$$\nabla_{\mathbf{g}} J(\underline{\mathbf{W}}, \mathbf{g}) = -2[(\underline{\mathbf{R}}_{za})^{'}]^{T}(\underline{\mathbf{R}}_{z}^{-1}\underline{\mathbf{R}}^{'}_{za})^{'}\mathbf{g} + 2(\underline{\mathbf{R}}_{aa})^{'}\mathbf{g} - \lambda \nabla_{\mathbf{g}} c(\mathbf{g}). \tag{3.57}$$

Thus, the crux of this novel technique is the appropriate rearrangement of matrices. An example is given below to illustrate the technique for $N_g = 3$ and $N_r = 3$. In view of the hexagonal bit-cells and the radially symmetric symbol response, we may choose a target that is also symmetric, *i.e.* the cross interferences to the nearest spots are equal for all the tracks per group. Further, we assume the interferences from beyond adjacent bits to be zero. Then, the target can be written as

$$\mathbf{G}_0 = \begin{bmatrix} g_0 & 0 & 0 \\ g_0 & g_0 & 0 \\ 0 & g_0 & g_0 \end{bmatrix}, \mathbf{G}_1 = \begin{bmatrix} g_1 & g_0 & 0 \\ g_0 & g_1 & g_0 \\ 0 & g_0 & g_1 \end{bmatrix}, \mathbf{G}_2 = \begin{bmatrix} g_0 & g_0 & 0 \\ 0 & g_0 & g_0 \\ 0 & 0 & g_0 \end{bmatrix}, \tag{3.58}$$

and the 1D form of the target matrix becomes $\mathbf{g} = [g_0 \quad g_1]^T$, where g_0 is value of the cross interference to the nearest spot, and g_1 is value of the self interference of the central spot. Let $1^{st}\underline{\mathbf{R}}_{za}$ represent the first column of $\underline{\mathbf{R}}_{za}$ and similar representations are applied to $2^{nd}\underline{\mathbf{R}}_{za}, 3^{rd}\underline{\mathbf{R}}_{za}, \cdots$. Then,

$$\underline{\mathbf{R}}'_{za,g_0} = \begin{bmatrix} \left[1^{st}\underline{\mathbf{R}}_{za} + 5^{th}\underline{\mathbf{R}}_{za} + 7^{th}\underline{\mathbf{R}}_{za} + 8^{th}\underline{\mathbf{R}}_{za}\right]^T \\ \left[1^{st}\underline{\mathbf{R}}_{za} + 2^{nd}\underline{\mathbf{R}}_{za} + 4^{th}\underline{\mathbf{R}}_{za} + 6^{th}\underline{\mathbf{R}}_{za} + 8^{th}\underline{\mathbf{R}}_{za} + 9^{th}\underline{\mathbf{R}}_{za}\right]^T \\ \left[2^{nd}\underline{\mathbf{R}}_{za} + 3^{rd}\underline{\mathbf{R}}_{za} + 5^{th}\underline{\mathbf{R}}_{za} + 9^{th}\underline{\mathbf{R}}_{za}\right]^T \end{bmatrix}^T \quad (3.59)$$

$$\underline{\mathbf{R}}'_{za,g_1} = \begin{bmatrix} \left[4^{th}\underline{\mathbf{R}}_{za}\right]^T \\ \left[5^{th}\underline{\mathbf{R}}_{za}\right]^T \\ \left[6^{th}\underline{\mathbf{R}}_{za}\right]^T \end{bmatrix}^T . \quad (3.60)$$

The rearranged equalizer matrices in this case are given as

$$(\underline{\mathbf{W}}^T)' = \begin{bmatrix} 1st \text{ row of } \underline{\mathbf{W}}^T & 2nd \text{ row of } \underline{\mathbf{W}}^T & 3rd \text{ row of } \underline{\mathbf{W}}^T \end{bmatrix}, \quad (3.61)$$

$$(\underline{\mathbf{R}}_{za}\underline{\mathbf{G}})' = \begin{bmatrix} 1st \text{ column of } \underline{\mathbf{R}}_{za}\underline{\mathbf{G}} \\ 2nd \text{ column of } \underline{\mathbf{R}}_{za}\underline{\mathbf{G}} \\ 3rd \text{ column of } \underline{\mathbf{R}}_{za}\underline{\mathbf{G}} \end{bmatrix}. \quad (3.62)$$

Then, we obtain $(\underline{\mathbf{R}}_{za})'$ as

$$(\underline{\mathbf{R}}_{za})' = \begin{bmatrix} \begin{bmatrix} 1st \text{ column of } \underline{\mathbf{R}}'_{za,g_0} \\ 2nd \text{ column of } \underline{\mathbf{R}}'_{za,g_0} \\ 3rd \text{ column of } \underline{\mathbf{R}}'_{za,g_0} \end{bmatrix} & \begin{bmatrix} 1st \text{ column of } \underline{\mathbf{R}}'_{za,g_1} \\ 2nd \text{ column of } \underline{\mathbf{R}}'_{za,g_1} \\ 3rd \text{ column of } \underline{\mathbf{R}}'_{za,g_1} \end{bmatrix} \end{bmatrix}. \quad (3.63)$$

When the channel input bits are uncorrelated, $\text{trace}(\underline{\mathbf{G}}^T\underline{\mathbf{R}}_{aa}\underline{\mathbf{G}}) = 14g_0^2 + 3g_1^2$. Therefore, we get $(\underline{\mathbf{R}}_{aa})' = \begin{bmatrix} 14 & 0 \\ 0 & 3 \end{bmatrix}$. Finally,

$$(\underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za})' = \begin{bmatrix} \begin{bmatrix} 1st \text{ column of } \underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za,g_0} \\ 2nd \text{ column of } \underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za,g_0} \\ 3rd \text{ column of } \underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za,g_0} \end{bmatrix} & \begin{bmatrix} 1st \text{ column of } \underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za,g_1} \\ 2nd \text{ column of } \underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za,g_1} \\ 3rd \text{ column of } \underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za,g_1} \end{bmatrix} \end{bmatrix} \quad (3.64)$$

3.3.3 2D Target Constraints

In view of the hexagonal bit-cells and the radially symmetric symbol response, as mentioned in the previous subsection, we may choose a target that is also symmetric, *i.e.* we assume that the cross interferences to the nearest spot are equal for all the tracks per group. Further, we assume the interferences from beyond adjacent bits to be zero. Then, the target can be written as ($N_g = 3, N_r = 5$)

$$\begin{aligned}
\mathbf{G}_0 &= \begin{bmatrix} g_0 & 0 & 0 & 0 & 0 \\ g_0 & g_0 & 0 & 0 & 0 \\ 0 & g_0 & g_0 & 0 & 0 \\ 0 & 0 & g_0 & g_0 & 0 \\ 0 & 0 & 0 & g_0 & g_0 \end{bmatrix}, \\
\mathbf{G}_1 &= \begin{bmatrix} g_1 & g_0 & 0 & 0 & 0 \\ g_0 & g_1 & g_0 & 0 & 0 \\ 0 & g_0 & g_1 & g_0 & 0 \\ 0 & 0 & g_0 & g_1 & g_0 \\ 0 & 0 & 0 & g_0 & g_1 \end{bmatrix}, \\
\mathbf{G}_2 &= \begin{bmatrix} g_0 & g_0 & 0 & 0 & 0 \\ 0 & g_0 & g_0 & 0 & 0 \\ 0 & 0 & g_0 & g_0 & 0 \\ 0 & 0 & 0 & g_0 & g_0 \\ 0 & 0 & 0 & 0 & g_0 \end{bmatrix},
\end{aligned} \tag{3.65}$$

and the 1D form of the target matrix becomes $\mathbf{g} = [g_0 \;\; g_1]^T$, where g_0 is the value of the cross interference to the nearest spot, and g_1 is the value of the self interference of the central spot. Even though we are considering symmetric targets here for illustration, the technique that converts the 2D target design problem into a 1D problem is applicable for any general target. For example, the technique can also be used to design the non-symmetric "casual ITI" constrained target to be discussed in the next chapter.

In this subsection, we design the targets for three different constraints. The first constraint used is $g_0 = 1$, which we call the "1D monic constraint". This constraint is widely used in 1D systems and is known to have good noise-whitening properties. Secondly, we use the "energy constraint". The rationale behind this constraint is as follows. For a single-bit error event $\mathbf{I}_e = [1\ 0\ \cdots\ 0]^T$ of length $N_e = 1$, neglecting noise coloration and channel misequalization, the resulting minimum $\mathrm{SNR}_{\mathrm{eff}}$ is given by

$$\mathrm{SNR}_{\mathrm{eff}} = \frac{\sum_k \left(\mathbf{G}_k \mathbf{I}_e\right)^T \left(\mathbf{G}_k \mathbf{I}_e\right)}{\sigma^2_{min}} = \frac{4g_0^2 + g_1^2}{\sigma^2_{min}} \tag{3.66}$$

where σ^2_{min} is the MMSE at the equalizer output. The "energy constraint" mentioned above refers to setting $4g_0^2 + g_1^2 = 1$ so that minimizing the MSE results in maximizing the $\mathrm{SNR}_{\mathrm{eff}}$. A similar argument for choosing a similar energy constraint in 1D data storage systems was made in [118]. The third constraint used is $g_1 = 1$, which we call the "2D monic constraint". The rationale behind this target constraint is as follows. Let the target response for a single spot in the spatial frequency domain be radially symmetric, *i.e.* $\mathbf{G}_{sf}(\phi, \rho) = \tilde{\mathbf{G}}_{sf}(\rho)$. Then, for the noise to be 2D white, $|\tilde{\mathbf{G}}_{sf}(\rho)|^2$ should be ideally of the form $\mu|\tilde{\mathbf{H}}_{sf}(\rho)|^2$, where μ is some constant of proportionality. Of all the targets with the same magnitude characteristics of $|\tilde{\mathbf{G}}_{sf}(\rho)|$, the one with the energy optimally concentrated near g_1 has the largest possible amplitude g_1. Equivalently, for a constrained value $g_1 = 1$, the target with the energy optimally concentrated near g_1 will yield the smallest possible μ and therefore the smallest possible noise enhancement. Since the target is symmetric as given in (3.65), the target has its energy almost concentrated near g_1. Consequently, making the target with the energy optimally concentrated near g_1 will not burden the finite length equalizer much. So setting $g_1 = 1$ is a reasonable target constraint.

For the 1D monic constraint, the constraint function is $c(\mathbf{g}) = 2(\underline{\mathbf{i}}^T\mathbf{g} - 1)$, where $\underline{\mathbf{i}} = [1\ \ 0]^T$. Then, the optimum target can be obtained as (using (3.57))

$$\mathbf{g} = \lambda \left[(\underline{\mathbf{R}}_{aa})' - [(\underline{\mathbf{R}}_{za})']^T (\underline{\mathbf{R}}_z^{-1} \underline{\mathbf{R}}'_{za})'\right]^{-1} \underline{\mathbf{i}} \tag{3.67}$$

where

$$\lambda = \frac{1}{\underline{\mathbf{i}}^T \left[(\underline{\mathbf{R}}_{aa})' - [(\underline{\mathbf{R}}_{za})']^T (\underline{\mathbf{R}}_z^{-1} \underline{\mathbf{R}}'_{za})'\right]^{-1} \underline{\mathbf{i}}}. \tag{3.68}$$

And it can be shown that the MMSE is equal to the Lagrange multiplier λ.

For the energy constraint, we have $c(\mathbf{g}) = \mathbf{g}^T(\underline{\mathbf{R}}_{ec})'\mathbf{g} - 1$, where $(\underline{\mathbf{R}}_{ec})'$ is such that $\sum_k (\mathbf{G}_k\mathbf{I}_e)^T (\mathbf{G}_k\mathbf{I}_e) = \mathbf{g}^T(\underline{\mathbf{R}}_{ec})'\mathbf{g}$. Using this in (3.57) and setting the gradients zero, we obtain

$$\left\{[(\underline{\mathbf{R}}_{ec})']^{-1/2}(\underline{\mathbf{R}}_{aza})'[(\underline{\mathbf{R}}_{ec})']^{-1/2}\right\}[(\underline{\mathbf{R}}_{ec})']^{1/2}\mathbf{g} = \lambda[(\underline{\mathbf{R}}_{ec})']^{1/2}\mathbf{g} \tag{3.69}$$

where $[(\underline{\mathbf{R}}_{ec})']^{1/2}$ is such that $\{[(\underline{\mathbf{R}}_{ec})']^{1/2}\}^T[(\underline{\mathbf{R}}_{ec})']^{1/2} = (\underline{\mathbf{R}}_{ec})'$ and $(\underline{\mathbf{R}}_{aza})' = [(\underline{\mathbf{R}}_{aa})' - [(\underline{\mathbf{R}}_{za})']^T(\underline{\mathbf{R}}_z^{-1}\underline{\mathbf{R}}'_{za})']$. It can be shown that the MMSE is given by the minimum eigenvalue of the matrix $\{[(\underline{\mathbf{R}}_{ec})']^{-1/2}(\underline{\mathbf{R}}_{aza})'[(\underline{\mathbf{R}}_{ec})']^{-1/2}\}$ and $[(\underline{\mathbf{R}}_{ec})']^{1/2}\mathbf{g}$ is the corresponding eigenvector.

For the 2D monic constraint, the solutions are identical to the 1D monic constraint case except for $\underline{\mathbf{i}} = [0 \;\; 1]^T$.

3.4 Performance Comparison of Different Targets

Computer simulations are carried out for the TwoDOS system shown in Fig. 2.2 with the channel length $N_h = 41$, equalizer length $N_w = 31$, target length $N_g = 3$, and $N_r = 5$ tracks per group. The target matrix $\underline{\mathbf{G}}$ is set up using the symmetry constraint, and the coefficients are given by (3.65). The normalized optical cutoff frequency with respect to the inverse of the center-to-center distance between adjacent bits is set to $1/\sqrt{3}$, which corresponds to the worst case where the largest allowable ISI and ITI occur in the TwoDOS system [44,103]. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = 10\log_{10}\left(\frac{\sum_{x,y}\breve{H}_s^2(x,y)}{\sigma^2}\right) \tag{3.70}$$

where $\sum_{x,y}\breve{H}_s^2(x,y)$ is the energy for a single spot with $\breve{H}_s(x,y)$ being the symbol response in Cartesian coordinates, and σ^2 is the variance of the additive white Gaussian noise (AWGN) on each individual track (*i.e.* $E[\theta_i^2(n)] = \sigma^2, i = 1, 2, \cdots, N_r$). In addition to AWGN, we also consider domain bloom and transition jitter for performance evaluation. We use four different PR targets for performance comparison: fixed PR target ($g_0 = 1, g_1 = 2$), 1D monic constrained target ($g_0 = 1$), energy constrained target

$(4g_0^2 + g_1^2 = 1)$ and 2D monic constrained target $(g_1 = 1)$. BER is used as the performance measure. The equalizer and target are designed with the approaches described in Sections 3.2 and 3.3 for each SNR.

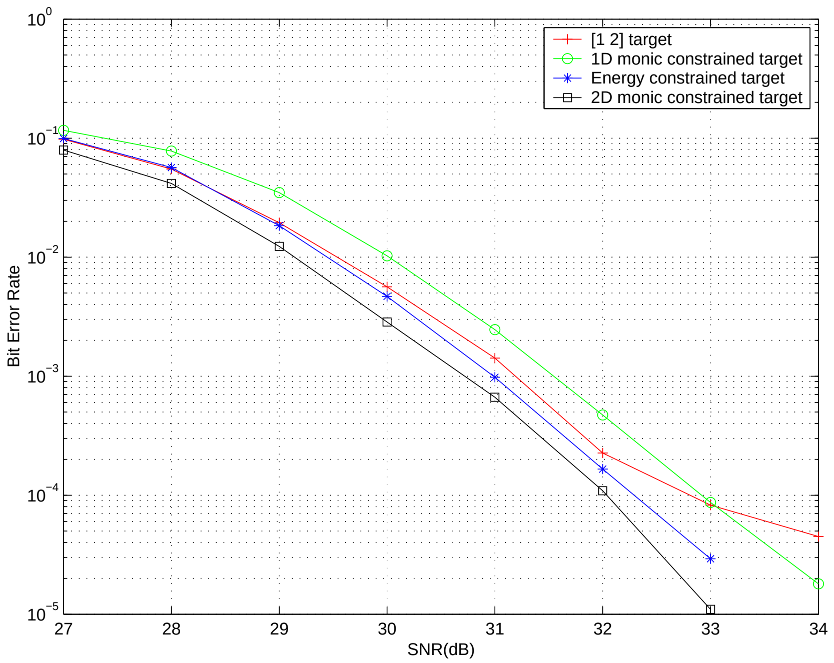

Figure 3.2: BER performance of 2D VD for different target constraints.

Fig. 3.2 shows the performance of 2D VD for the four different targets. Observe that the 1D monic constrained target performs worse than all the other targets when SNR is smaller than 33 dB, whereas the fixed PR [1 2] target becomes the worst for SNR $<$ 33 dB. At high SNRs, residual ISI dominates the MSE. Hence, jointly designing the target and equalizer, rather than fixed target approach, is preferable at higher SNRs since noise is not significant at these SNRs. For this reason, for SNR $>$ 33 dB, the 1D monic constrained target outperforms the fixed PR target. Fig. 3.2 also shows that the energy constraint is a reasonable target constraint. More importantly, the target with 2D monic constraint performs best among the four targets at all the SNRs, thus indicating that concentrating the energy around g_1 is a good choice for symmetric targets (e.g. see (3.65)).

Table 3.1: Noise correlation at equalizer output for different targets.

SNR	Spatial difference along the track	0	1	2	3	4
32 dB	Fixed PR target	1	0.2811	-0.1390	0.0548	-0.0083
	1D monic constrained target	1	0.2903	-0.1446	0.0492	-0.0038
	2D monic constrained target	1	0.2320	-0.1059	0.0714	-0.0266
	Energy constrained target	1	0.2607	-0.1259	0.0639	-0.0169
34 dB	Fixed PR target	1	0.2595	-0.1365	0.0655	-0.0061
	1D monic constrained target	1	0.2545	-0.1306	0.0684	-0.0100
	2D monic constrained target	1	0.2094	-0.0811	0.0773	-0.0329
	Energy constrained target	1	0.2312	-0.1043	0.0761	-0.0239

To further understand the performance trends observed in Fig. 3.2, we examine the noise correlation at the 2D VD input for different targets, since the performance of 2D VD is known to be influenced significantly by noise correlation. Moreover, the 2D equalizer could result in significant noise correlation. This problem becomes even more severe for shorter targets (e.g. $N_g = 3$), in addition to the fact that the MMSE design approach focuses on minimizing the total noise power, not the noise correlation. The normalized noise correlation along the third track is shown in Table 3.1 for the four different targets at SNR = 32 dB and SNR = 34 dB. Similar trends in the correlation are observed for other tracks as well as in the cross-track direction. Of all the targets, we see that the 2D monic constrained target results in least noise correlation at both SNRs, which is consistent with its best BER performance among the four targets. Further, comparing the noise correlation produced by the fixed PR target with the 1D monic constrained target at SNR = 32 dB and SNR = 34 dB, we can roughly understand why the BER curves for these two targets cross each other around SNR = 33 dB in Fig. 3.2. Finally, the relative noise correlation obtained for the energy constraint target with respect to the other targets is consistent with the BER performance observed in Fig. 3.2.

Table 3.2 shows the values of g_1 (normalized by g_0) for the four targets at different

Table 3.2: Normalized g_1 with respect to g_0 for different target constraints.

SNR	32	33	34	35
Fixed PR Target	2	2	2	2
1D monic constrained target	1.9468	1.9901	2.0310	2.0696
2D monic constrained target	2.2784	2.2828	2.2901	2.2997
Energy constrained target	2.1150	2.1411	2.1668	2.1919

SNRs. Comparing Table 3.2 with Fig. 3.2, it can be concluded that the target that has the smallest normalized g_1 results in the worst BER performance and vice versa. This observation reconfirms that the 2D monic constraint is a reasonable target constraint since it causes the energy to be concentrated near g_1, thus resulting in the largest normalized g_1 compared with the other targets. The 1D monic constraint can be considered as causing the target with the energy to be concentrated near g_0 and as a result it minimizes the normalized g_1. For this reason, the 1D monic constraint is not a good target constraint for 2D channels.

We did further investigations to see if the BER performance can be further improved (beyond that in Fig. 3.2) by removing the symmetry constraint on the target. This addresses the situation where the channel matrices are track dependent in the TwoDOS system. Since the 2D monic constraint resulted in the best BER performance in Fig. 3.2, we designed asymmetric targets under the constraint $g_1 = 1$. We considered two kinds of asymmetry. First, we broke the assumption that the values of cross interferences to the adjacent bits are constant. Therefore, we used six different values to represent the values of cross interference from the six adjacent bits. Secondly, we broke the assumption that the target responses for different tracks are the same. Therefore, we used 26 different values to represent the values of cross interferences from the adjacent bits and for five tracks. Simulations showed that the BER performances (not shown here) of these two asymmetric targets are similar to that of the 2D monic constraint target in Fig. 3.2. This further reconfirms the effectiveness of the simple 2D monic constraint. Note that here we consider the TwoDOS case where the largest allowable ISI and ITI occur. In other TwoDOS cases where ISI and ITI are smaller, the channel matrices become less track dependent. Then, applying the symmetry constraint and the subsequent 2D monic constraint target is more desirable in such cases.

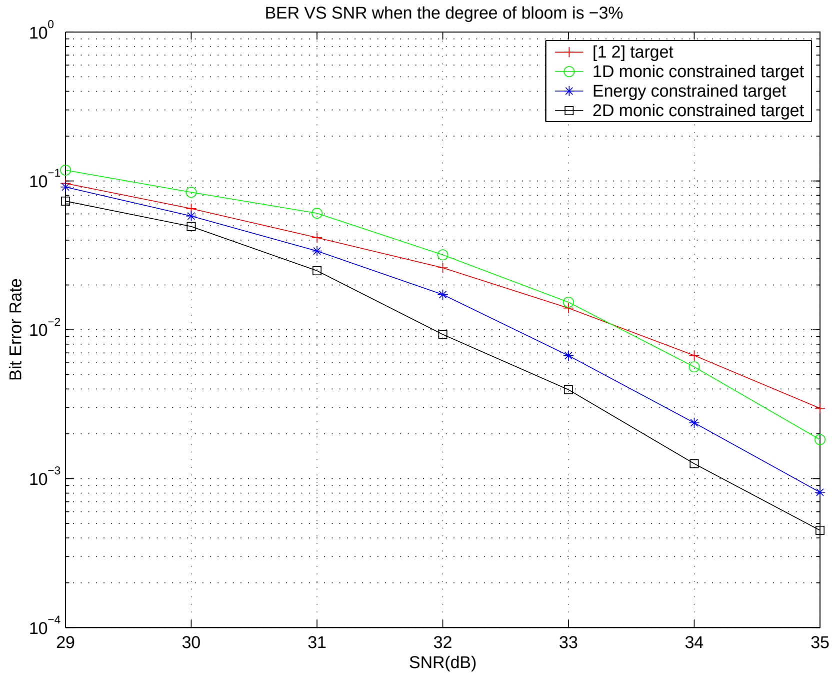

Figure 3.3: BER performance for different target constraints with -3% domain bloom.

Fig. 3.3 and Fig. 3.4 compare the performance of different targets as a function of SNR in the presence of domain bloom. The equalizers and targets used are the same as that in Fig. 3.2 and information of domain bloom is not used in the detector. This is a practical consideration since in many cases we may not know the degree of domain bloom. As illustrated, the 2D monic constrained target still outperforms all the other targets for both the under- and over-etching cases. Further, the BER curves in these figures indicate that the influence of the value of normalized g_1 on BER in the presence of domain bloom remain the same as that in the absence of domain bloom, *i.e.* larger g_1 results in better performance and vice versa.

Fig. 3.5 shows the performance of the different targets in the presence of transition

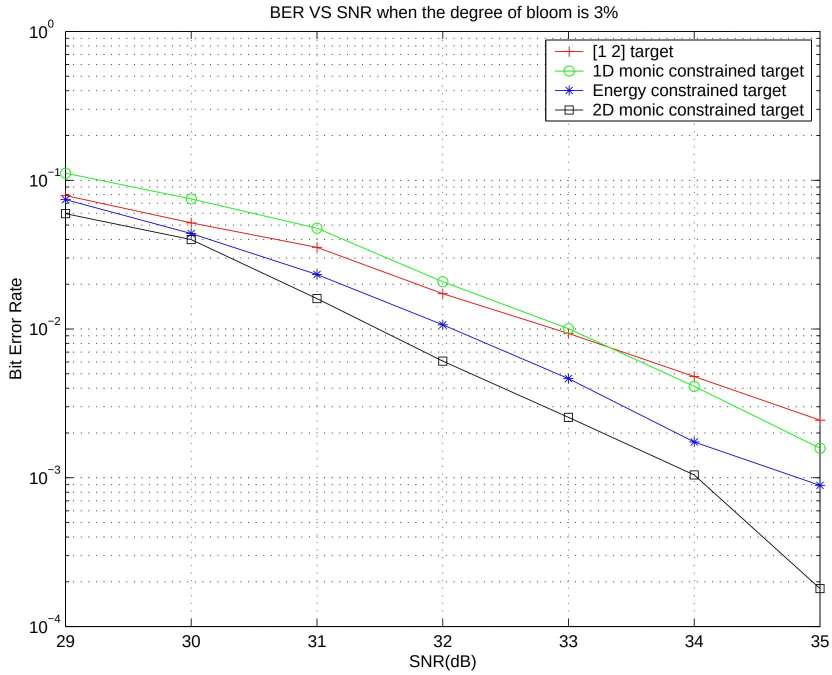

Figure 3.4: BER performance for different target constraints with 3% domain bloom.

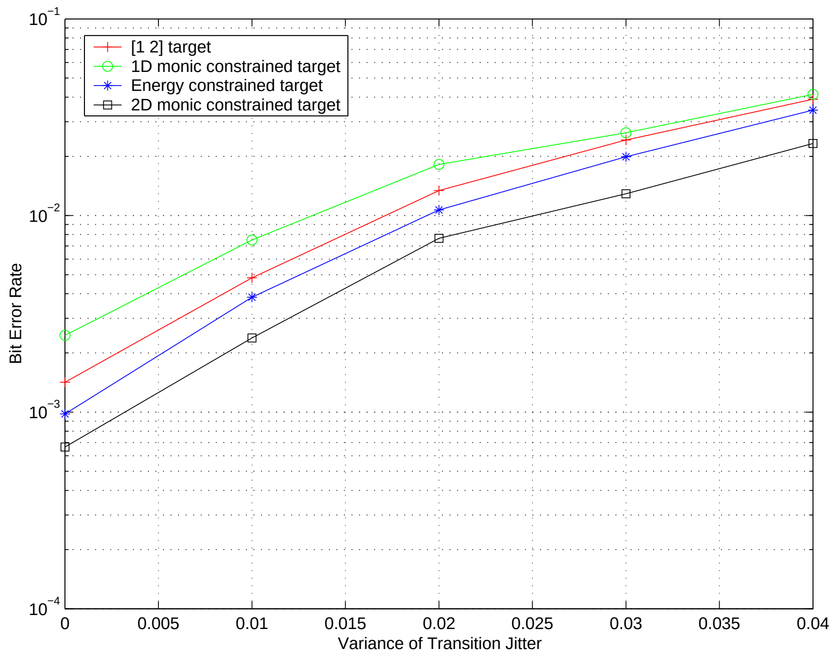

Figure 3.5: BER performance for different target constraints in the presence of transition jitter when SNR=31 dB. The transition jitter is normalized with respect to the radius of the pit hole R.

jitter whose probability density function is Gaussian with zero mean and variance σ_t^2, where σ_t is specified as a percentage of the radius of the pit hole R. In this case, the equalizer and target are designed using the modified correlation matrices given by (3.30)–(3.33). Observe that the 2D monic constrained target still outperforms other targets. In conclusion, the 2D monic constraint is a suitable target constraint in TwoDOS with or without domain bloom and transition jitter.

3.5 Conclusions

In this chapter, we have introduced two techniques to reduce the complexity of 2D VD by means of prefiltering. One technique is to design the 2D MMSE equalizer for a given 2D PR target. This technique does account for domain bloom, transition jitter, as well as correlated data and additive noise. The other technique is to jointly design the equalizer and target based on the MMSE approach to improve the performance of the 2D VD. Instead of directly imposing a constraint on the 2D PR target to avoid the trivial solution, we have proposed a novel technique which converts the 2D target design problem into a 1D problem. Also, we have introduced a computationally efficient analytic approach to evaluate different targets. By concentrating energy near the central bit, we have developed an appropriate target constraint called "2D monic constraint" that results in good bit detection performance for the 2D VD.

As TwoDOS demands a large number of tracks in a group to achieve high track density, prefiltering techniques may not sufficiently decrease the complexity of the 2D VD. In the next chapter, we will introduce the "quasi-1D VD", which uses the cross-track decisions to further reduce the complexity of the 2D VD.

Chapter 4

Quasi-1D Viterbi Detector

4.1 Introduction

The perpetual push for higher track density necessitates the two-dimensional optical storage (TwoDOS) systems to have large number of tracks in a single group. In the current stage, the number of tracks is chosen to be 11 within the group [103]. As mentioned in the previous chapter, the complexity of two-dimensional (2D) Viterbi detector (VD) grows exponentially with both the target length N_g and number of tracks N_r in a single group. Hence, truncating the channel memory by means of prefiltering techniques does not sufficiently reduce the complexity of 2D VD for the current TwoDOS system. For example, in the last chapter, though we have shortened the channel memory by setting $N_g = 3$, it is by far impractical because the number of states for the full-fledged 2D VD will reach 2^{22} for $N_r = 11$. For this reason, in this chapter, we develop a quasi-one-dimensional (quasi-1D) VD, which exploits the cross-track decisions as the feedback to facilitate the implementation of reduced-complexity 2D Viterbi-like detectors for systems with large number of tracks per group. A brief review of prior work in detectors with sequence feedback is first given here for the purpose of stating the main definitions and assumptions relevant to this work as well as for the sake of clarifying the notation.

This chapter is organized as follows. Section 4.2 briefly reviews the prior work on detectors with sequence feedback. Then, Section 4.3 develops a computationally efficient detector called the quasi-1D VD for TwoDOS by the use of proposed causal ITI constrained target. The performance of the quasi-1D VD is evaluated in Section 4.4 by simulation. Finally, Section 4.5 concludes this chapter.

4.2 Review of Detection Techniques with Sequence Feedback

4.2.1 Decision Feedback Equalization

Decision feedback equalization is a nonlinear detection technique that is quite popular in digital communication systems [15, 21]. Fig. 4.1 shows the block digram of a discrete-time decision feedback equalizer (DFE). In the figure, h_k is the discrete-time channel symbol response, $\theta(n)$ is the additive white Gaussian noise (AWGN) with variance σ^2, and w_k and f_k represent the taps of the forward and feedback equalizer, respectively. The forward equalizer shapes the channel into a prescribed target g_k, which is constrained to be causal and the first tap g_0 is constrained to be one. Feedback equalizer has a strictly causal impulse response f_k that should match g_k for all $k \geq 1$ in order to cancel the causal intersymbol interference (ISI), *i.e.* the ISI due to the symbols that have already been detected. By removing the causal ISI, the DFE uses the threshold comparator to make the bit decision based on the input of the slicer. Though the DFE is the optimum detector that has no detection delay [24], its performance lags behind that of the VD because of the following two main reasons.

- **Error propagation:** Any decision errors at the output of the slicer will cause a corrupted estimation of the causal ISI, which is to be generated by the feedback equalizer. The result is that a single error causes the detector to be less tolerant of the noise for a number of future decisions. This phenomenon is referred to as the *error propagation* and degrades the performance of the detector.

- **Energy reduction:** Even in the absence of error propagation, the DFE is still suboptimum compared to the VD in terms of performance. This is because in the decision process, the DFE subtracts the causal ISI and thus ignores the signal energy embedded in this causal ISI component. In other words, some signal energy that is beneficial for the optimum detection is neglected. The adverse effect on the detection performance is referred to as the *energy reduction.* To minimize the energy reduction effect due to neglecting the energy of causal ISI, the target is designed to have minimum-phase characteristics, *i.e.* the energy of the target is optimally concentrated near the time origin.

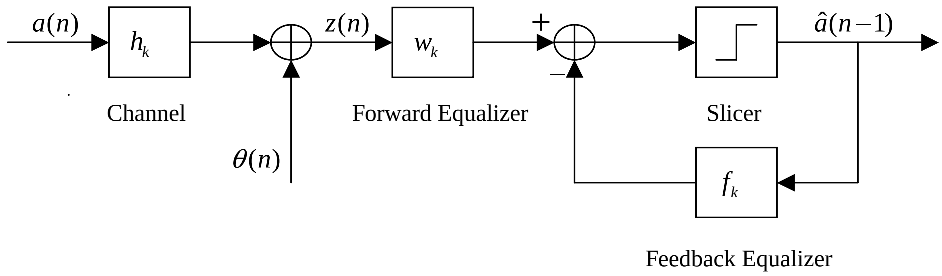

Figure 4.1: Block diagram of a discrete-time decision feedback equalizer.

4.2.2 Fixed-Delay Tree Search

Unlike the DFE that makes the bit decision instantly, the fixed-delay tree search (FDTS) detection technique makes the bit decision after a delay of D [10]. In this technique, the bit decision is based on a sequence of $D+1$ input samples before the detector and uses the maximum-likelihood (ML) decision rule for the bit decision with a delay of D. The ML decision exploits partly or all of the signal energy embedded in the causal ISI components, and thus reduces the energy reduction effect compared to the DFE. The choice of parameter D, is limited by the compromise between performance and complexity. If $D+1$ is smaller than the target length N_g, the FDTS is referred to as the fixed delay tree search with decision feedback (FDTS/DF) [136]. In fact, the FDTS can be considered as a generalization of the DFE since the FDTS is essentially equivalent to the DFE when $D = 0$.

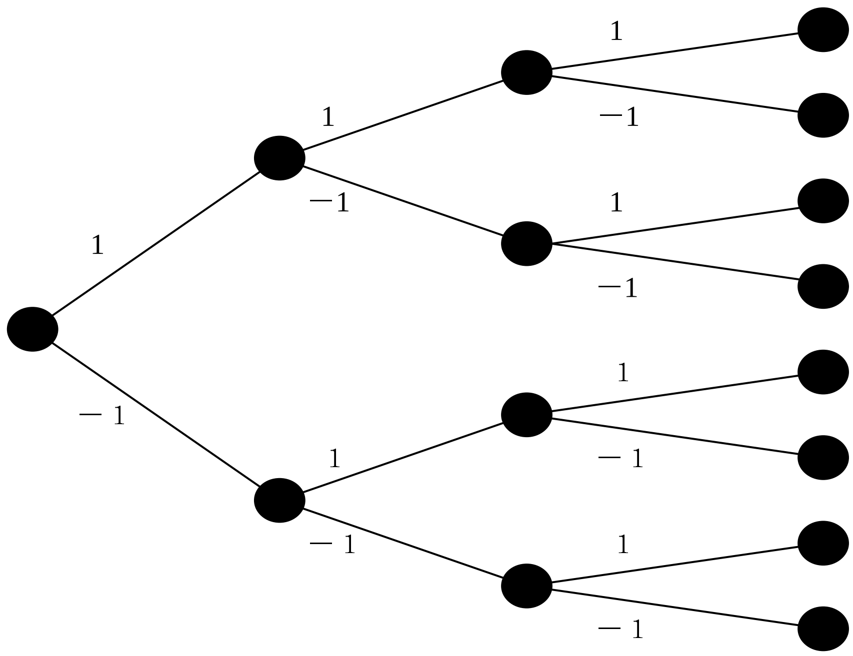

Figure 4.2: Tree representation with depth $D = 2$ for the uncoded binary channel input data.

Similar to the DFE, the FDTS first uses the forward equalizer to shape the channel into a known target. Then, the noiseless input of the detector is $d(n) = \sum_{i=0}^{N_g-1} g_i a(n-i)$, where g_i $(i = 0, 1, \cdots, N_g - 1)$ represent the coefficients of the target whose length is N_g, and $a(n)$ is the channel input bit at time index n. The FDTS uses a fixed-depth ML decision rule implemented as a tree search algorithm. The tree representation with depth $D = 2$ is shown in Fig. 4.2 for illustration. Each branch corresponds to one input bit at a particular time. A sequence of branches through the tree diagram is referred to as a path. Each possible path corresponds to one input sequence and vice versa. At time index n, the tree diagram consists of $D + 1$ bits. Thus, at each time index, the trellis contains 2^{D+1} paths that represent all the possible 2^{D+1} input sequences.

As mentioned in Chapter 1, detection based on the smallest Euclidian distance between the detector input $z(n)$ and the desired noiseless detector input $d(n)$ is optimum in the ML sense when the noise component of the detector input is white and Gaussian.

Thus, similar to the trellis diagram that corresponds to the VD, the Euclidian distance $[z(n) - d(n)]^2$ is defined as the branch metric for each branch, and the summation of the branch metrics associated with each path is called the path metric. Since the FDTS performs ML detection based on a sequence of samples, it chooses the path whose path metric is minimum as the most likely transmitted sequence and releases the first bit associated with this path as the detected bit. More specifically, the FDTS operates recursively as follows [136]:

1. **Initial condition**: At the end of $(n-1)$th step, the tree structure has a depth of $D-1$. Each path retains the path metric obtained from the previous iteration.

2. **Path extension**: At the nth step, the tree structure is extended such that the depth is increased to D. The new input sample $z(n)$ is used to compute the branch metric $[z(n)-d(n)]^2$ for each extended branch. To compute $d(n)$ for each extended branch, the possible channel input bit sequence consists of the preceding $N_g - 1$ input bits lying on the path leading to that extended branch. If the channel memory $N_g - 1$ exceeds the detection delay D, the already detected bits are also used to compute $d(n)$.

3. **Path selection**: After computing all the path metrics for the extended paths, the first bit of the path that has the smallest path metric is selected and released as the detected bit. Then, half of the total paths that are incompatible with the detected bit are discarded. As a result, the tree structure that remains has a depth of $D-1$.

As time progresses, the root node moves along the ML path and a fixed-size identical tree structure is maintained at each time index. Therefore, the complexity of the FDTS is kept constant for each time index. Similar to the VD, the ML decision rule makes the FDTS unduly complicated if D is large. An efficient and simple realization of the FDTS for systems using runlength-limited (RLL) $(1, k)$ codes can be found in [111, 130].

4.2.3 Sequence Detection with Local Feedback

Many detection techniques with sequence feedback, such as the DFE and FDTS/DF, use the detected bits as the input of the feedback equalizer, resulting in the error propagation problem. Nevertheless, this problem can be reduced by resorting to local feedback [63,81]. The local feedback is based on the trellis structure that is introduced in Chapter 1, and uses the path memory associated with the current state instead of the past decisions to estimate the causal ISI. The local feedback guarantees that the branch metric of the correct path is the ML metric, as long as it is discarded in favor of some incorrect path [63]. As a result, it improves the performance of those detectors with sequence feedback at the price of requiring a large memory to store paths associated with each state.

4.3 Quasi-1D Viterbi Detector

4.3.1 Complexity of 2D VD

In Chapter 3, we used 2D PR equalization to shape the 2D channel into a known 2D target with controlled ISI and intertrack interference (ITI). These controlled ISI and ITI are left to be handled by the 2D VD. The noiseless input of the 2D VD is given by $\mathbf{d}(n) = \sum_{i=0}^{N_g-1} \mathbf{G}_i \mathbf{a}(n-i)$, where $\mathbf{G}_i$ $(i = 0, 1, \cdots, N_g - 1)$ is the target matrix whose length is N_g, and $\mathbf{a}(n)$ is the channel input vector at time index n. As indicated earlier, the complexity of 2D VD grows exponentially with both the target length N_g and number of tracks N_r in a single group. For a better understanding, the trellis structure for the case of target length $N_g = 3$ and number of tracks per group $N_r = 2$ is shown in Fig. 4.3. In this figure, the '+' and '−' represent the bits '+1' and '−1', respectively. The trellis is assumed to start at the node S_0, and then becomes steady at instant $n = 3$ (*i.e.* $n = N_g$). Here, the labels of states represent the channel memory and number of tracks per groups associated with the paths that pass through these states. At time index n, each state consists of N_r $(N_g - 1)$ bits (*i.e.* $\{\breve{\mathbf{a}}(n-1), \breve{\mathbf{a}}(n-2), \cdots, \breve{\mathbf{a}}(n-N_g+1)\}$). Thus, at each time index, the trellis contains $2^{N_r(N_g-1)}$ states. At time index n, each branch specifies the channel memory associated with the state that the branch originates from and the

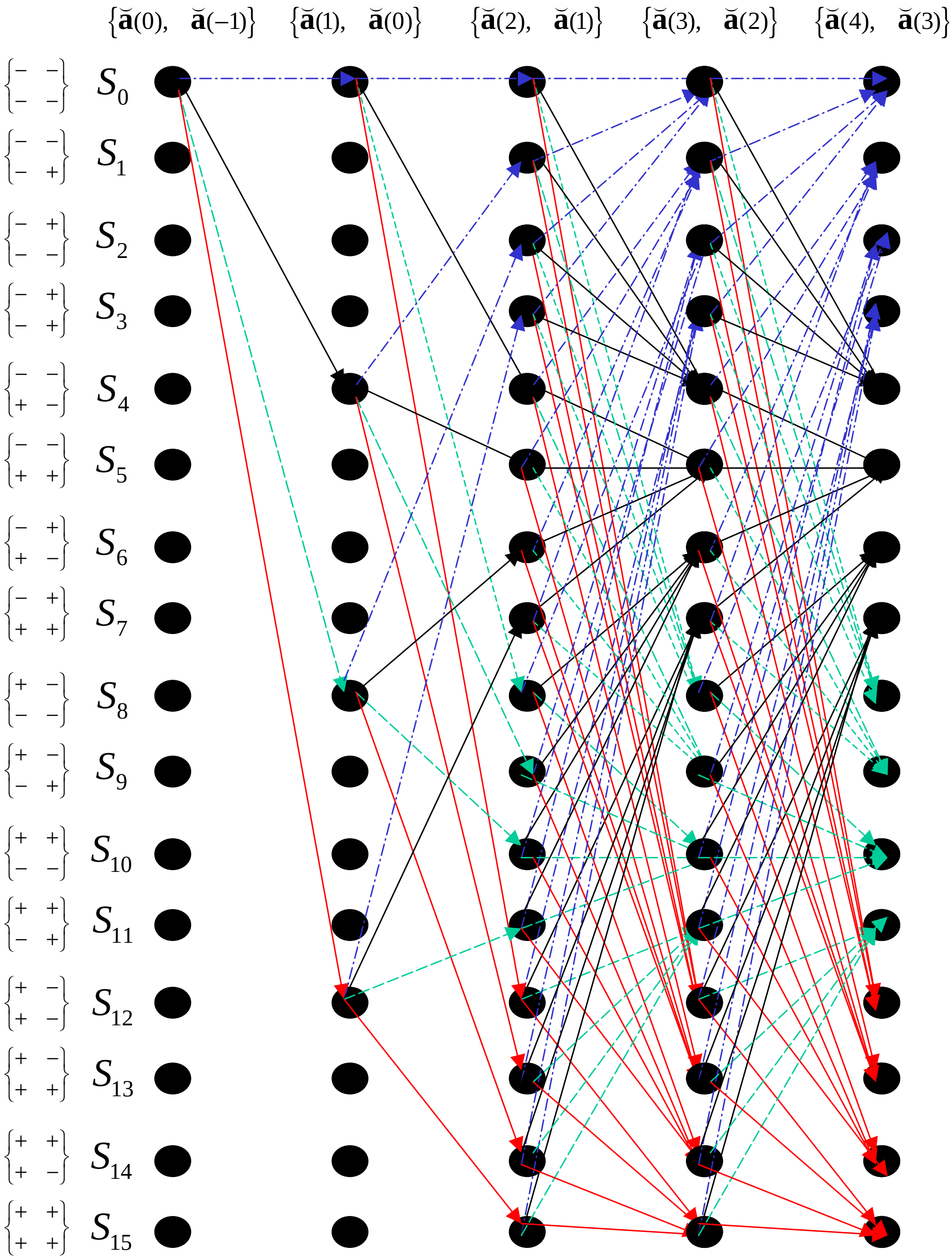

Figure 4.3: Trellis structure for a channel with $N_g = 3$ and $N_r = 2$.

possible channel input vector $\breve{\mathbf{a}}(n)$. Therefore, each branch corresponds to one possible noiseless detector input $\mathbf{d}(n) = \sum_{i=0}^{N_g-1} \mathbf{G}_i\breve{\mathbf{a}}(n-i)$. For the binary channel input bit, each state possesses 2^{N_r} incoming and 2^{N_r} outgoing branches and thus there are totally $2^{N_rN_g}$ incoming and $2^{N_rN_g}$ of outgoing branches for each time index of the trellis.

Comparing Fig. 4.3 and Fig. 1.3, it is clear that even in this simple 2D case, the trellis of 2D case is much more complicated than the one-dimensional (1D) case though the target length is the same. Thus, the practical implementation of the 2D Viterbi-like detector for large N_r also requires the significant reduction of the complexity arising from the cross-track direction. In [84], a technique using the Viterbi detector track-by-track, as well as the decision feedback to estimate the ITI between tracks was proposed. We call this detector the DFE-VD. It uses a set of sub-2D VDs, each corresponding to one track. In the bit decision process for a given track, the known bits just above (or below) the current track are used as the feedback to calculate part of the ITI. These known bits can be previously detected bits, or can be zeros if the upper (or lower) track is the guard-band. The branch metric is then computed by subtracting the effect of these known bits. However, in this track-by-track technique, the ITI from either only the upper track(s) or only the lower track(s) estimated, and the remaining ITI estimations are still dependent on the trellis states. As a result, the number of states should be larger than that of 1D VD with the same target length. Moreover, this redundant complexity will not benefit performance much since the detector makes the detection based still only on the input samples from the current single track. An improved detector is the stripe-wise Viterbi detector (SWVD) [44] [103]. This detector consists of a set of sub-2D VDs, each dealing with one stripe that consists of a limited number of tracks. The number of stripes is equal to that of tracks in a single group. The preliminary decisions from one sub-2D VD is used for estimating the ITI in the next sub-2D VD, which is shifted up (or down) by one track. This procedure is continued for all the stripes and the full procedure from bottom to top (or top to bottom) of the group is considered to be one iteration. Note that at least two iterations are required in order to estimate the ITI from both upper and lower tracks. Unlike the DFE-VD that resorts to the trellis states to estimate the ITI from the lower (or upper) track(s), the SWVD uses the preliminary decisions from the previous iteration to estimate the ITI from the lower (or upper) track(s). This additional decision feedback not only reduces the complexity but also improves the performance compared with the DFE-VD since its decisions exploit the input information from both upper and lower track(s) as well as that from current. However, the use of iterations increases complexity as well as latency. Our new proposal, whereas, is a non-iterative reduced-complexity detector that is applicable to any 2D system.

4.3.2 Causal ITI Target

In this subsection, we introduce the causal ITI target as a starting point for the development of our reduced-complexity 2D Viterbi-like detectors. Conventionally, the causal and anticausal ISI are referred to as the ISI from the past and future bit decisions, respectively [24]. Similarly, we refer to the causal and anticausal ITI as the ITI resulting from the lower and upper tracks, respectively. The concept of causal ITI was first used in the multichannel DFE [78]. Similar as shown in Fig. 4.1, this multichannel DFE consists of a multichannel forward filter, a multichannel feedback filter, and a decision block. The multichannel forward filter is designed to constrain the channel to be causal ISI and ITI. The multichannel feedback filter is designed to remove the causal ISI based on the previous bit decisions. The causal ITI is left to be handled by the decision block. Motivated by this, we propose the causal ITI target such that the 2D target matrices are constrained to be the right triangular matrices. It should be noted that this target is the basis for the development of our reduced-complexity 2D Viterbi-like detectors.

As a starting point for our development, we first examine the suitability of the causal ITI target in TwoDOS. Fig. 4.4 shows the performance of full-fledged 2D VD for four different targets when N_r is five and target length N_g is three. In the figure, the diagonal elements of G_0 in the causal ITI target are constrained to be 1s to avoid trivial solutions of the target and equalizer. We use a fixed 2D target with elements [1 2] and 2D monic constrained target, which are reasonable targets described in the last chapter for Two-

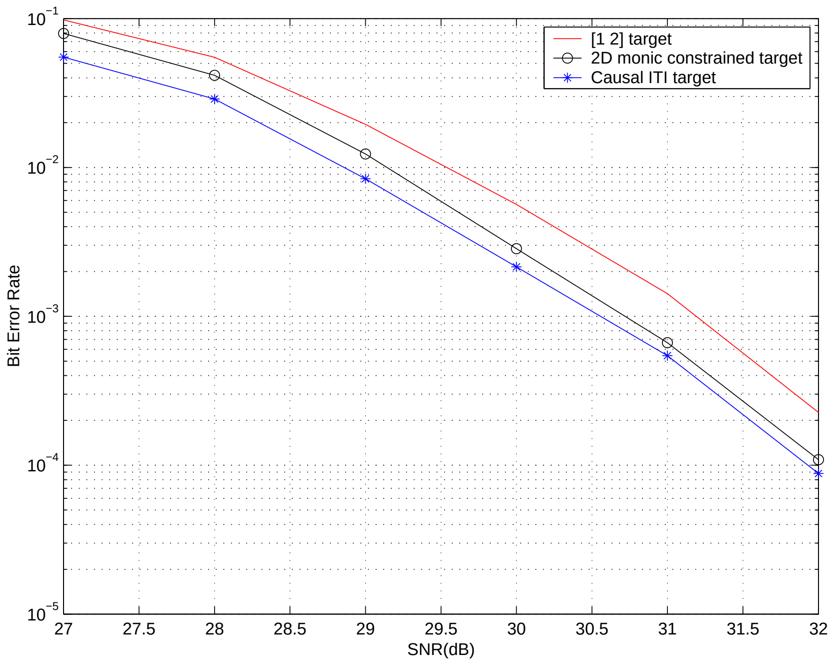

Figure 4.4: BER performance for different target constraints.

DOS, as reference targets. Note that we impose a symmetry constraint, which constrains all the tracks within the same group to suffer the same amount of ITI, in the design of the 2D monic constrained target. In other words, after the finite length equalizer, all the tracks within the same group ideally suffer the same amount of ITI. However, due to the presence of guard-bands serving as boundaries of the group, before the finite length equalizer, not all the tracks suffer the same amount of ITI. In addition, the 2D monic constrained target only allows ITI from adjacent tracks. Therefore, the symmetry constraint will burden the design of finite length equalizer and result in residual ISI and ITI. Note that the causal ITI target does not have this symmetry constraint, and allows ITI not only from the adjacent tracks. Therefore, compared with the 2D monic constrained target, the causal ITI target burdens the finite length equalizer less and is expected to achieve better performance. From Fig. 4.4, it is shown that the causal ITI target outperforms all the targets at every SNR. This result indicates that it is reasonable to use the causal ITI target for TwoDOS. More importantly, based on this target, we propose some reduced-complexity 2D Viterbi-like detectors that are quite different from DFE-VD and SWVD since the latter two detectors suffer ITI from both lower and upper tracks.

4.3.3 Principle of Quasi-1D VD

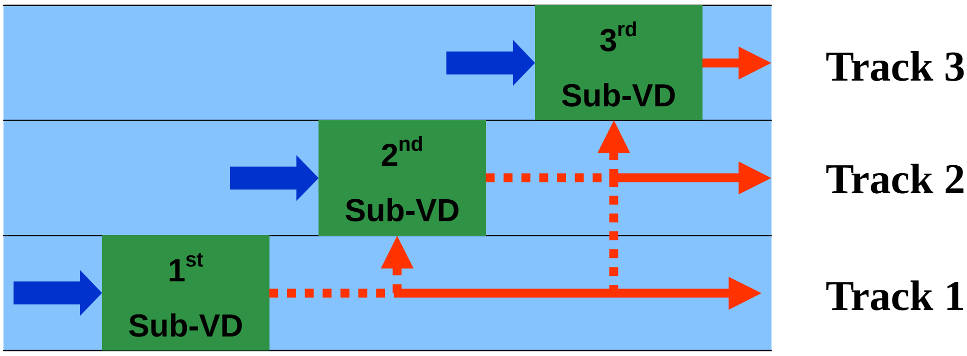

Figure 4.5: Principle of the quasi-1D VD. The solid lines represent the input and output of sub-VDs, the dashed lines represent the feedback coming from the output of the previous sub-VDs.

Since the causal ITI target contains ITI only from the lower tracks, the bits in the upper tracks will not affect the desired output. Based on this idea, a set of 1D VDs are used to detect the bits, each deals with one track. More specifically, as shown in Fig. 4.5, the first 1D VD that deals with the lowest track is processed with no delay and the bits are detected after a delay D. The second 1D VD that deals with the second lowest track is processed with the delay D in order to use the detected bits from the lowest track to estimate all the ITI in the second lowest track. The third 1D VD that deals with the third lowest track is processed with a delay D after the second 1D VD, and the detected bits from the lowest two tracks are used to estimate the ITI in the third lowest track. This procedure continues for all the tracks. Since the bits detection does not need to consider the interference from the upper tracks, this detector is distinct from the DFE-VD and SWVD. Compared with the DFE-VD, this detector has less computational complexity since fewer states are needed for bit detection. More importantly, the quasi-1D VD has better BER performance since it uses all, while DFE-VD uses part, of the input information that is needed in the cross-track direction. As illustrated in Fig. 4.6, the quasi-1D VD outperforms the DFE-VD significantly no matter what target is chosen for the DFE-VD. Compared with the SWVD, as mentioned previously, it has much lower complexity since it has no iterative procedures.

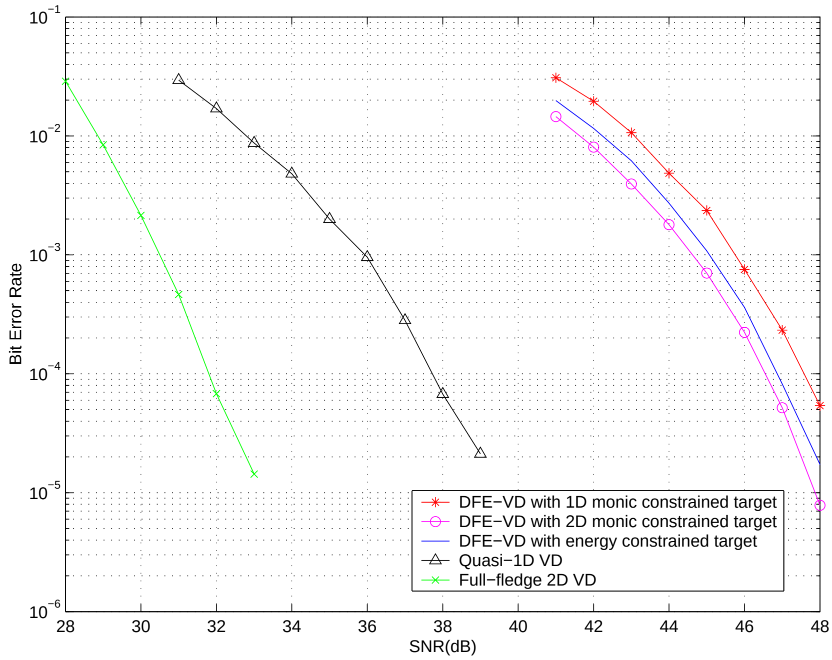

Figure 4.6: Performance comparison of different detection techniques.

- *Link with QR Detector*

  Our quasi-1D VD is developed for the TwoDOS system, which is a multiple-input multiple-output system having a large temporal span of the channel. Obviously, this quasi-1D VD is applicable to multiple-input multiple-output systems having an arbitrary temporal span of the channel. In many wireless communication systems, the multiple-input multiple-output channel is assumed to be flat-fading [68, 175], *i.e.* the temporal span $N_h = 1$. In such systems, the channel is characterized by a matrix, instead of a sequence of matrices in the TwoDOS system. Let N_1 and N_2 represent the number of transmit and receive antennas, respectively, in multiple-input multiple-output wireless communication systems. Then, the channel output vector at a given time is given by

$$\mathbf{z} = \mathbf{H}\mathbf{a} \tag{4.1}$$

where $\mathbf{z}$ and $\mathbf{a}$ are the $(N_2 \times 1)$ channel output vector, and $(N_1 \times 1)$ channel input vector, respectively, $\mathbf{H}$ is the $(N_2 \times N_1)$ flat-fading channel matrix. For the sake of simplicity, the time index is ignored here. Then, QR decomposition of the channel matrix yields $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q}$ is an $(N_2 \times N_1)$ orthonormal matrix constructed to make the $(N_1 \times N_1)$ matrix $\mathbf{R}$ right triangular [129]. Premultiplying the channel output vector $\mathbf{z}$ with $\mathbf{Q}^H$, the resulting vector $\hat{\mathbf{z}}$ is given by

$$\hat{\mathbf{z}} = \mathbf{Q}^H \mathbf{z} = \mathbf{Ra}. \tag{4.2}$$

Note that if the noise in $\mathbf{z}$ is additive white Gaussian noise (AWGN), the noise in $\hat{\mathbf{z}}$ remains AWGN since $\mathbf{Q}^H\mathbf{Q}$ is an $(N_1 \times N_1)$ identity matrix. Comparing $\mathbf{R}$ with the causal ITI target discussed in the previous subsection, we find that $\mathbf{R}$ can be seen as a special case of causal ITI targets. Then, like the quasi-1D VD, the first element from the bottom of the channel input vector $\mathbf{a}$ is first detected. The detected element is used to estimate interferences for the detection of the second element from the bottom of $\mathbf{a}$. This procedure continues until all the elements in $\mathbf{a}$ are detected. This detector is commonly referred to as the QR detector and has been investigated in multiple-input multiple-output flat-fading channels [51, 193, 200]. The QR detector is also applicable in multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems [123, 194], which will be discussed in Part II of the thesis, since the channel at each subcarrier of MIMO-OFDM systems is considered as a multiple-input multiple-output flat-fading channel. Note that our proposed quasi-1D VD is suitable for any multiple-input multiple-output channel with arbitrary positive N_h, while the QR detector is only applicable for multiple-input multiple-output flat-fading channel, *i.e.* $N_h = 1$. Therefore, the QR detector is considered as a special case of our proposed quasi-1D VD.

4.4 Performance of Quasi-1D VD

As shown in Fig. 4.5, though the quasi-1D has much lower complexity than the DFE-VD and SWVD, it causes significant detraction from optimality. We consider three factors

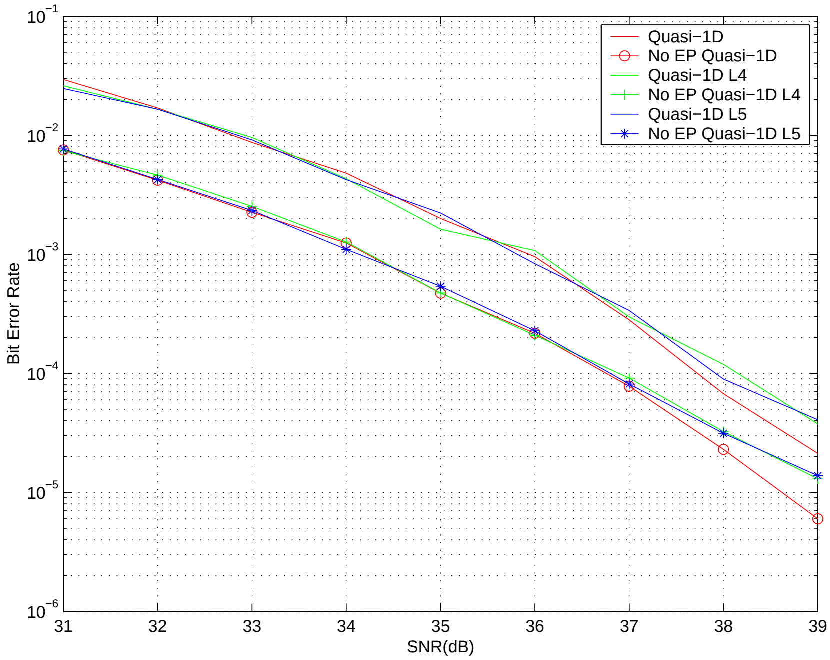

Figure 4.7: BER performance of quasi-1D VD with different target lengths.

that affect the performance of quasi-1D VD: target length, error propagation and energy reduction. In Fig. 4.7, "L4" and "L5" represent that the lengths of targets are four and five, respectively. Otherwise, the length of target is three. "No EP" means detectors without suffering error propagation. In simulation, "No EP" is achieved by use of correct input bits to estimate ITI. The length of the equalizer is 31 in all the simulations. As illustrated in Fig. 4.7, the BER performance is not significantly improved by increasing the target length. Further investigation shows that all the elements in target matrices G_3 and G_4 approach zero, therefore confirming that there is no need to increase the channel memory beyond two. Fig. 4.7 also shows that the error propagation degrades performance by 1 dB for BER is 10^{-4}. Thus, the energy reduction should be the dominant factor that degrades the performance.

4.5 Conclusions

In this chapter, we have first briefly reviewed prior work on the detectors with sequence feedback. Then, by constraining the target with causal ITI, we have developed a quasi-1D VD, which uses a computationally efficient technique whose complexity grows only linearly with the number of tracks. This is a significant complexity reduction compared to the conventional 2D VD whose complexity grows exponentially with the number of tracks. We have shown that the quasi-1D VD improves over the DFE-VD and SWVD in terms of complexity. Further, we have shown that the widely known QR detector is a special case of our proposed quasi-1D VD. However, we have found that the quasi-1D VD still causes significant detraction from optimality in the TwoDOS system. Therefore, effective compensation techniques are needed to ensure reliable data recovery. To achieve this goal, we have investigated the factors that might degrade the performance. Our simulation results implied that the energy reduction is the dominant factor that degrades the performance of the quasi-1D VD. Therefore, in the next chapter, we develop some effective techniques to reduce the effect of this energy reduction problem. In addition, the effect of error propagation still needs to be minimized since it degrades the performance by roughly 1 dB when BER is 10^{-4}. However, increasing the target length beyond three is of no practical value for the TwoDOS system since it hardly improves the performance while introducing excessive complexity.

Chapter 5

Generalized 2D Viterbi Detector

5.1 Introduction

In the two-dimensional optical storage (TwoDOS) system, due to the absence of guard-bands between tracks and the hexagonal arrangement of bit-cells, the received signals contain severe intertrack interference (ITI) and intersymbol interference (ISI). As a result, the traditional threshold-like detector that is widely used in the current optical storage systems cannot be used in the TwoDOS system. We choose the two-dimensional (2D) Viterbi-like detector because of its superior detection performance in the presence of severe ITI and ISI. However, a full-fledged 2D Viterbi detector (VD) is by far impractical for the current TwoDOS system (with large number of tracks per group) since its complexity grows exponentially with the channel memory and number of tracks per group. Unfortunately, further simplifications of the 2D VD always result in performance degradation. For example, in the previous chapter, the quasi-one-dimensional (quasi-1D) VD significantly reduces the complexity of a full-fledged 2D VD but suffers a considerable performance loss. Thus, the main objective in this chapter is to reduce the complexity of the 2D Viterbi-like detector without paying a large penalty in terms of performance. The detector proposed here can be regarded as a generalized 2D VD since the conventional full-fledged 2D VD, QR detector, and our proposed quasi-1D VD are special cases of this detector. We also address some novel techniques that provide this generalized 2D VD with more flexibility to deal with its focal problem, *i.e.* the balance between the performance and complexity.

This chapter is organized as follows. Section 5.2 proposes a generalized 2D VD called FDTS-DF/VD. Then, Section 5.3 provides a theoretical platform to evaluate the performance of the FDTS-DF/VD. Section 5.4 develops the RFDTS-DF/VD to further reduce the complexity of the FDTS-DF/VD. Some novel targets specifically for the FDTS-DF/VD and RFDTS-DF/VD are presented in Section 5.5. Finally, Section 5.6 concludes this chapter.

5.2 Principle of Generalized 2D VD

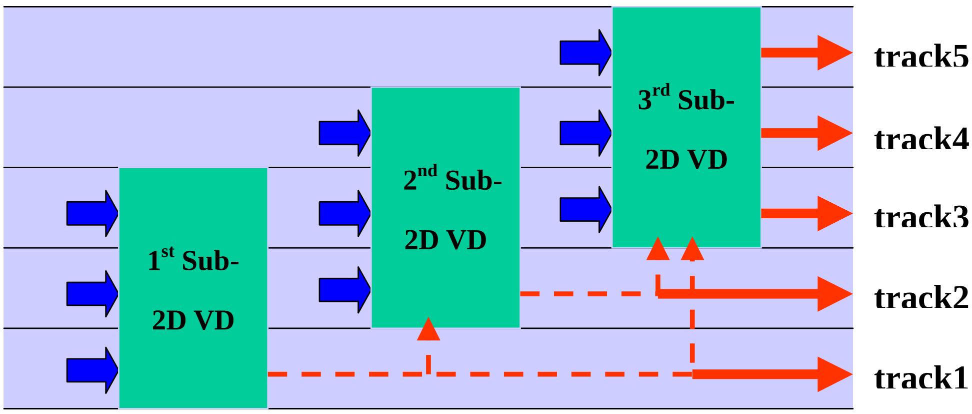

Figure 5.1: Principle of FDTS/DF-VD with $N_{sr} = 3$ and $N_r = 5$. The solid lines represent the input and output of sub-2D VDs, the dashed lines represent the feedback coming from the output of the previous sub-2D VDs.

The basic idea of this novel detector is to divide the full-fledged 2D VD, which deals with N_r (*i.e.* number of tracks per group) tracks simultaneously, into a set of sub-2D VDs. This detector is based on the casual ITI target introduced in the previous chapter. Each sub-2D VD deals with a subgroup of N_{sr} tracks, and the detected bits from lower tracks are used to estimate all the causal ITI in the current subgroup. More specifically, we use the first sub-2D VD to deal with the lowest N_{sr} tracks and the detected bits in the lowest track are the final decisions with a delay of D. The second sub-2D VD deals with N_{sr} tracks that are shifted up by one track and starts after the first sub-2D VD with the delay D. Its outputs are the final decisions in the second lowest track. This procedure continues until all the bits are detected. It should be noted that the output of each sub-2D VD is limited to the bottom track of its corresponding subgroup except for the last sub-2D VD, whose outputs correspond to all the N_{sr} tracks of its corresponding subgroup (see Fig. 5.1). Since in the cross-track direction we essentially use the fixed delay tree search with decision feedback (FDTS/DF) [136], we call this detector the FDTS/DF-VD. Clearly, the conventional 2D full-fledged VD ($N_{sr} = N_r$) and the quasi-1D VD ($N_{sr} = 1$) are special cases of this generalized 2D VD. Further, as mentioned in the previous chapter, the QR detector is the special case of the quasi-1D VD. Thus, the QR detector is also a special case of the FDTS/DF. The flexibility introduced by the parameter N_{sr} allows us to design systems with varying performance and complexity.

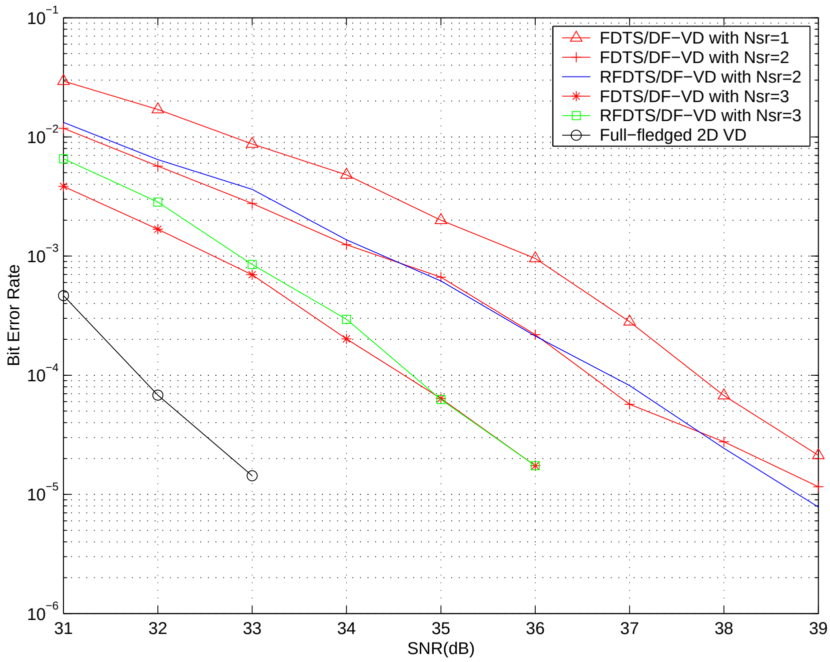

Figure 5.2: BER performance comparison of different 2D bit detectors.

As shown in the previous chapter, the quasi-1D VD, which is the simplest FDTSDF-VD, provides a rather poor performance. The main reason for the considerable performance degradation is that the quasi-1D VD ignores the signal energy embedded in the causal ITI components, which are rather large for TwoDOS due to the absence of guard-band between tracks. Therefore, we can increase N_{sr} to achieve better performance. Fig. 5.2 compares the bit error rate (BER) performances of different 2D bit detectors and shows that setting $N_{sr} = 3$ is sufficient enough to yield good performance. Note that here we consider the TwoDOS case where the largest allowable ISI and ITI occur. In other TwoDOS cases where ISI and ITI are smaller, the performance loss of different 2D bit detectors compared with the full-fledged 2D VD would become less. Thus, using N_{sr} smaller than 3 might still suffice to achieve good performance. From the above, setting $N_{sr} = 3$ is sufficient enough to yield good performance for all the TwoDOS cases when $N_r = 5$.

5.3 Performance Analysis of FDTS/DF-VD

The exact performance of the FDTS/DF-VD is difficult to analyze because of the decision feedback. However, as mentioned in the previous section, error propagation degrades the performance with 1 dB when BER is 10^{-4}. Therefore, we first investigate the FDTS/DF-VD that is free of error propagation. The effect of error propagation can be assessed through simulations.

Recalling (2.9), the channel output vector resulting from the parallel read-out at time index n is given by

$$\mathbf{z}(n) = \sum_{k=0}^{N_h-1} \mathbf{H}_k \mathbf{a}(n-k) + \boldsymbol{\theta}(n) \tag{5.1}$$

where $\mathbf{z}(n) = [z_1(n),\ z_2(n),\ \cdots,\ z_{N_r}(n)]^T$, $z_k(n)$ denotes the received signal component from the kth track, $\mathbf{a}(n) = [a_1(n),\ a_2(n),\ \cdots,\ a_{N_r}(n)]^T$, $a_k(n) \in \{-1,\ 1\}$ denotes the data bit written on the kth track, $\boldsymbol{\theta}(n) = [\theta_1(n),\ \theta_2(n),\ \cdots,\ \theta_{N_r}(n)]^T$ and $\theta_k(n)$ denotes the noise picked up from the kth track, for $k = 1, 2, \cdots, N_r$. Further, let $\mathbf{G}_k$ ($k = 0,\ 1,\ \cdots,\ N_g - 1$) and $\mathbf{W}_k$ ($k = 0,\ 1,\ \cdots,\ N_w - 1$) represent $N_r \times N_r$ coefficient matrices of partial response (PR) target and equalizer, respectively. We assume that the channel response $\mathbf{H}_k$ ($k = 0, 1, \cdots, N_h - 1$) is time-invariant, and the PR target $\mathbf{G}_k$ is known in this subsection.

For notation convenience, define $\underline{\mathbf{a}} = [\mathbf{a}^T(n), \mathbf{a}^T(n+1), \cdots, \mathbf{a}^T(n+N_e-1)]^T$, $\underline{\hat{\mathbf{a}}} = [\hat{\mathbf{a}}^T(n), \hat{\mathbf{a}}^T(n+1), \cdots, \hat{\mathbf{a}}^T(n+N_e-1)]^T$, where $\hat{\mathbf{a}}(n)$ denotes the detected channel input vector corresponding to $\mathbf{a}(n)$. Further, we define a 2D error event vector $\underline{\mathbf{e}} = [\mathbf{e}^T(n), \mathbf{e}^T(n+1), \cdots, \mathbf{e}^T(n+N_e-1)]^T$ with $\mathbf{e}(n) = (\mathbf{a}(n) - \hat{\mathbf{a}}(n))/2$. Then, $\underline{\mathbf{e}}$ is a 2D error event of length N_e if

1. $\mathbf{e}(n) \neq 0$ and $\mathbf{e}(n+N_e-1) \neq 0$,
2. the length of strings of zero vectors in $\underline{\mathbf{e}}$ does not exceed $N_g - 1$, and
3. $\hat{\mathbf{a}}(n+i) = \mathbf{a}(n+i)$ for $-N_f \leq i < 0$ and $N_e \leq i \leq N_e + N_f - 1$.

Here N_f is called the "error-free interval" and $N_f \geq N_g - 1$, where N_g is the target length. It should be noted that in the absence of error propagation, only the bits from the $(k+1)$th to the $(k+N_{sr})$th track affect the performance of the $(k+1)$th sub-2D VD, and the elements of $\mathbf{e}(n)$ that correspond to other tracks should be zero. Let $W(\underline{\mathbf{e}})$ be the number of error bits in the corresponding output track(s) of a sub-2D VD for the error event $\underline{\mathbf{e}}$. Then, the BER of the kth sub-2D VD is given by

$$\mathrm{BER}^k = \sum_{\underline{\mathbf{e}} \in E_s} \sum_{\underline{\mathbf{a}} \in S_e} W(\underline{\mathbf{e}}) \Pr(\underline{\mathbf{a}}) \Pr(error|\underline{\mathbf{a}}) \tag{5.2}$$

where E_s is the set of all possible error events, S_e is the set of all possible data patterns that support the error event $\underline{\mathbf{e}}$, $\Pr(\underline{\mathbf{a}})$ is the probability that the true recorded data is $\underline{\mathbf{a}}$, and $\Pr(error|\underline{\mathbf{a}})$ is the conditional probability that the sub-2D VD detects $\underline{\hat{\mathbf{a}}}$ as the recorded data when the true recorded data is $\underline{\mathbf{a}}$.

Let $a_j(n)$ and $(\mathbf{A}*\mathbf{B})_{n,j}$ denote the jth element of vector $\mathbf{a}_n$ and $(\mathbf{A}*\mathbf{B})_n$, respectively, and '$*$' refers to 2D convolution given by $(\mathbf{A}*\mathbf{B})_k = \sum_n \mathbf{A}_n \mathbf{B}_{k-n}$ unless otherwise specified. According to the principle of VD, $\Pr(error|\underline{\mathbf{a}})$ can be upper-bounded by the probability that the metric corresponding to path $\underline{\hat{\mathbf{a}}}$ is smaller than that corresponding to path $\underline{\mathbf{a}}$. That is

$$\sum_{i=0}^{N_e+N_g-2} \sum_{j=1+k}^{N_{sr}+k} [x_j(n+i) - (\mathbf{G} * \hat{\mathbf{a}})_{n+i,j}]^2$$

$$< \sum_{i=0}^{N_e+N_g-2} \sum_{j=1+k}^{N_{sr}+k} [x_j(n+i) - (\mathbf{G} * \mathbf{a})_{n+i,j}]^2 \tag{5.3}$$

where $k+1$ is the index of the sub-2D VD, and $x_j(n)$ is the equalizer output in the jth track

$$x_j(n) = \left[\sum_{i=0}^{N_w-1} \mathbf{W}_i \mathbf{z}(n-i) \right]_j = \Big\{ [(\mathbf{G} + \mathbf{M}) * \mathbf{a}]_n + (\mathbf{W} * \boldsymbol{\theta})_n \Big\}_j \tag{5.4}$$

where $\mathbf{M}_0$, $\mathbf{M}_1$, $\cdots$, $\mathbf{M}_{N_h+N_w-2}$ represent the matrices of the 2D residual ISI channel, *i.e.* $\mathbf{M}_n = (\mathbf{W} * \mathbf{H})_n - \mathbf{G}_n$, and $\boldsymbol{\theta}(n)$ is a 2D white Gaussian noise vector with variance σ^2. Using (5.4), we can express (5.3) as

$$d^2(\underline{\mathbf{e}}) + \sum_{i=0}^{N_e+N_g-2} \sum_{j=1+k}^{N_{sr}+k} \tilde{e}_j(n+i)[(\mathbf{M} * \mathbf{a})_{n+i,j} + (\mathbf{W} * \boldsymbol{\theta})_{n+i,j}] < 0 \tag{5.5}$$

where $d^2(\underline{\mathbf{e}}) = \sum_{i=0}^{N_e+N_g-2} \sum_{j=1+k}^{N_{sr}+k} (\mathbf{G} * \mathbf{e})^2_{n+i,j}$ and $\tilde{\mathbf{e}}(n) = (\mathbf{G} * \mathbf{e})_n$. For given $\underline{\mathbf{a}}$ and $\underline{\mathbf{e}}$, the left hand side of the inequality in (5.5) is a Gaussian random variable with mean

$$m_u = d^2(\underline{\mathbf{e}}) + m_{me} \tag{5.6}$$

where

$$m_{me} = \sum_{i=-(N_h+N_w-2)}^{N_e+N_g-2} \sum_{j=1+k}^{N_{sr}+k} \left(\mathbf{M}_-^T * \tilde{\mathbf{e}}\right)_{n+i,j} a_j(n+i) \tag{5.7}$$

and variance

$$\sigma_u^2 = \sigma^2 \sum_{i=-(N_w-1)}^{N_e+N_g-2} \sum_{j=1+k}^{N_{sr}+k} \left(\mathbf{W}_-^T * \tilde{\mathbf{e}}\right)^2_{n+i,j}. \tag{5.8}$$

Thus, we get

$$\Pr(error|\underline{\mathbf{a}}) \leq Q\left(\frac{m_u}{\sigma_u}\right) \tag{5.9}$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-x^2/2} dx$ is the tail probability of Gaussian distribution. By substituting (5.9) into (5.2), we get the BER of the $(k+1)$th sub-2D VD as

$$\mathrm{BER}^{k+1} \leq \sum_{\underline{\mathbf{e}} \in E_s} \sum_{\underline{\mathbf{a}} \in S_e} W(\underline{\mathbf{e}}) \Pr(\underline{\mathbf{a}}) Q\left(\frac{m_u}{\sigma_u}\right). \tag{5.10}$$

Based on the BER expression for the sub-2D VD given above, the overall BER of the FDTS/DF-VD in the absence of the error propagation is upper bounded by the summation of the BER of all the sub-2D VDs. In Figure 5.3, the analysis performance for FDTS/DF-VD with no error propagation is realized by the use of the above upper bound. Also, we only use limited number of the channel input data patterns to realize the theoretical performance analysis for the FDTS/DF-VD with no error propagation (see Subsection 3.3.1 for details). As illustrated, the theoretical BER performance is asymptotic to the simulation BER performance, thus indicating that the theoretical approach described above is suitable to analyze the FDTS/DF-VD without error propagation. Further, Figure 5.3 shows that the error propagation degrades performance by 1 dB for BER is 10^{-4}. Compared with Fig. 4.7, it seems that, unlike the energy reduction effect, the error propagation effect is independent of N_{sr}.

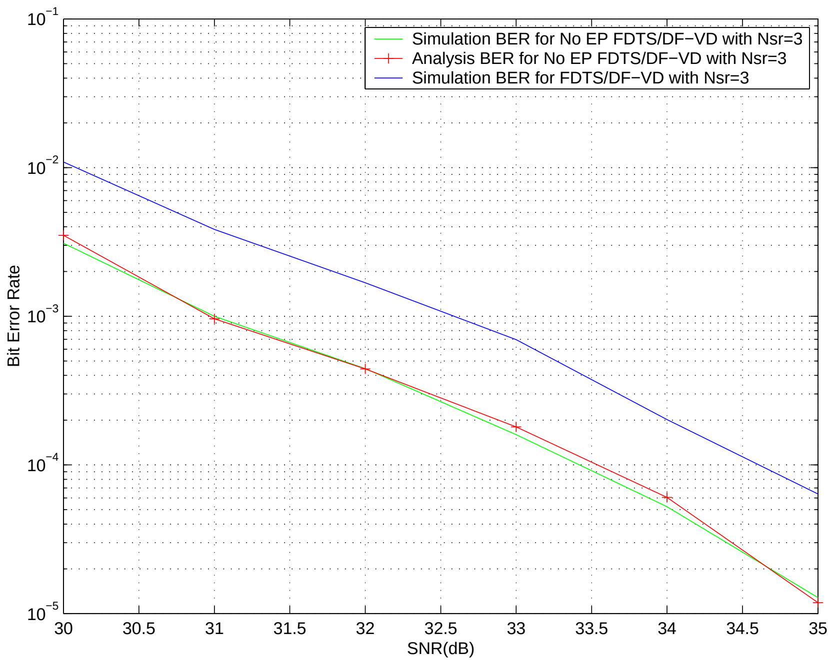

Figure 5.3: Theoretical and simulation BER performance for FDTS/DF-VD.

5.4 Reduced-Complexity FDTS/DF-VD

For the FDTS/DF-VD, adjacent two sub-2D VDs have $N_{sr} - 1$ tracks in common. However, it is hard to tell which sub-2D VD is more reliable in detecting the bits of these

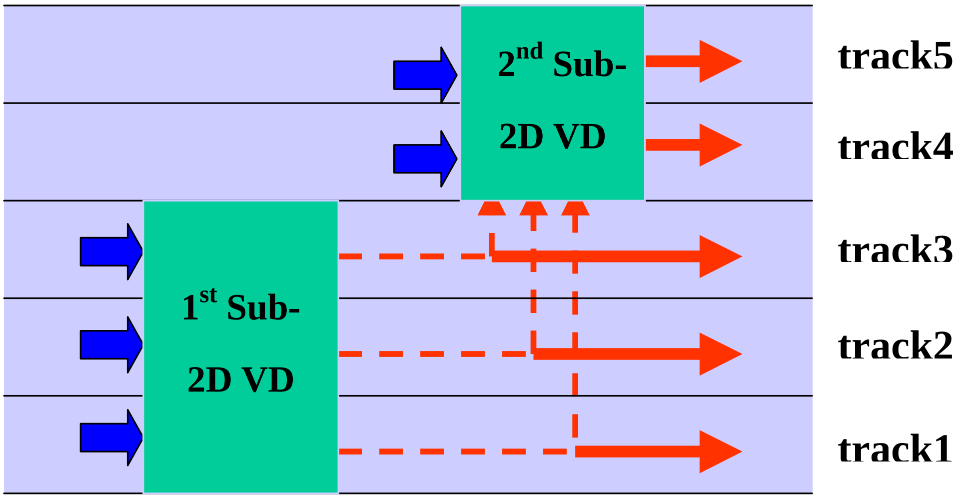

Figure 5.4: Principle of reduced-complexity FDTS/DF-VD with $N_{sr} = 3$ and $N_r = 5$. The solid lines represent the input and output of sub-2D VDs, the dashed lines represent the feedback coming from the output of the previous sub-2D VDs. The last sub-2D VD only deals with two tracks.

tracks. For this reason and in order to further reduce the complexity, we make each track correspond to one unique sub-2D VD. More specifically, the first sub-2D VD deals with the lowest N_{sr} tracks and detected bits from these tracks are all final decisions, the second sub-2D VD deals with the subgroup that is shifted up by N_{sr} tracks and detects all the bits in this subgroup. Same procedures are applied to the third sub-2D VD and above. Note that the last sub-2D VD only deals with N_{msr} (N_{msr}=N_r modulo N_{sr}) tracks, as illustrated in Fig. 5.4. We call this detector the reduced-complexity FDTS/DF-VD (RFDTS/DF-VD) since it uses less number of sub-2D VDs. From Fig. 5.2, we find that for the same N_{sr}, RFDTS/DF-VD and FDTS/DF-VD perform almost comparably. Note that here we consider the TwoDOS case where the largest allowable ISI and ITI occur. In other TwoDOS cases where ISI and ITI are smaller, the two detectors would perform more similarly. In view that RFDTS/DF-VD can further reduce the complexity compared to FDTS/DF-VD, it is an attractive reduced-complexity detector for all the TwoDOS cases when $N_r = 5$.

Table 5.1: Complexity of different 2D detectors.

2D Detectors	Complexity (General)	Complexity (Example)
Full-fledged 2D VD	$O(2^{N_r(N_g-1)})$	$O(2^{10})$
FDTS/DF-VD	$(N_r - N_{sr} + 1)O(2^{N_{sr}(N_g-1)})$	$3O(2^6)$
RFDTS/DF-VD	$\lceil N_r/N_{sr} \rceil O(2^{N_{sr}(N_g-1)})$	$2O(2^6)$
Quasi-1D VD	$N_r O(2^{(N_g-1)})$	$5O(2^2)$

Table 5.1 lists the complexity of different 2D detectors. In the column 'Complexity (Example)', we set $N_r = 5$, $N_{sr} = 3$, and $N_g = 3$ to compute the corresponding complexity for different detectors. This setting is typically used in our simulation. The function $\lceil \alpha \rceil$ in the table is the ceiling function that returns the smallest integer not less than α. From the table, it is clear that all the detectors developed in this and previous chapter (*i.e.* FDTS/DF-VD, RFDTS/DF-VD, and quasi-1D VD) require much less computational complexity compared with the conventional full-fledged 2D VD.

5.5 Target Design for FDTS/DF-VD

As discussed in Section 5.2, better performance of the FDTS/DF-VD can be achieved by increasing N_{sr}. However, the computational complexity of the FDTS/DF-VD is dramatically increased with the increase of N_{sr} since the complexity of the detector grows exponentially with N_{sr}. For this reason, the practical FDTS/DF-VD should constrain N_{sr} to a certain value (*e.g.* three). But this constraint may sometimes not meet the stringent performance requirement[1]. Recalling that a suitable target can improve the detection performance without leading to the increase of detection complexity, in this section, we propose several new targets that improve over conventional causal ITI target in terms of detection performance or target design complexity. It should be noted that these targets are specific for the FDTS/DF-VD and RFDTS/DF-VD, not for the full-

[1]Even though N_{sr} can increase with the advances of circuits and processors, maintaining the stringent performance requirement will become harder with the increase of storage capacity. Therefore, the continuous demand for increased storage capacity will always impose a constraint on N_{sr}.

fledged 2D VD. Consequently, these targets may even degrade the performance of the full-fledged 2D VD.

5.5.1 Truncated Causal ITI Target

For a conventional causal ITI target, all of its 2D target matrices are the right triangular matrices and the diagonal elements of G_0 are constrained to be 1s. This target causes a given track to suffer the ITI only from the tracks below it. However, not all tracks below contributes equal interference to the current track. For the TwoDOS case where the largest ISI and ITI occur, we still found that only the two nearest tracks below the track cause significant ITI. Therefore, in the target matrices, we only take the elements that correspond to the two nearest lower tracks into account. All the remaining causal ITI terms are constrained to be zero. For example, if the target length N_g is three and number of tracks per group N_r is five, the target matrices can be written as,

$$\mathbf{G}_0 = \begin{bmatrix} 1 & \times & \times & 0 & 0 \\ 0 & 1 & \times & \times & 0 \\ 0 & 0 & 1 & \times & \times \\ 0 & 0 & 0 & 1 & \times \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5.11}$$

$$\mathbf{G}_1 = \begin{bmatrix} \times & \times & \times & 0 & 0 \\ 0 & \times & \times & \times & 0 \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \end{bmatrix}, \tag{5.12}$$

$$\mathbf{G}_2 = \begin{bmatrix} \times & \times & \times & 0 & 0 \\ 0 & \times & \times & \times & 0 \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \end{bmatrix} \tag{5.13}$$

where '×' represents the elements that need to be optimized under a given criterion (*e.g.* minimum mean square error). We call this target truncated causal ITI target since it is essentially a causal ITI target with truncated channel memory across the track. This truncation leads to performance degradation of the full-fledge 2D VD. Nevertheless, it makes the FDTS/DF-VD suffer less error propagation effect since it prevents the error propagation from the lower tracks whose separation between the current track is beyond two tracks. In a word, this target benefits FDTS/DF-VD by reducing error propagation effect at the price of introducing the performance loss due to the truncation.

5.5.2 Symmetric Truncated Causal ITI Target

As shown in Section 3.4, a target with the symmetry constraint results in trivial performance loss. Thereupon, we conduct an investigation on the symmetric truncated causal ITI target, *i.e.* we assume that ITI is track independent. For example, if the target length N_g is three and number of tracks per group N_r is five, the target matrices can be written as,

$$\mathbf{G}_0 = \begin{bmatrix} 1 & g_0 & g_1 & 0 & 0 \\ 0 & 1 & g_0 & g_1 & 0 \\ 0 & 0 & 1 & g_0 & g_1 \\ 0 & 0 & 0 & 1 & g_0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5.14}$$

$$\mathbf{G}_1 = \begin{bmatrix} g_2 & g_3 & g_4 & 0 & 0 \\ 0 & g_2 & g_3 & g_4 & 0 \\ 0 & 0 & g_2 & g_3 & g_4 \\ 0 & 0 & 0 & g_2 & g_3 \\ 0 & 0 & 0 & 0 & g_2 \end{bmatrix}, \tag{5.15}$$

$$\mathbf{G}_2 = \begin{bmatrix} g_5 & g_6 & g_7 & 0 & 0 \\ 0 & g_5 & g_6 & g_7 & 0 \\ 0 & 0 & g_5 & g_6 & g_7 \\ 0 & 0 & 0 & g_5 & g_6 \\ 0 & 0 & 0 & 0 & g_5 \end{bmatrix}. \tag{5.16}$$

In this case, the 1D vector target $\mathbf{g} = [g_0, g_1, \cdots, g_7]^T$, where g_i $(i = 0, 1, \cdots, 7)$ are the distinct non-zero elements in the target matrices. This target has two advantages: First, it reduces the complexity of target design process shown in Section 3.3 since the target vector has only eight elements while the traditional causal ITI target has seventy-eight elements when the number of tracks per group $N_r = 5$ and target length $N_g = 3$. Second, it ensures that all the sub-2D VDs have equal noiseless output for a given state. Therefore, all the sub-2D VDs can share the same architecture to yield the noiseless output and this can be realized by a simple table look-up operation, *e.g.* by means of a random-access memory. However, since the channel response is not symmetric for different tracks, this predefined "symmetric" target additionally burdens the finite length equalizer and leads to detection performance degradation.

5.5.3 Simulation Results

As illustrated in Figure 5.5, both new targets degrade the performance of the 2D VD because of the truncation loss, and symmetric truncated causal ITI target results in worst performance for the 2D VD since its symmetry property additionally burdens the finite-length equalizer. Figure 5.5 also shows that these two new targets do not lead to better performance of the FDTS/DF than the conventional causal ITI target. The main reason is that the performance loss due to truncation counteracts the gains of error propagation compensations. Thus, these two targets can not manifest its superiority in reducing the error propagation in the simulations. Nevertheless, both targets can result in reduction in the latency due to the error propagation. Furthermore, they lead to lower complexity in the implementation of FDTS/DF-VD as well as the process of target design. In view that the symmetric truncated causal ITI target results in the lowest

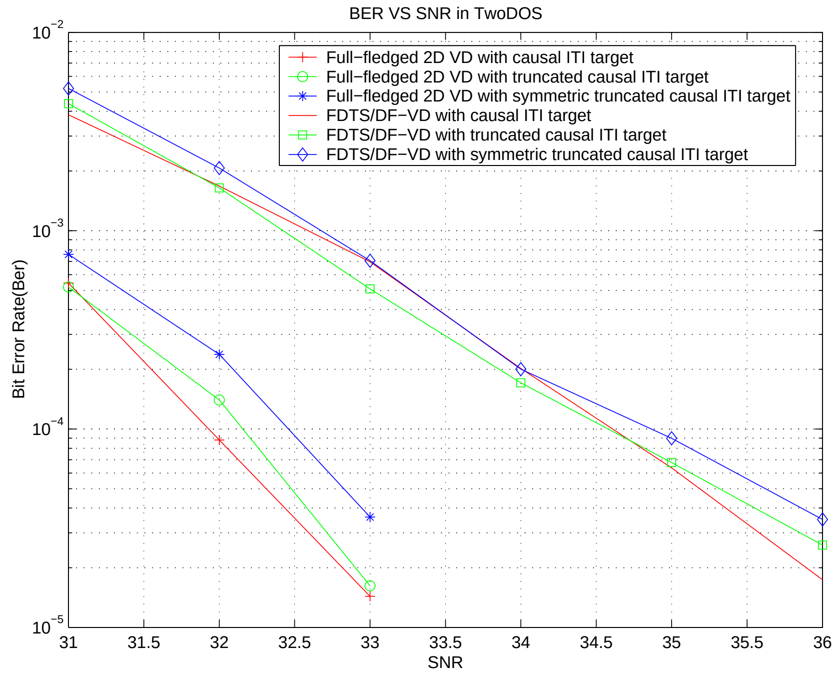

Figure 5.5: BER performance for 2D VD and FDTS/DF-VD with different targets.

complexity while still maintaining good detection performance, we may choose it for the FDTS/DF. Therefore, by judiciously choosing the target and number of tracks under consideration in the FDTS/DF-VD, we can develop a reduced-complexity 2D Viterbi-like detector that facilitates the high-speed TwoDOS implementation without paying a large penalty in detection performance.

5.6 Conclusions

In this chapter, we have proposed a generalized 2D VD called FDTS-DF/VD. The conventional full-fledged 2D VD, QR detector, and our proposed quasi-1D VD can all seen be viewed as special cases of this detector. The detector breaks up the full-fledged 2D VD into a set of sub-2D VDs, each of which deals with a subgroup of tracks (with number N_{sr}). By constraining the target to have causal ITI, all the ITI outside the current subgroup can be estimated and removed by the detected bits from the lower tracks. The flexibility introduced by the parameter N_{sr} allows us to design detectors with varying performance and complexity. We have provided a theoretical platform to evaluate the performance of FDTS-DF/VD that is free of error propagation. Additionally, we have introduced the RFDTS-DF/VD to further reduce the complexity of the FDTS-DF/VD. We have also presented several other novel targets specifically for the FDTS-DF/VD and RFDTS-DF/VD. Our simulation results indicate that by judiciously choosing the target and number of tracks under consideration in the FDTS/DF-VD, we can develop a reduced-complexity 2D Viterbi-like detector that facilitates the high-speed TwoDOS implementation while incurring only a minor penalty in detection performance.

Part II

MIMO-OFDM Systems

In Part I of the thesis, we have concentrated on developing reduced-complexity and high-performance detectors for multiple-input multiple-output systems and have shown that they are effective in the two-dimensional optical storage (TwoDOS) system. In Part I, we have assumed that the knowledge of channel coefficients is available to the receiver. This assumption is realistic to optical storage systems such as TwoDOS but may not hold for wireless communication systems because a system placed in different environments may experience completely different fading behaviors. Therefore, channel estimation is quite important in multiple-input multiple-output wireless communication systems and is the focus of Part II of this thesis.

In the field of wireless communication, multiple-input multiple-output technology is usually referred to as MIMO technology, which has emerged as one of the most significant technical breakthroughs in modern communications after the theoretical work by Foschini [68] and Teletar [175]. Generally, MIMO technology can be used to achieve two objectives: spatial diversity and space-division multiplexing. Spatial diversity, which provides the receiver with several (ideally independent) replicas of the transmitted signals, may rely on the use of space-time coding [13,176,177] to improve the link reliability by transmitting different representations of the same data stream on different transmit antennas. Space-division multiplexing, by comparison, achieves a higher data rate by transmitting independent data streams on different transmit antennas simultaneously [68]. It is believed that MIMO technology is the one of the most likely technologies to achieve data rates in excess of 1 Gb/s [106].

Orthogonal frequency division multiplexing (OFDM) is an attractive multicarrier technology for data transmission over frequency-selective fading channels [190]. This technology was first introduced in 1966 [32], and has been used in various wireless communication systems such as the digital audio broadcasting (DAB) [2, 191], digital video broadcasting (DVB) [1,155,160], and wireless local area networks (WLAN) [3,6]. OFDM technology divides a frequency-selective (wideband) channel into a number of flat fading (narrowband) subchannels, each of which having a distinct subcarrier. This renders the equalizer design particularly simple since only a constant has to be inverted for each flat fading subchannel. In addition, the use of overlapping subcarriers allows the realization of high spectral efficiency.

The complementary benefits of MIMO and OFDM technologies motivate the investigation of systems that combine both technologies [12, 163, 173]. Such systems are referred to as MIMO-OFDM systems and they serve as a potential candidate for the next generation wireless communication systems operating in frequency-selective fading channels [100,173,198]. One particular application of MIMO-OFDM systems is the next generation wireless local area network (WLAN) system. This system is mainly deployed in richly scattered multipath indoor environments, which are ideal for achieving high capacities [68,73]. Aiming to define a standard for MIMO-OFDM based WLAN systems named IEEE 802.11n, the IEEE set up a new task group in January 2004. Till now, several proposals that support data rates up to 108 Mb/s have been submitted for the IEEE802.11n standard. Further, a data rate of 540 Mb/s is envisaged as an optional part of the IEEE 802.11n standard by 2007 [105].

Channel estimation for MIMO-OFDM systems is particularly challenging because of the large number of channel coefficients to be estimated. Early MIMO-OFDM channel estimation techniques treat channels as spatially uncorrelated. However, in many situations, MIMO-OFDM channels tend to be spatially correlated, for example, due to limited scattering. Prior knowledge of this channel spatial correlation and the channel frequency correlation can be exploited by using the linear minimum mean square error (LMMSE) technique. However, the complexity of the full-fledged LMMSE technique, which utilizes both the channel spatial and frequency correlation, is quite high. Further, prior knowledge of channel spatial and frequency correlation is not always available to the receiver. The least squares (LS) technique circumvents these problems but gives much poorer performance. Therefore, it is important to develop reduced-complexity, suboptimal, approximate LMMSE channel estimation techniques that allow a good trade-off between the performance, complexity, and availability of channel stochastic information (e.g. channel correlation or power) for MIMO-OFDM systems. This is the focus of Part II of the thesis.

The organization of Part II is as follows. Chapter 6 provides a overview of channel estimation techniques for conventional OFDM systems with single transmit and receive antennas. This chapter serves as the technical basis on channel estimation techniques for MIMO-OFDM systems. Then, Chapter 7 describes MIMO-OFDM systems in the angle domain, where the channel model lends itself to a physical interpretation. This physically oriented angle-domain channel representation inspires our work for MIMO-OFDM channel estimation techniques. Chapter 8 proposes several channel instantaneous power based angle-domain channel estimation techniques for the case where the channel stochastic information is not available to the receiver. Finally, Chapter 9 develops LMMSE-based angle-domain channel estimation techniques for the case where the channel stochastic information is known to the receiver.

Chapter 6

Channel Estimation for OFDM Systems

6.1 Introduction

As coherent demodulation that requires and utilizes the knowledge of channel coefficients can achieve a 3 dB performance gain compared with differential demodulation [154], it is quite commonly adopted in orthogonal frequency division multiplexing (OFDM) systems. Therefore, accurate and robust channel estimation that permits the realization of coherent demodulation is very important in order to ensure reliable data recovery. Channel estimation can be realized either by pilot-aided or blind techniques. Pilot-aided channel estimation techniques [48] utilize the received signals and specific known transmitted signals, while blind channel estimation techniques [53, 83] use the received signals and the stochastic information (*e.g.* second order statistics) of transmitted and received signals, to estimate the channel coefficients. Compared with pilot-aided techniques, blind techniques save on the use of pilots and can thus increase the spectral efficiency. However, blind techniques require the prior knowledge of stochastic information of the transmitted and received signals. Further, they always result in poorer performance compared with pilot-aided techniques. Therefore, although blind techniques achieve spectral efficiency, in our work, we still focus on pilot-aided techniques given additionally the fact that preamble-based OFDM systems have been widely used in many applications and standards. This chapter provides the technical background of pilot-aided channel estimation techniques for single-input single-output (SISO)-OFDM systems and serves as the basis for the channel estimation techniques for multiple-input multiple-output (MIMO)-OFDM systems.

This chapter is organized as follows. Section 6.2 briefly illustrates OFDM systems. Then, Section 6.3 introduces typical pilot arrangements in OFDM systems. Section 6.4 describes conventional pilot-aided channel estimation techniques for OFDM systems. Finally, Section 6.5 concludes this chapter.

6.2 OFDM Systems

The basic idea of OFDM technology is to use a number of carriers (*i.e.* N_d) to divide the frequency-selective fading channels into N_d parallel and ideally independent flat-fading subchannels. These carriers are referred to as subcarriers and made orthogonal to each other by appropriately choosing the frequency spacing between them. Thus, OFDM technology is a kind of multicarrier technology. OFDM technology has several appealing advantages. First, it can easily handle intersymbol interference (ISI) in frequency-selective fading channels. In such channels, the single carrier technology has to use a multiple-tapped time-domain equalizer [119, 154] to suppress ISI. OFDM technology can avoid this by only using a simple one-tap frequency-domain equalizer for each subcarrier. Second, OFDM technology is robust to impulsive interferences in the time domain. This is because the energy of short time-domain interferences will spread over the entire frequency bandwidth and thus detrimental effects resulting from impulsive interferences are significantly reduced for each subcarrier. In addition, OFDM technology achieves higher spectral efficiency compared to the conventional frequency division multiplexing technique, which disallows the spectral overlap between the subcarriers.

In a typical OFDM system, the high rate signals to be transmitted are first grouped into blocks of N_d data signals at the transmitter. These blocks are called frequency-

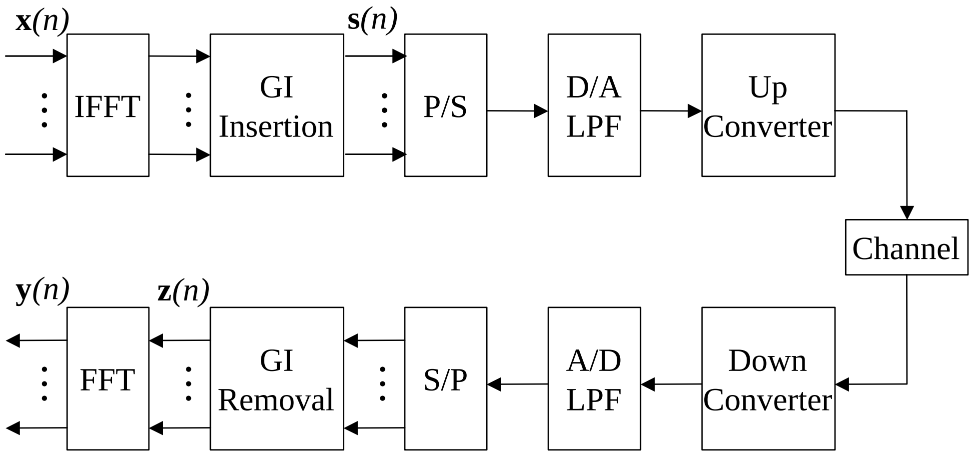

Figure 6.1: Block diagram of a typical OFDM channel.

domain OFDM symbols and the nth group is represented by the vector $\mathbf{x}(n) = [x_0(n), \cdots, x_k(n), \cdots, x_{N_d-1}(n)]^T$ as shown in Fig. 6.1, where k and n denote the indices of the subcarrier and OFDM symbol, respectively. Next, an inverse discrete Fourier transform (IDFT) is applied to each OFDM symbol. The IDFT at the transmitter and discrete Fourier transform (DFT) at the receiver, respectively, serve to modulate and demodulate the data on the orthogonal subcarriers. At the IDFT output (*i.e.* the time domain), a CP of length N_g that is a copy of the last part of the symbol, is inserted at the beginning of each symbol to prevent ISI and its length is assumed to be not shorter than the channel length. The resulting nth time-domain OFDM symbol is represented by $\mathbf{s}(n) = [s_0(n), s_1(n), \cdots, s_{N_d+N_g-1}(n)]^T$, where

$$s_m(n) = \frac{1}{N_d} \sum_{k=0}^{N_d-1} x_k(n) e^{j2\pi k(m-N_g)/N_d} \tag{6.1}$$

for $m = 0, 1, \cdots, N_d + N_g - 1$.

The samples $\{s_m(n)\}$ are sent through a frequency-selective fading channel, which can be represented by an equivalent discrete-time linear finite duration channel impulse response (CIR) given by $\mathbf{c}(n) = [c_0(n), c_1(n), \cdots, c_{N_h-1}(n)]^T$, where N_h is the length of the time-domain channel. By performing an N_d-point DFT on the received OFDM symbols after removing the CP from each OFDM symbol at the receiver, we get the corresponding frequency-domain symbols. The resulting frequency-domain data sample at the kth subcarrier in the nth OFDM symbol is given by

$$y_k(n) = h_k(n)x_k(n) + \vartheta_k(n) \tag{6.2}$$

where $h_k(n)$ is the kth sample of the N_d-point DFT of $\mathbf{c}(n)$, and the noise $\vartheta_k(n)$ is assumed to be additive white Gaussian (AWGN). In a compact notation, the received block of consecutive N_c OFDM symbols in the frequency domain can be written as

$$\underline{\mathbf{y}} = \underline{\mathbf{X}}\,\mathbf{h} + \underline{\boldsymbol{\vartheta}} \tag{6.3}$$

where $\underline{\mathbf{y}} = [\mathbf{y}(n), \mathbf{y}(n+1), \cdots, \mathbf{y}(n+N_c-1)]^T$ with $\mathbf{y}(n) = [y_0(n), y_1(n), \cdots, y_{N_d-1}(n)]^T$, $\underline{\boldsymbol{\vartheta}} = [\boldsymbol{\vartheta}(n), \boldsymbol{\vartheta}(n+1), \cdots, \boldsymbol{\vartheta}(n+N_c-1)]^T$ with $\boldsymbol{\vartheta}(n) = [\vartheta_0(n), \vartheta_1(n), \cdots, \vartheta_{N_d-1}(n)]^T$, $\mathbf{h} = [h_0, h_1, \cdots, h_{N_d-1}]^T$ is the channel transfer function (CTF), and $\underline{\mathbf{X}} = [\mathbf{X}(n), \mathbf{X}(n+1), \cdots, \mathbf{X}(n+N_c-1)]^T$ with $\mathbf{X}(n) = \mathrm{diag}[x_0(n), x_1(n), \cdots, x_{N_d-1}(n)]$. The index n is omitted from the CTF $\mathbf{h}$ because we assume that the channel remains time-invariant over N_c symbols. This assumption is satisfied, for example, in IEEE 802.11a systems for N_c=2.

6.3 Pilot Arrangements in OFDM systems

In the literature, there exist two types of frequency-domain pilot arrangements in OFDM systems for the implementation of pilot-aided channel estimation techniques: block-type and comb-type pilot arrangements. The first type (see Fig. 6.2) is realized by inserting pilots into all of the subcarriers within a periodical time interval and the estimated channel coefficients obtained from one period are used for further signal processing (*e.g.* equalization, detection *etc.*) until the next period of pilots is received. This type of pilot arrangement is especially suitable for slowly fading channels. The second type (see Fig. 6.3) is performed by inserting pilots into a certain number of subcarriers of each OFDM symbol and use the estimated channel coefficients at these pilot subcarriers to interpolate channel coefficients at the remaining subcarriers (*i.e.* data subcarriers). Compared with the first type, this type is more suitable for rapidly time-varying channels

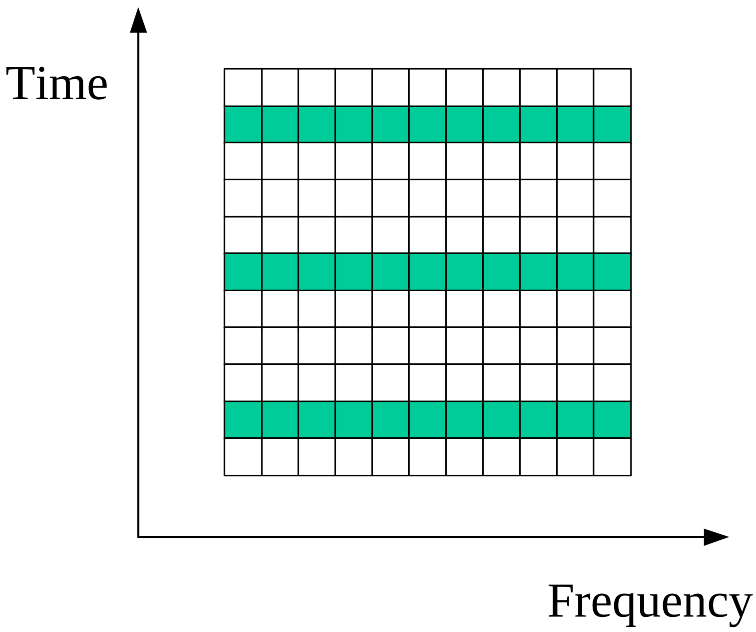

Figure 6.2: Block-type pilot arrangement. The solid and hollow squares represent the pilot symbols and data symbols, respectively.

in that it can track time variation within each OFDM symbol. However, it may suffer an irreducible estimation error floor due to the interpolation.

Interpolation techniques for the comb-type pilot arrangements can be realized in the following ways:

- **Piecewise-Constant Interpolation:**

  The piecewise-constant interpolation technique is the simplest technique in which the estimated channel coefficient at a single pilot subcarrier f_p is used for all the data subcarriers that are near the pilot subcarrier f_p. This technique uses the assumption that the fading on pilot subcarriers and data subcarriers are totally correlated [42]. This assumption requires the separation between pilot subcarriers and data subcarriers to be sufficiently small to maintain high correlation. Consequently, a relatively large number of pilot subcarriers is required for reliable estimation, but this requirement is undesirable to achieve a high spectral efficiency.

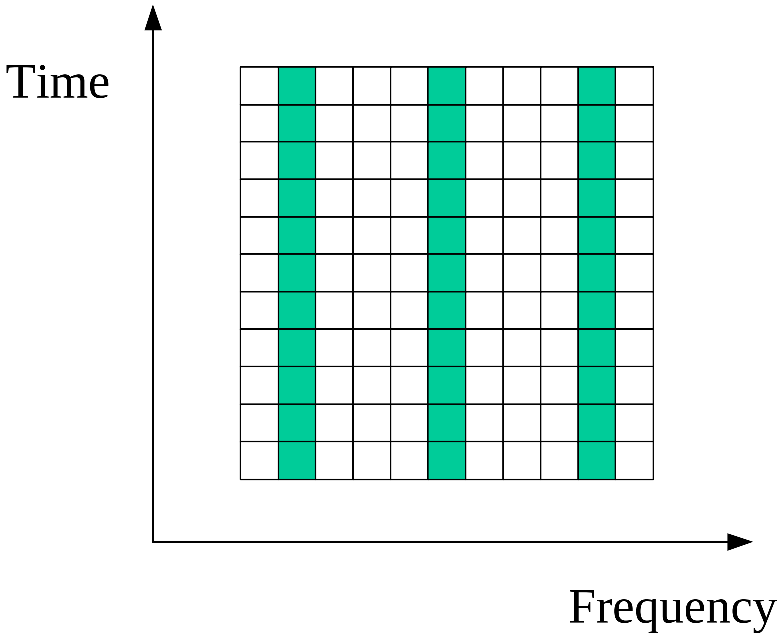

Figure 6.3: Comb-type pilot arrangement. The solid and hollow squares represent the pilot symbols and data symbols, respectively.

- **Piecewise-Linear Interpolation:**

  The piecewise-linear interpolation technique uses the estimated channel coefficients at two consecutive pilot subcarriers to determine channel coefficients at data subcarriers that are between the two pilot subcarriers. Let f_p and f_{p+1} denote the index of two pilot subcarriers, respectively, the estimated channel coefficient $\tilde{h}_k$ at the kth subcarrier that is between the adjacent two pilots subcarriers is given by

  $$\tilde{h}_k = \left(\frac{\tilde{h}_{f_{p+1}} - \tilde{h}_{f_p}}{f_{p+1} - f_p} \right) (k - f_p) + h_{f_p} \tag{6.4}$$

  where $\tilde{h}_{f_p}$ and $\tilde{h}_{f_{p+1}}$ are the estimated channel coefficients at two pilot subcarriers, respectively.

  Compared to the piecewise-constant interpolation technique, this technique looses the stringent requirement that pilot subcarriers and data subcarriers should be totally correlated and thus the separation between pilot subcarriers and data subcarriers can be made larger. Consequently, fewer pilots are required and the spectral efficiency can be increased. Further, this technique may yield considerable performance improvement because more estimated channel coefficients at pilot subcarriers are utilized during the interpolation [158]. However, the first-order interpolation does not always reflect the actual channel characteristics since most of the practical channel transfer functions between adjacent pilot subcarriers are not line segments and thus will result in an irreducible estimation error floor.

- **Second-Order Interpolation:**

  The second-order interpolation technique uses the estimated channel coefficients at three consecutive pilot subcarriers to determine channel coefficients at data subcarriers that are between the pilot subcarriers. Let f_{p-1},f_p and f_{p+1} denote the index of these three pilot subcarriers, respectively. Then, the estimated channel coefficient $\tilde{h}_k$ at the kth subcarrier that is between the f_p^{th} and f_{p+1}^{th} pilots subcarriers is given by [92]

$$\tilde{h}_k = \beta_1 \tilde{h}_{f_{p-1}} + \beta_0 \tilde{h}_{f_p} + \beta_{-1} \tilde{h}_{f_{p+1}} \tag{6.5}$$

  where

$$\begin{cases} \beta_1 & = & \alpha(\alpha+1)/2, \\ \beta_0 & = & -(\alpha-1)(\alpha+1), \\ \beta_{-1} & = & \alpha(\alpha-1)/2, \end{cases}$$

  and $\alpha = (k - f_p)/N_d$.

  The second-order interpolation technique usually performs better than the piecewise-linear interpolation technique since the former fits the channel transfer function better and exploits more estimated channel coefficients at pilot subcarriers for interpolation.

- **Transformed-Domain Based Interpolation:**

  The transformed-domain based interpolation technique utilizes the estimated channel coefficients at all the pilot subcarriers for interpolation. Let N_p denote the number of pilots in a single OFDM symbol. Then, an N_p-point DFT is performed to transform the estimated channel coefficients at pilot subcarriers from the frequency domain into one transformed domain. The resulting transformed channel coefficients corresponding to those at pilot subcarriers for the nth OFDM symbol is given by

$$G_m(n) = \sum_{k'=0}^{N_p-1} \tilde{h}_{k'}(n)\ e^{-j2\pi ml/N_p} \tag{6.6}$$

for $m = 0, 1, \cdots, N_p-1$, where m is the index of samples in the transformed domain. After applying a low-pass filter whose cutoff frequency is p_c, the resulting channel coefficients in the transformed domain become

$$\tilde{G}_m(n) = \begin{cases} G_m(n), & \text{if } 0 \leq m \leq p_c \text{ or } N_p - p_c \leq m \leq N_p - 1, \\ 0, & \text{otherwise.} \end{cases} \tag{6.7}$$

After this low-pass filtering operation, only $2(P_c + 1)/N_p$ of the noise remains. Then, padding $N_d - N_p$ zeros at the high-frequency region of the transformed domain [202]:

$$\hat{G}_m(n) = \begin{cases} \tilde{G}_m(n), & \text{if } 0 \leq m \leq p_c, \\ 0, & \text{if } p_c < m < N - p_c, \\ \tilde{G}_{m-N_d+N_p}(n), & \text{if } N_d - p_c \leq m \leq N_d - 1. \end{cases} \tag{6.8}$$

Performing an N_d-point IDFT to the zero-padded channel coefficients, the estimated channel coefficients at all the data and pilot subcarriers are obtained. In fact, if we change the order of DFT and IDFT operations, the transformed domain becomes the time domain. This interpolation technique is also applicable in OFDM systems [48].

Since both the low-pass filtering operation in the transformed domain and the use of all the estimated channel coefficients at pilot subcarriers for interpolation can reduce the noise effect, the transformed-domain based interpolation technique could yield better performance than the above mentioned techniques at the price of higher complexity [202]. However, the performance of this technique is limited at higher SNRs if the low-pass filtering operation discards some channel information. Furthermore, this technique cannot be considered a pure interpolation technique because after the IFFT operation, the estimated channel coefficients at pilot subcarriers may not be the same as those before the DFT operation.

- **Low-Pass Interpolation:**

The low-pass interpolation technique is another technique that utilizes estimated channel coefficients at all the pilot subcarriers. In this technique, zeros are inserted between each pair of adjacent estimated channel coefficients at pilot subcarriers to form the input vector so that the number of input vector reaches N_d, which is the number of the total pilot and data subcarriers. This operation creates a higher-rate signal whose spectrum is the same as the original over the original bandwidth, but has images of the original spectrum centered on multiples of the original sampling rate. Then, a low-pass finite impulse response (FIR) filter is applied to eliminate the images as well as some high-frequency region according to the cutoff frequency, and the ideal values of the interpolated points (*i.e.* the channel coefficients at data subcarriers) are obtained. Finally, the interpolation is performed such that the original estimated channel coefficients at pilot subcarriers remain unchanged and the mean square error (MSE) between the interpolated and ideal values is minimized. This minimum MSE (MMSE) based technique makes the low-pass interpolation technique perform better than the transformed-domain based technique, as verified by simulation in [48].

- **Windowed DFT Based Interpolation:**

The transformed-domain based and low-pass interpolation techniques exploit the fact that the channel power is concentrated on a limited number of channel coefficients in the transformed domain. Then, a low-pass filter (LPF) is used to filter the noise outside these channel coefficients. However, for non-sample-spaced channels, *i.e.* when not all the multipath time delays are located at integer multiples of the sampling time, it is widely known that transforming the channel coefficients from the frequency domain into the transformed domain by the use of DFT or IDFT operations will result in aliasing effects [195]. In other words, the channel power is not concentrated in a limited number of channel coefficients but leaks into all the other channel coefficients. This will cause a portion of the channel power to be lost by the LPF operation and thus result in an irreducible channel estimation error floor. For this reason, a frequency-domain windowing operation can be used to reduce the aliasing effects [195]. As illustrated in Fig. 6.4, a window such as the Hanning window is applied to the estimated channel coefficients at pilot subcarriers in order to keep the channel power concentrated as much as possible in a few number of channel coefficients in the transformed domain. Then, after the N_d-point IDFT, a linear transformation such as Wiener filtering is used to suppress the noise in the transformed domain. Later, zeros are padded similarly as shown in (6.8) so that the number of processed samples increases to N_d. Finally, an N_d-point DFT is used and the estimated frequency-domain channel coefficients are obtained by removing the effects of data windowing. In [195], it was shown that this windowed DFT based interpolation technique can achieve much better performance than the transformed-domain based technique for non-sample-spaced channels.

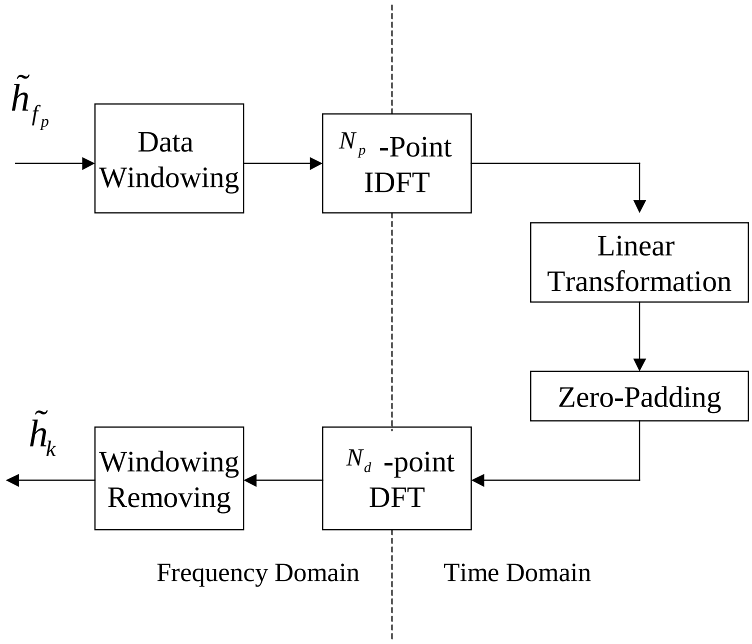

Figure 6.4: Block diagram of the windowed-DFT based interpolation.

- **DCT Based Interpolation:**

The above transformed-domain based interpolation, low-pass interpolation, and windowed DFT based interpolation techniques can all be considered as DFT based interpolation techniques. Recently, a discrete cosine transform (DCT) based interpolation technique is introduced to deal with power leakage in non-sample-spaced channels [196, 197]. This technique is motivated by the fact that the performance of DFT based interpolation techniques will be degraded when there are abrupt variations between the estimated channel coefficients at the first and last pilot subcarriers. This is because the discontinuance in periodical boundaries before the IDFT or DFT operations will result in high-frequency components in the transformed domain. These high-frequency components will generate significant aliasing effects especially in the non-sample-spaced channels. Though this undesirable discontinuous edge effect can be alleviated by the windowed DFT based interpolation technique, the effect cannot be eliminated. To deal with this problem, a DCT based interpolation technique was proposed in [196, 197] by exploiting the fact that the operation of an N_d-point DCT is equivalent to extending the original N_d points to $2N_d$ points by mirror extension, followed by a $2N_d$-point DFT of the extended data with some magnitude and phase compensations. As a result, the waveform of the $2N_d$ points will be smoother and more continuous in the boundary between consecutive periods. Therefore, the DCT based interpolation technique can eliminate the discontinuous edge effect and thus lead to less power leakage in the non-sample-spaced channels compared with the DFT based interpolation techniques.

The DCT based interpolation technique can be summarized as follows. First, the estimated channel coefficients at pilot subcarriers are transformed from the frequency domain to the transformed domain by an N_p-point DCT operation. Note that the DCT operation will introduce an index dependent phase shift in the transformed domain compared with the DFT operation. Then, the linear transformation (*e.g.* LPF or Wiener filter) and zero-padding are applied in the transformed domain. Finally, the N_d-point extendible inverse DCT (EIDCT) is used instead of conventional inverse DCT (IDCT) to compensate for the phase shift in the transformed domain, and the estimated channel coefficients at data subcarriers are thus obtained in the frequency domain [196]. Clearly, the DCT based interpolation technique improves over DFT based interpolation techniques in terms of performance for non-sample-spaced OFDM channels. However, the use of EIDCT results in relatively high complexity. One simplified DCT based interpolation technique is also introduced in [197] by defining a $2N_p$ vector based on the estimated channel coefficients at pilot subcarriers for the compatibility with the fast DCT and IDCT. However, the main disadvantage of this simplified technique is that the estimated channel coefficients at the data subcarriers that are beyond the last pilot subcarrier will go down to zero.

In addition to the 1D pilot arrangements, there also exist 2D pilot arrangements where pilots are placed in both the time and frequency domains [87, 88, 147, 183]. The minimum pilot spacings in time and frequency domains are determined by the Doppler frequency and delay spread, respectively. The 2D pilot arrangements can significantly reduce the number of pilots and thus achieve spectral efficiency. However, they require 2D interpolation techniques, whose complexity is always quite high, and thus have limited application. For example, the 2D pilot arrangements are not typically used in wireless local area networks (WLANs) because of two reasons. First, the interpolation is not required in the time domain because the transmission duration for a packet is short enough such that the channel is usually assumed to be constant during each packet transmission. Second, the interpolation in the time domain makes the channel estimation not real-time, which will ultimately decrease the data rates of WLANs. Therefore, the 2D pilot arrangements will not be considered in this thesis.

6.4 Pilot-Aided Channel Estimation Techniques

The previous section discussed pilot arrangements and reviewed interpolation techniques in comb-type pilot arrangements by assuming that the estimated channel coefficients at pilot subcarriers are already available. In this section, we start to introduce techniques that are used to estimate these channel coefficients at pilot subcarriers. These pilot-aided channel estimation techniques are applicable in both block-type and comb-type pilot arrangements. For the sake of simplicity, we focus on the block-type pilot arrangement where the number of channel coefficients to be estimated is equal to the number of total subcarriers (*i.e.* N_d). The estimation techniques can be easily extended to comb-type pilot arrangements, but this will not be covered in this section.

Generally speaking, pilot-aided channel estimation is based on either the least squares (LS) [19] or the linear minimum mean square error (LMMSE) technique [57,124]. The essential difference between these two types of techniques is that the channel coefficients are treated as deterministic but unknown constants in the former, and as random variables of a stochastic process in the latter. Compared with LS-based techniques, LMMSE-based techniques yield better performance because they additionally exploit and require the prior knowledge of channel correlation. However, the channel correlation is sometimes not *a priori* known, which makes LMMSE-based techniques infeasible.

As discussed in Chapter 1, the pilot-aided channel estimation techniques for OFDM systems can be divided into three categories: frequency-domain techniques, time-domain techniques, and DFT-based techniques. Frequency-domain techniques treat the frequency-domain channel coefficients as the parameters to be estimated. Time-domain techniques treat the time-domain channel coefficients as the parameters to be estimated. The estimated time-domain channel coefficients are then transformed into the frequency-domain ones. DFT-based techniques also treat the frequency-domain channel coefficients as the parameters to be estimated. But unlike the frequency-domain techniques, they transform the estimated channel coefficients from the frequency domain into the time domain, where the noise filtering process is performed, and finally back to the frequency domain by the use of IDFT and DFT operations.

In this section, we treat the LS-based and LMMSE-based channel estimation techniques in different subsections. Within each subsection, we treat frequency-domain and DFT-based techniques separately. The relation of time-domain techniques with frequency-domain and DFT-based techniques will also be treated within each subsection.

6.4.1 LS Estimation Techniques

LS estimation techniques treat channel coefficients as deterministic but unknown constants. For the assumed signals that are determined by the pilots and unknown channel coefficients, and the given signals that are observations of the assumed signals corrupted by the noise and inaccuracies due to the mismatch between the estimated and actual channel coefficients, LS techniques choose coefficients that minimizes the square error between the assumed signals and given signals as the estimated channel coefficients [110].

- **Frequency-Domain LS Channel Estimation**

  Recall from (6.3) that the received block of consecutive N_c OFDM symbols in the frequency domain is given by

$$\underline{\mathbf{y}} = \underline{\mathbf{X}}\,\mathbf{h} + \underline{\boldsymbol{\vartheta}}. \tag{6.9}$$

  Then, according to the principle of LS estimation techniques, the square error $J_{\mathrm{LS}}(\mathbf{h})$ between the assumed signals and given signals is defined as

$$J_{\mathrm{LS}}(\mathbf{h}) = \left(\underline{\mathbf{y}} - \underline{\mathbf{X}}\,\mathbf{h}\right)^H \left(\underline{\mathbf{y}} - \underline{\mathbf{X}}\,\mathbf{h}\right). \tag{6.10}$$

  Taking the partial derivative of $J_{\mathrm{LS}}(\mathbf{h})$ with respect to $\mathbf{h}$ and setting it to zero, the estimated frequency-domain channel coefficients based on this frequency-domain LS channel estimation technique are given by

$$\mathbf{h}_{\mathrm{LS}} = \left(\underline{\mathbf{X}}^H \underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}^H \underline{\mathbf{y}} \tag{6.11}$$

  where the superscripts 'H' and '-1' denote the Hermitian and inverse operation, respectively, and $(\underline{\mathbf{X}}^H \underline{\mathbf{X}})^{-1}\underline{\mathbf{X}}^H$ is the Moore-Penrose inverse of $\underline{\mathbf{X}}$. Further, the MSE between the estimated and actual channel coefficient per subcarrier for a given set of pilots or detected data $\underline{\mathbf{X}}$ is given by

$$\begin{aligned}
\mathrm{MSE}_{\mathrm{LS}|\underline{\mathbf{X}}} &= \frac{1}{N_d} E\left[||\mathbf{h}_{\mathrm{LS}} - \mathbf{h}||^2\right] \\
&= \frac{1}{N_d} E\left\{\left[\left(\underline{\mathbf{X}}^H \underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}^H \underline{\boldsymbol{\vartheta}}\right]^H \left[\left(\underline{\mathbf{X}}^H \underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}^H \underline{\boldsymbol{\vartheta}}\right]\right\} \\
&= \frac{\sigma_f^2}{N_d} \mathrm{trace}\left[\left(\underline{\mathbf{X}}^H \underline{\mathbf{X}}\right)^{-1}\right]
\end{aligned} \tag{6.12}$$

where σ_f^2 is the variance of the AWGN. Furthermore, if we replace $(\underline{\mathbf{X}}^H\underline{\mathbf{X}})^{-1}$ with $E[(\underline{\mathbf{X}}^H\underline{\mathbf{X}})^{-1}]$, the average MSE between the estimated and actual channel coefficient per subcarrier is given by

$$\mathrm{MSE}_{\mathrm{LS}} = \frac{\beta}{N_c\mathrm{SNR}} \tag{6.13}$$

where $\beta = E(|x_k(n)|^2)E(1/|x_k(n)|^2)$, N_c is the number of consecutive OFDM symbols and $\mathrm{SNR} = E(|x_k(n)|^2)/\sigma_f^2$. Since N_c is in the denominator of $\mathrm{MSE}_{\mathrm{LS}}$ as shown in (6.13), it can be easily seen that the performance of the frequency-domain LS channel estimation technique can be improved by using more consecutive OFDM symbols at the price of higher complexity.

The main advantage of the frequency-domain LS channel estimation technique is its low complexity since $\underline{\mathbf{X}}$ in (6.11) is a diagonal matrix such that the inverse $(\underline{\mathbf{X}}^H\underline{\mathbf{X}})^{-1}$ will not result in high complexity. However, its performance is always poorer than the DFT-based LS channel estimation, which will be discussed next.

- **DFT-Based LS Channel Estimation**

Usually, the length of the CIR $\mathbf{c}(n)$ is relatively small compared with the length of an OFDM symbol. The DFT-based LS channel estimation technique utilizes this by assuming that the temporal span of the channel is concentrated over a small number of coefficients, and thus the noise in the coefficients beyond the channel length can be removed in the time domain. This noise removal results in a performance improvement compared to the frequency-domain LS channel estimation technique. Defining the $N_d \times N_d$ unitary FFT matrix as

$$\mathbf{F} = \frac{1}{\sqrt{N_d}} \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & e^{-j2\pi/N_d} & \cdots & e^{-j2\pi(N_d-2)/N_d} & e^{-j2\pi(N_d-1)/N_d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & e^{-j2\pi(N_d-2)/N_d} & \cdots & e^{-j2\pi(N_d-2)(N_d-2)/N_d} & e^{-j2\pi(N_d-2)(N_d-1)/N_d} \\ 1 & e^{-j2\pi(N_d-1)/N_d} & \cdots & e^{-j2\pi(N_d-1)(N_d-2)/N_d} & e^{-j2\pi(N_d-1)(N_d-1)/N_d} \end{bmatrix} \tag{6.14}$$

and the unitary IFFF matrix as $\mathbf{F}^H$ or $\mathbf{F}^{-1}$, the estimated channel coefficients based on the DFT-based LS channel estimation technique are given by [52]

$$\mathbf{h}_{\text{DFT-LS}} = \mathbf{F}_1\mathbf{F}_1^H\mathbf{h}_{\mathrm{LS}} \tag{6.15}$$

where $\mathbf{F}_1$ contains the first $\tilde{N}_h$ columns of the unitary FFT matrix $\mathbf{F}$ and $\tilde{N}_h$ is the estimated channel length. From (6.15), it is clear that the estimated channel coefficients using the frequency-domain LS technique are first transformed from the frequency domain to the time domain via the IDFT operation. The resulting estimated CIR has the length of $\tilde{N}_h$, the remaining $N_d - \tilde{N}_h$ taps of the CIR are set to zero and the noise in these taps is accordingly removed. Then, the DFT operation is performed and the estimated frequency-domain channel coefficients are finally obtained. When all the channel power is maintained during the noise removal process, the average MSE between the estimated and actual channel coefficient per subcarrier is given by

$$\mathrm{MSE}_{\mathrm{LS}} = \frac{\tilde{N}_h \beta}{N_d N_c \mathrm{SNR}}. \tag{6.16}$$

Comparing (6.13) with (6.16), we find that the DFT-based LS channel estimation technique improves over the frequency-domain LS channel estimation technique when all the channel power is maintained during the noise removal process since $\tilde{N}_h$ is always smaller than N_d. The DFT-based LS channel estimation technique is well suited for sample-spaced channels. However, its performance will be deteriorated for non-sample-spaced channels where the channel power will not be concentrated on a small number of coefficients [19, 56]. Further, this technique has higher complexity compared to the frequency-domain LS channel estimation technique due to the use of DFT and IDFT operations.

– *Link with Time-Domain and DFT-Based LS Channel Estimation*

  The time-domain LS technique is commonly referred to as the time-domain maximum likelihood (ML) technique. This technique treats the time-domain channel coefficients as the parameters to be estimated [19, 52, 138]. The estimated time-domain channel coefficients are then transformed to the estimated channel coefficients at pilot subcarriers by the use of DFT operation. Let $\mathbf{F}_{11}$ and $\mathbf{F}_{12}$ contain the first N_p and remaining rows of $\mathbf{P}^H\mathbf{F}_1$. Then, using the time-domain ML technique, the estimated channel coefficients at pilot subcarriers are given by [52]

$$\mathbf{h}_{\mathrm{TML}} = \mathbf{F}_2\mathbf{F}_2^H\mathbf{h}_{\mathrm{LS}} \tag{6.17}$$

where $\mathbf{F}_2$ contains the first N_h columns of the unitary FFT matrix $\mathbf{F}$. Comparing (6.15) with (6.17), it is clearly that the time-domain and DFT-based LS channel estimation techniques are equivalent when $N_h = \tilde{N}_h$. This is expected as both techniques consider the temporal span of the channel to be concentrated over a small number of coefficients such that the noise in the coefficients beyond the channel length can be removed.

6.4.2 LMMSE Estimation Techniques

LMMSE estimation techniques treat channel coefficients as random variables of a stochastic process. Though they are more complex to implement compared with LS estimation techniques and require the prior knowledge of channel correlation, they can achieve better performance [19]. Therefore, LMMSE estimation techniques are used in many systems where performance is a primary concern. For example, LMMSE estimation techniques are quite commonly used when pilots are arranged in a comb-type fashion since the precision of estimated channel coefficients at pilot subcarriers highly determines the performance of channel interpolation at data subcarriers. However, the channel correlation may not always be available to the receiver in practice.

- **Frequency-Domain LMMSE Channel Estimation:**

  Denoting a linear transformation of the received signals at pilot subcarriers as $\mathbf{A}$, the estimated channel coefficients using linear techniques are given by

$$\mathbf{h}_{\mathrm{linear}} = \mathbf{A}\underline{\mathbf{y}}. \tag{6.18}$$

  Based on the principle of LMMSE, we need to find the optimum $\mathbf{A}_o$ such that the MSE

$$J_{\mathrm{linear}}(\mathbf{A}) = E[||\mathbf{h}_{\mathrm{linear}} - \mathbf{h}||^2] \tag{6.19}$$

is minimized. The estimated channel coefficients based on the LMMSE channel estimation technique are then given by

$$\mathbf{h}_{\text{LMMSE}} = \mathbf{A}_o \underline{\mathbf{y}}. \tag{6.20}$$

To find the optimum $\mathbf{A}_o$, substituting (6.18) and (6.9) into (6.19) yields

$$J_{\text{LMMSE}}(\mathbf{A}) = \text{trace}\left[\mathbf{A}\left(\sigma_f^2 \mathbf{I}_{N_d \times N_c} + \underline{\mathbf{X}}\mathbf{R}\underline{\mathbf{X}}^H\right)\mathbf{A}^H + \mathbf{R} - \mathbf{A}\underline{\mathbf{X}}\mathbf{R} - \mathbf{R}\underline{\mathbf{X}}^H\mathbf{A}^H\right] \tag{6.21}$$

where $\mathbf{R} = E[\mathbf{h}\mathbf{h}^H]$ is the channel correlation matrix of $\mathbf{h}$, and $\mathbf{I}_{N_d}$ is an $(N_d \times N_c) \times (N_d \times N_c)$ identity matrix. Differentiating J_{LMMSE} with respect to $\mathbf{A}^H$ and setting the result to zero, we get the optimum $\mathbf{A}_o$ and estimated channel coefficients as

$$\mathbf{A}_o = \mathbf{R}\underline{\mathbf{X}}^H(\sigma_f^2\ \mathbf{I}_{N_d \times N_c} + \underline{\mathbf{X}}\mathbf{R}\underline{\mathbf{X}}^H)^{-1} \tag{6.22}$$

and

$$\mathbf{h}_{\text{LMMSE}} = \mathbf{R}\left[\mathbf{R} + \sigma_f^2\left(\underline{\mathbf{X}}^H\underline{\mathbf{X}}\right)^{-1}\right]^{-1}\mathbf{h}_{\text{LS}}, \tag{6.23}$$

respectively. Equation (6.23) clearly shows that the estimated channel coefficients based on the frequency-domain LMMSE technique can be obtained from estimated channel coefficients based on the frequency-domain LS technique through a linear transformation $\mathbf{B}_t$ given by

$$\mathbf{B}_t = \mathbf{R}\left[\mathbf{R} + \sigma_f^2\left(\underline{\mathbf{X}}^H\underline{\mathbf{X}}\right)^{-1}\right]^{-1}. \tag{6.24}$$

This transformation uses the prior knowledge of $\mathbf{R}$ and σ_f^2, and hence makes the frequency-domain LMMSE technique improve over the frequency-domain LS technique in terms of performance at the cost of higher computational complexity. In addition, the MSE per subcarrier for a given set of pilots or detected data $\mathbf{X}$ is given by

$$\text{MSE}_{\text{LMMSE}|\underline{\mathbf{X}}} = \frac{1}{N_d}\text{trace}\left\{\mathbf{R}\left[\mathbf{R} + \sigma_f^2\left(\underline{\mathbf{X}}^H\underline{\mathbf{X}}\right)^{-1}\right]^{-1}\sigma_f^2\left(\underline{\mathbf{X}}^H\underline{\mathbf{X}}\right)^{-1}\right\}. \tag{6.25}$$

By using the singular value decomposition (SVD) of $\mathbf{R}$, we obtain

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \tag{6.26}$$

where $\mathbf{U}$ is a unitary matrix containing the eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N_d-1}$ on its diagonal. Then, the average MSE per subcarrier is given by

$$\mathrm{MSE}_{\mathrm{LMMSE}} = \frac{\beta}{N_c N_d \mathrm{SNR}} \sum_{k=0}^{N_d-1} \frac{\lambda_k}{\lambda_k + \frac{\beta}{N_c \mathrm{SNR}}}. \tag{6.27}$$

– *Link with Time-Domain and Frequency-Domain LMMSE Channel Estimation*
  As the time-domain and frequency-domain channel correlation is interchangeable in OFDM systems, the time-domain and Frequency-Domain LMMSE technique achieve the same performance. Thus, we will not separately cover the time-domain LMMSE channel estimation technique in this section.

- **DFT-Based LMMSE Channel Estimation**

Similar to the LS channel estimation technique, the estimator channel coefficients based on the DFT-based LMMSE channel estimation technique are given by [56]

$$\mathbf{h}_{\text{DFT-LMMSE}} = \mathbf{F}_1 \mathbf{R}_{\tilde{\mathbf{c}}} \left[\mathbf{R}_{\tilde{\mathbf{c}}} + \sigma_f^2 \left(\mathbf{F}_1^H \underline{\mathbf{X}}^H \underline{\mathbf{X}} \mathbf{F}_1 \right)^{-1} \right]^{-1} \left(\mathbf{F}_1^H \mathbf{F}_1 \right)^{-1} \mathbf{F}_1^H \mathbf{h}_{\mathrm{LS}} \tag{6.28}$$

where $\mathbf{R}_{\tilde{\mathbf{c}}} = E[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^H]$ is the channel correlation matrix of $\tilde{\mathbf{c}}$ and $\tilde{\mathbf{c}}$ is the CIR with the estimated channel length $\tilde{N}_h$. Similar to the DFT-based LS channel estimation technique, the DFT-based LMMSE channel estimation technique also assumes that the temporal span of the channel is concentrated over a small number of coefficients, and thus the noise in the coefficients beyond the channel length is removed in the time domain. When all the channel power is maintained during the noise removing process, the average MSE between the estimated and actual channel coefficient per subcarrier is given by

$$\mathrm{MSE}_{\text{DFT-LMMSE}} = \frac{\beta}{N_c N_d \mathrm{SNR}} \sum_{k=0}^{\tilde{N}_h-1} \frac{\lambda_k}{\lambda_k + \frac{\beta}{N_c \mathrm{SNR}}}. \tag{6.29}$$

If $\tilde{N}_h$ is smaller than N_h, there exists an MSE error floor given by

$$\mathrm{MSE}_{\text{DFT-LMMSE, floor}} = \frac{1}{N_c N_d} \sum_{l=\tilde{N}_h}^{N_h-1} \gamma_l \tag{6.30}$$

where γ_l is the diagonal elements of $\mathbf{R}$ in decreasing order with l being the index.

Comparing (6.29) with (6.16), we see that the DFT-based LMMSE technique improves over the DFT-based LS technique since $\sum_{k=0}^{\tilde{N}_h-1} \lambda_k/(\lambda_k + \beta/N_c\text{SNR})$ is always smaller than $\tilde{N}_h$. Like the DFT-based LMMSE channel estimation technique, the DFT-based LMMSE channel estimation technique is well suited for sample-spaced channels. Its performance will be deteriorated for non-sample-spaced channels where the channel power will not be concentrated within a small number of coefficients [19, 56]. Compared to the frequency-domain LMMSE technique, the complexity of the DFT-based LMMSE technique is reduced because the size of the matrix $(\mathbf{F}_1^H \underline{\mathbf{X}}^H \underline{\mathbf{X}} \mathbf{F}_1)^{-1}]^{-1}(\mathbf{F}_1^H \mathbf{F}_1)$, which is required to be inverted, is reduced. In addition, the DFT-based LMMSE technique can achieve the same performance as the frequency-domain LMMSE technique when $\tilde{N}_h = N_h$ because the channel correlation in both cases are interchangeable. However, the performance of the DFT-based LMMSE technique will be degraded when $\tilde{N}_h$ is smaller than N_h as part of the channel correlation is ignored. Therefore, the main advantage of the DFT-based LMMSE technique is its flexibility in the trade-off between performance and complexity by the choice of estimated channel length $\tilde{N}_h$.

- **SVD based Channel Estimation**

The SVD based channel estimation technique is the low-rank approximation of the LMMSE technique [57]. This technique first reduces the complexity of the LMMSE technique by replacing $(\underline{\mathbf{X}}^H \underline{\mathbf{X}})^{-1}$ with its expectation $E[(\underline{\mathbf{X}}^H \underline{\mathbf{X}})^{-1}]$ in (6.23) such that the matrix inversion is not needed every time. Then, the estimated channel coefficients based on this simplified LMMSE channel estimation technique are given by

$$\mathbf{h}_{\text{SLMMSE}} = \mathbf{R}\left(\mathbf{R} + \frac{\beta}{N_c\text{SNR}}\mathbf{I}_{N_d}\right)^{-1}\mathbf{h}_{\text{LS}}. \tag{6.31}$$

Applying SVD on $\mathbf{R}$ by substituting (6.26) into (6.31) yields

$$\mathbf{h}_{\text{SLMMSE}} = \mathbf{U}\mathbf{\Lambda}\left(\mathbf{\Lambda} + \frac{\beta}{N_c\text{SNR}}\mathbf{I}_{N_d}\right)^{-1}\mathbf{U}^H\mathbf{h}_{\text{LS}}. \tag{6.32}$$

Then, if we retain the largest $\tilde{N}_h$ eigenvalues and ignore all the remaining ones, the resulting estimated channel coefficients are given by

$$\hat{\mathbf{h}}_{\text{SVD}} = \mathbf{U}_1\mathbf{\Lambda}'\left(\mathbf{\Lambda}' + \frac{\beta}{N_c\text{SNR}}\mathbf{I}_{\tilde{N}_h}\right)^{-1}\mathbf{U}_1^H\mathbf{h}_{\text{LS}} \tag{6.33}$$

where $\mathbf{U}_1$ contains the first $\tilde{N}_h$ columns of $\mathbf{U}$, $\mathbf{\Lambda}'$ contains the upper left $\tilde{N}_h \times \tilde{N}_h$ corner of $\mathbf{\Lambda}$ and $\mathbf{I}_{\tilde{N}_h}$ is an $\tilde{N}_h \times \tilde{N}_h$ identity matrix. This low-rank technique is referred to as the SVD based channel estimation technique. It first transforms the estimated channel coefficients based on the frequency-domain LS channel estimation technique from the frequency domain into a transformed domain. Then, Wiener filtering is performed on this transformed domain under the assumption that the channel power is concentrated within its first $\tilde{N}_h$ coefficients. The resulting estimated channel coefficients are finally transformed back to the frequency domain. It is clear from (6.33) that this SVD based channel estimation technique results in low computational complexity since the matrices in its inversion operation are diagonal and the size is reduced to $\tilde{N}_h \times \tilde{N}_h$. The average MSE per subcarrier based on the SVD channel estimation technique is given by [57]:

$$\mathrm{MSE}_{\mathrm{SVD}} = \frac{1}{N_d} \sum_{k=0}^{\tilde{N}_h - 1} \left[\lambda_k (\kappa_k - 1)^2 + \frac{\beta}{N_c \mathrm{SNR}} \kappa_k^2 \right] \tag{6.34}$$

when $\tilde{N}_h - 1$ is chosen such that $\lambda_k = 0$ $(k \geq \tilde{N}_h)$, where

$$\kappa_k = \begin{cases} \lambda_k / (\lambda_k + \beta / N_c \mathrm{SNR}), & \text{if } k \in [0, \tilde{N}_h - 1], \\ 0, & \text{otherwise.} \end{cases} \tag{6.35}$$

If $\tilde{N}_h$ is smaller than N_h, the resulting MSE error floor is given by

$$\mathrm{MSE}_{\mathrm{SVD,\ floor}} = \frac{1}{N_c N_d} \sum_{l=\tilde{N}_h}^{N_d - 1} \lambda_l. \tag{6.36}$$

– *Link Between DFT and SVD Based Channel Estimation*

The DFT and SVD based channel estimations are closely related because they both linearly transform the estimated channel coefficients from the frequency domain into another domain. It is obvious that if the channel correlation matrix can be expressed as

$$\mathbf{R} = \mathbf{F} \mathbf{\Lambda}_f \mathbf{F}^H \tag{6.37}$$

where $\mathbf{\Lambda}_f$ is a diagonal matrix, the DFT and SVD based channel estimation techniques are equivalent. Obviously, sample-spaced channels satisfy the condition indicated in (6.37). Further, it is easy to show that (6.29) is equivalent to (6.34) if the channel power is concentrated within the first $\tilde{N}_h$ coefficients in the time domain as well as in the SVD-transformed domain. Therefore, in such cases, the DFT and SVD based channel estimation techniques achieve the same performance. However, if part of the channel power is leaked into other coefficients, it is seen from (6.36) and (6.30) that the MSE error floors of these two techniques are not the same: $\mathrm{MSE}_{\mathrm{SVD,\ floor}}$ is always smaller than $\mathrm{MSE}_{\mathrm{DFT\text{-}LMMSE,\ floor}}$ since the channel power in the SVD-transformed domain is maximized in its first $\tilde{N}_h$ coefficients. This result indicates that the performance of the SVD based channel estimation technique will not be worse than that of the DFT-based channel estimation technique. Furthermore, the matrix that requires the inversion operation is diagonal in (6.33) but that in (6.28) may not be diagonal. This observation suggests that the SVD based channel estimation technique is more efficient to implement when the channel correlation R is constant, which is easily satisfied, *e.g.* in indoor environments.

6.5 Conclusions

We have reviewed some recent developments in pilot-aided channel estimation techniques for OFDM systems in this chapter. Because of the broad application of OFDM systems, it is hard to tell in general which technique is the best. However, the advantages and disadvantages for each technique have been provided so that we can choose techniques for a specific OFDM application. In order to meet various requirements of MIMO-OFDM systems, both the LS and LMMSE techniques will be covered in this thesis. Before we start to describe these techniques, we first introduce the angle-domain MIMO-OFDM systems in the next chapter. Both the current and next chapter serve as the basis for the development of our reduced-complexity angle-domain channel estimation techniques for MIMO-OFDM systems.

Chapter 7

Angle-Domain MIMO-OFDM Systems

7.1 Introduction

Chapter 6 has introduces orthogonal frequency division multiplexing (OFDM) systems and the associated channel estimation techniques. Recent research trends have shown that combining the multiple-input multiple-output (MIMO) technology with orthogonal frequency division multiplexing (OFDM) can help to achieve spatial diversity and/or space-division multiplexing gain [12, 163, 173, 198]. In this chapter, we start to describe MIMO-OFDM systems in the angle domain, which motives the development of our novel channel estimation techniques to be discussed in the following two chapters.

A typical MIMO-OFDM channel is conceived as the unique link between the transmitted and noiseless received signals. The corresponding model is referred to as the *array-domain* channel model. The array-domain channel is treated as spatially uncorrelated in most of previous pilot-aided channel estimation techniques for MIMO-OFDM systems (*e.g.* [17, 168, 173]) possibly due to the fact that early MIMO studies assume the array-domain channel to be spatially uncorrelated (*e.g.* [68, 175]). We call these techniques least squares (LS)-based techniques in the array domain. However, in many realistic scenarios, the MIMO-OFDM channel tends to be spatially correlated, for example, due to antenna spacing constraints and limited scattering [132, 166, 167]. For these spatially correlated MIMO-OFDM systems, the LMMSE-based techniques in the array domain, which exploit and require the prior knowledge of channel spatial correlation, yield better performance than LS-based techniques in the array domain [59, 133, 200]. But when the channel spatial correlation is not *a priori* known, these techniques are not applicable. In such cases, to improve the performance of conventional LS-based techniques in the array domain, we investigate techniques in the *angle domain*, where the channel model lends itself to a simple physical interpretation.

As introduced in Chapter 1, the angle domain is used to characterize the MIMO channel in a physically oriented fashion. The angle-domain representation of MIMO channels uses antenna beamforming patterns with different main lobes to characterize the physical propagation environment [162, 182]. For a MIMO system with N_t transmit and N_r receive antennas, the angle domain is partitioned into $(N_t \times N_r)$ angle-domain bins. Each angle-domain bin corresponds to the cross-section of one transmit lobe and one receive lobe from the beamforming pattern. Then, multiple unresolvable physical paths that occur in the angle-domain bin can be approximately aggregated into one resolvable path, and the paths from other directions will have little effect on this resolvable path because they originate or end at other lobes. Consequently, different physical paths approximately contribute to different angle-domain bins, and the channel coefficients in different angle-domain bins can be assumed to be approximately spatially uncorrelated. Further, as indicated in [182], when some angle-domain bins contain few physical paths due to limited scattering, the corresponding channel coefficients should approach zero. Based on these two special properties for the angle-domain channel coefficients, we develop several novel channel estimation techniques for MIMO-OFDM systems in the following two chapters.

Note that in MIMO-OFDM systems, we classify the signals and channels in two domain types. The first type is represented by either array or angle domain, the latter one is represented by either time or frequency domain. Hereinafter, we will explicitly state which representation is used for each domain type unless indicated otherwise. For example, the angle-time domain means that the angle and time representations are used for the above two domain types, respectively. When we state the representation for only one domain type, we mean that both representations for the other domain type are applicable unless indicated otherwise. For example, the angle domain refers to either the angle-time or angle-frequency domain. Further, the group of angle-time domain bins that correspond to the identical angular lobes (but with different time indices) is called the angle-time domain *beam*. Similarly, the group of angle-frequency domain bins that correspond to the identical angular lobes (but with different subcarrier indices) is called the angle-frequency domain beam.

This section is organized as follows. Section 7.2 describes the conventional array-domain representation of MIMO-OFDM systems. Then, Section 7.3 describes the angle-domain representation of MIMO-OFDM systems. Finally, Section 7.4 discusses the pilot design for the ease of the direct implementation of angle-domain channel estimation techniques to be discussed in the following two chapters.

7.2 MIMO-OFDM Systems

For the sake of convenience and convention, we refer to the time-domain and frequency-domain as the array-time domain and array-frequency domain, respectively, in this section. In a typical MIMO-OFDM system with N_t transmit and N_r receive antennas as shown in Fig. 7.1, the high rate symbols to be transmitted are first grouped into blocks of N_d data symbols at the transmitter. These groups are called frequency-domain OFDM symbols and the nth group at the (i_t)th transmitter is represented by the vector $\mathbf{x}_{i_t}(n) = [x_{i_t}(0,n),\, x_{i_t}(1,n),\, \cdots,\, x_{i_t}(N_d-1,n)]^T$, where i_t and n denote the indices of the transmitter, and OFDM symbol, respectively. Next, an inverse discrete Fourier transform (IDFT) block is applied to each OFDM symbol at each transmitter. The IDFT block at the transmitter and the discrete Fourier transform (DFT) block at the receiver serve to modulate and demodulate the data on the orthogonal subcarriers, respectively. At the

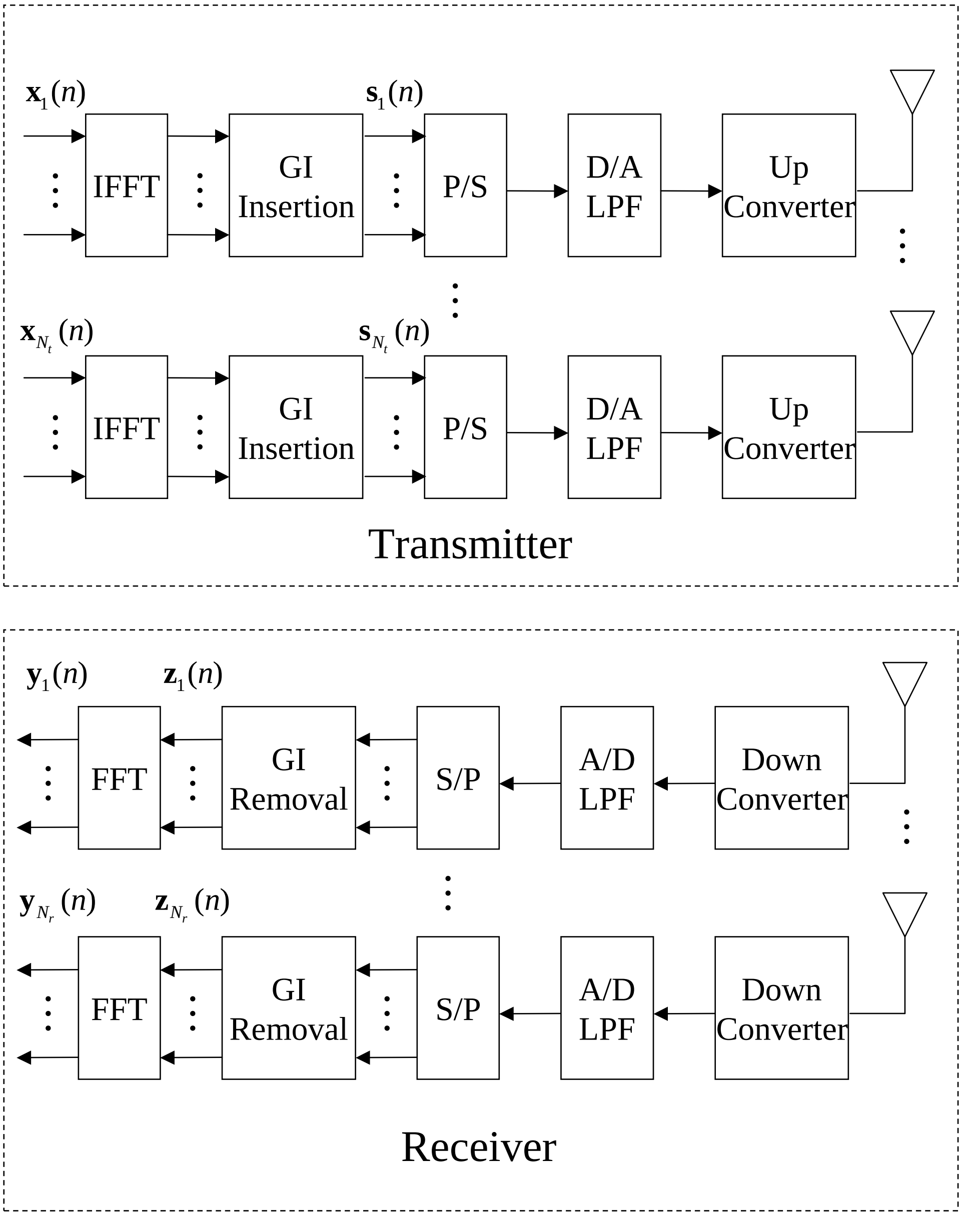

Figure 7.1: Block diagram of a typical MIMO-OFDM system with N_t transmit and N_r receive antennas.

IDFT output (*i.e.* in the time domain), a cyclic prefix (CP) of length N_g as a copy of the last part of the current OFDM symbol, is inserted at the beginning of each symbol to avoid ISI, and its length N_g is assumed to be not shorter than the channel length. The resulting nth time-domain OFDM symbol at the (i_t)th transmitter is represented by $\mathbf{s}_{i_t}(n) = [s_{i_t}(0, n),\ s_{i_t}(1, n),\ \cdots,\ s_{i_t}(N_d + N_g - 1, n)]^T$, where the m th sample is given by

$$s_{i_t}(m, n) = \frac{1}{N_d} \sum_{l=0}^{N_d-1} x_{i_t}(l, n) e^{j2\pi l(m-N_g)/N_d} \tag{7.1}$$

for $m = 0, 1, \cdots, N_d + N_g - 1$.

The samples $\{s_{i_t}(m, n)\}$ are sent through a frequency-selective fading channel, which can be represented by an equivalent discrete-time linear finite duration channel impulse response (CIR) given by a sequence of channel matrices $\mathbf{C}(l)$ for $l = 0, 1, \cdots, N_h - 1$, where N_h is the temporal span of the MIMO channel and $\mathbf{C}(l)$ is an $N_r \times N_t$ matrix whose (i_r, i_t)th element $c_{i_r,i_t}(l)$ represents the channel coefficients from the (i_t)th transmit antenna to the (i_r)th receive antenna at delay l, *i.e.*

$$\mathbf{C}(l) = \begin{bmatrix} c_{0,0}(l) & c_{0,1}(l) & \cdots & c_{0,N_t-1}(l) \\ c_{1,0}(l) & c_{1,1}(l) & \cdots & c_{1,N_t-1}(l) \\ \vdots & \vdots & \ddots & \vdots \\ c_{N_r-1,0}(l) & c_{N_r-1,1}(l) & \cdots & c_{N_r-1,N_t-1}(l) \end{bmatrix}. \tag{7.2}$$

For the sake of simple description, $\mathbf{C}(l)$ represents the channel matrix at each integer time index. This representation is obvious for the sample-spaced channels. For the non-sample-spaced channels, $\mathbf{C}(l)$ can be obtained via the interpolation as shown in [19]. Here the index 'n' is omitted from $\mathbf{C}(l)$ as we assume that the channel remains time-invariant. When the transmitter and receiver are perfectly synchronized, the received OFDM symbols become free of ISI when the CP is removed from each symbol at all the N_r receivers. By performing an N_d-point DFT on the resulting symbols at each receiver, we get the corresponding frequency-domain symbols. The resulting frequency-domain data sample at the kth subcarrier and (i_r)th receiver in the nth OFDM symbol is given by

$$y_{i_r}(k,n) = \sum_{i_t=0}^{N_t-1} h_{i_r,i_t}(k) x_{i_t}(k,n) + \vartheta_{i_r}(k,n) \tag{7.3}$$

where $h_{i_r,i_t}(k)$ is the channel coefficient at the kth subcarrier from the (i_t)th transmitter to the (i_r)th receiver, and the noise $\vartheta_{i_r}(k,n)$ is assumed to be additive white Gaussian (AWGN) with variance σ_f^2. In a compact notation, the received block of the nth OFDM symbols at the kth subcarrier can be written as

$$\mathbf{y}(k,n) = \mathbf{H}(k)\ \mathbf{x}(k,n) + \boldsymbol{\vartheta}(k,n) \tag{7.4}$$

where $\mathbf{y}(k,n) = [y_0(k,n),\ y_1(k,n),\ \cdots,\ y_{N_r-1}(k,n)]^T$, $\mathbf{x}(k,n) = [x_0(k,n),\ x_1(k,n),\ \cdots,\ x_{N_t-1}(k,n)]^T$, $\boldsymbol{\vartheta}(k,n) = [\vartheta_0(k,n),\ \vartheta_1(k,n),\ \cdots,\ \vartheta_{N_r-1}(k,n)]^T$ and

$$\mathbf{H}(k) = \begin{bmatrix} h_{0,0}(k) & h_{0,1}(k) & \cdots & h_{0,N_t-1}(k) \\ h_{1,0}(k) & h_{1,1}(k) & \cdots & h_{1,N_t-1}(k) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r-1,0}(k) & h_{N_r-1,1}(k) & \cdots & h_{N_r-1,N_t-1}(k) \end{bmatrix} \tag{7.5}$$

is the channel transfer function (CTF) matrix at the kth subcarrier. Further, the received nth OFDM symbol in the time domain is given by

$$\mathbf{z}(l,n) = \sum_{m=0}^{N_h-1} \mathbf{C}(m)\mathbf{s}(l-m,n) + \mathbf{u}(l,n) \tag{7.6}$$

where $\mathbf{s}(l,n) = [s_0(l,n),\, s_1(l,n),\, \cdots,\, s_{N_t-1}(l,n)]^T$, $\mathbf{z}(l,n) = [z_0(l,n),\, z_1(l,n),\, \cdots,\, z_{N_r-1}(l,n)]^T$, and $\mathbf{u}(l,n) = [u_0(l,n),\, u_1(l,n),\, \cdots,\, u_{N_r-1}(l,n)]^T$ are the transmitted, received, and noise vectors, respectively, at the lth sample of the nth OFDM symbol, $s_{i_t}(l,n)$, $z_{i_r}(l,n)$ and $u_{i_r}(l,n)$ represent the time-domain transmitted signal at the (i_t)th transmitter, the received signal at the (i_r)th receiver, and the additive white Gaussian noise (AWGN) with variance σ_f^2 at the (i_r)th receiver, respectively, at the lth sample of the nth OFDM symbol. If the columns of $\mathbf{C}(l)$ and $\mathbf{H}(k)$ are stacked into the vectors $\mathbf{c}(l)$ and $\mathbf{h}(k)$, respectively, the time- and frequency-domain variables are related by

$$\underline{\mathbf{y}}(n) = (\mathbf{F} \otimes \mathbf{I}_{N_r})\underline{\mathbf{z}}(n), \tag{7.7}$$

$$\underline{\mathbf{x}}(n) = (\mathbf{F} \otimes \mathbf{I}_{N_t})\underline{\mathbf{s}}(n), \tag{7.8}$$

$$\underline{\mathbf{h}} = (\mathbf{F} \otimes \mathbf{I}_{N_r \times N_t})\underline{\mathbf{c}}, \tag{7.9}$$

$$\underline{\mathbf{c}} = (\mathbf{F}^H \otimes \mathbf{I}_{N_r \times N_t})\underline{\mathbf{h}} \tag{7.10}$$

where $\underline{\mathbf{y}}(n) = [\mathbf{y}^T(0,n),\ \mathbf{y}^T(1,n),\ \cdots,\ \mathbf{y}^T(N_d-1,n)]^T$, $\underline{\mathbf{z}}(n) = [\mathbf{z}^T(0,n),\ \mathbf{z}^T(1,n),\ \cdots,$ $\mathbf{z}^T(N_d-1,n)]^T$, $\underline{\mathbf{x}}(n) = [\mathbf{x}^T(0,n), \mathbf{x}^T(1,n), \cdots, \mathbf{x}^T(N_d-1,n)]^T$, $\underline{\mathbf{s}}(n) = [\mathbf{s}^T(0,n), \mathbf{s}^T(1,n),$ $\cdots,\ \mathbf{s}^T(N_d-1,n)]^T$, $\underline{\mathbf{h}} = [\mathbf{h}^T(0),\ \mathbf{h}^T(1),\ \cdots,\ \mathbf{h}^T(N_d-1)]^T$, $\underline{\mathbf{c}} = [\mathbf{c}^T(0),\ \mathbf{c}^T(1),\ \cdots,$ $\mathbf{c}^T(N_h-1), \mathbf{0}^T_{1\times[(N_d-N_h)N_tN_r]}$, $\otimes$ and H denote the Kronecker product and Hermitian transpose, respectively, $\mathbf{F}$, $\mathbf{I}_N$ and $\mathbf{0}_{N_1\times N_2}$ are the $N_d \times N_d$ unitary Fourier matrix, $N \times N$ identity matrix and $N_1 \times N_2$ zero matrix, respectively. Note that (7.10) is obtained from (7.9) since $(\mathbf{F}\otimes\mathbf{I}_{N_r\times N_t})^{-1} = (\mathbf{F}^{-1}\otimes\mathbf{I}_{N_r\times N_t}) = (\mathbf{F}^H\otimes\mathbf{I}_{N_r\times N_t})$.

7.3 Angle-Domain MIMO-OFDM Systems

In this section, we start to represent MIMO-OFDM systems in the angle domain. After establishing the angle-domain representation of MIMO-OFDM systems, we design suitable pilots in the next section for the purpose of direct implementation of our angle-domain channel estimation techniques. This section is an extension of the work in [182] for MIMO flat-fading systems.

7.3.1 Angle-Time Domain MIMO-OFDM Systems

As the channel models for MIMO-OFDM systems are commonly introduced in the array-time domain [36,61], we start to represent the corresponding angle-time domain MIMO-OFDM systems from (7.6). Later, we will also introduce the angle-frequency domain MIMO-OFDM systems. Suppose there is an arbitrary number of physical paths between the transmit and receive antennas at time l; the ith path has an attenuation of a_i with an angle ϕ_i^t ($\Omega_i^t := \sin\phi_i^t$) and ϕ_i^r ($\Omega_i^r := \sin\phi_i^r$) for the transmit and receive antennas, respectively. Then, $\mathbf{C}(l)$ is given by

$$\mathbf{C}(l) = \sum_i a_i^b \mathbf{e_r}(\Omega_i^r)\mathbf{e_t}^H(\Omega_i^t) \tag{7.11}$$

where

$$a_i^b \ := \ a_i\sqrt{N_tN_r}\exp\left(\frac{\mathrm{j}2\pi d_i}{\lambda_c}\right), \tag{7.12}$$

$$\mathbf{e_r}(\Omega_i^r) := \frac{1}{\sqrt{N_r}} \begin{bmatrix} 1 \\ \exp\left[\mathrm{j}(2\pi\Delta_r\Omega_i^r)\right] \\ \vdots \\ \exp\left[\mathrm{j}(N_r-1)(2\pi\Delta_r\Omega_i^r)\right] \end{bmatrix}, \tag{7.13}$$

$$\mathbf{e_t}(\Omega_i^t) := \frac{1}{\sqrt{N_t}} \begin{bmatrix} 1 \\ \exp\left[\mathrm{j}(2\pi\Delta_t\Omega_i^t)\right] \\ \vdots \\ \exp\left[\mathrm{j}(N_t-1)(2\pi\Delta_t\Omega_i^t)\right] \end{bmatrix}, \tag{7.14}$$

where the superscript 'H' denotes the Hermitian transpose, $\{N_t,\ \Delta_t,\ \mathbf{e_t}(\Omega_i^t)\}$ and $\{N_r,\ \Delta_r,\ \mathbf{e_r}(\Omega_i^r)\}$ are the number of antennas, the separation between adjacent antennas normalized by λ_c, and the array response vectors[1], respectively, for the transmit and receive antennas, respectively, λ_c is the wavelength of the center frequency in the whole signal bandwidth, and d_i is the distance between the last transmit and receive antennas along path i. For notational convenience, we ignore the time index l in these three variables. We also assume that the fractional bandwidth is small such that λ_c, $\mathbf{e_r}(\Omega_i^r)$, and $\mathbf{e_t}(\Omega_i^t)$ are approximated to be unchanged over the whole signal bandwidth. Note that there exist three implicit assumptions when the array response vectors are defined in (7.13) and (7.14). First, all the transmit and receive antennas have the same polarization and radiation patterns, and the geometry of the antenna array is the uniform linear array. Second, the spacing between the antennas at the transmitter or receiver is much smaller than the distances between the scatterers and antenna arrays so that the paths from a given scatterer to all the transmit or receive antennas are approximated to be parallel. This is a typical assumption for the analysis of MIMO-OFDM systems [62, 80, 182]. Third, the mutual couplings between adjacent antennas, and those between the antenna and its surroundings, are ignorable.

As from [182], the orthonormal bases for the angle-time transmitted and received

[1] In literature, the array response vector has been also alternatively called the array vector, the array steering vector, the array propagation vector and the array manifold vector.

signals are given by

$$\boldsymbol{\xi}_t := \left\{ \mathbf{e_t}(0), \mathbf{e_t}\left(\frac{1}{L_t}\right), \cdots, \mathbf{e_t}\left(\frac{N_t - 1}{L_t}\right) \right\}, \tag{7.15}$$

and

$$\boldsymbol{\xi}_r := \left\{ \mathbf{e_r}(0), \mathbf{e_r}\left(\frac{1}{L_r}\right), \cdots, \mathbf{e_r}\left(\frac{N_r - 1}{L_r}\right) \right\}, \tag{7.16}$$

respectively, where $L_t = N_t \Delta_t$ and $L_r = N_r \Delta_r$ are the normalized antenna array lengths of the transmitter and receiver, respectively. Let $\mathbf{U}_t$, $\mathbf{U}_r$ be the unitary matrices whose columns are the basis vectors in (7.15) and (7.16), respectively. They are unitary FFT matrices when the geometry of the antenna array is assumed to be the uniform linear array as described in [182]. Then, we can transform the lth samples of the nth transmitted and received OFDM symbol from the array-time domain into the angle-time domain by

$$\mathbf{s}^{\mathrm{a}}(l, n) \quad := \quad \mathbf{U}_t^H \mathbf{s}(l, n) \tag{7.17}$$

and

$$\mathbf{z}^{\mathrm{a}}(l, n) \quad := \quad \mathbf{U}_r^H \mathbf{z}(l, n), \tag{7.18}$$

respectively, where the superscript 'a' denotes the angle-domain variables. From (7.6), we obtain the angle-time domain MIMO-OFDM system equation as

$$\mathbf{z}^{\mathrm{a}}(l, n) = \sum_{m=0}^{N_h - 1} \mathbf{C}^{\mathrm{a}}(m) \mathbf{s}^{\mathrm{a}}(l - m, n) + \mathbf{u}^{\mathrm{a}}(l, n) \tag{7.19}$$

where

$$\begin{aligned} \mathbf{C}^{\mathrm{a}}(l) \quad &:= \quad \begin{bmatrix} c^{\mathrm{a}}_{0,0}(l) & c^{\mathrm{a}}_{0,1}(l) & \cdots & c^{\mathrm{a}}_{0,N_t-1}(l) \\ c^{\mathrm{a}}_{1,0}(l) & c^{\mathrm{a}}_{1,0}(l) & \cdots & c^{\mathrm{a}}_{1,N_t-1}(l) \\ \vdots & \vdots & \ddots & \vdots \\ c^{\mathrm{a}}_{N_r-1,0}(l) & c^{\mathrm{a}}_{N_r-1,1}(l) & \cdots & c^{\mathrm{a}}_{N_r-1,N_t-1}(l) \end{bmatrix} \\ &= \quad \mathbf{U}_r^H \mathbf{C}(l) \mathbf{U}_t \end{aligned} \tag{7.20}$$

is the angle-time domain channel matrix and

$$\mathbf{u}^{\mathrm{a}}(l, n) = \mathbf{U}_r^H \mathbf{u}(l, n) \tag{7.21}$$

is the angle-time domain noise vector that still has an independent and identically distributed (i.i.d.) multivariate complex normal distribution. As explained in [182], due to the finite number of antennas, multiple unresolvable physical paths can be appropriately aggregated into one resolvable path with gain $c^{\mathrm{a}}_{k_r,k_t}(l)$. This gain uniquely corresponds to the channel coefficient for the (k_r, k_t) angle domain bin at delay l. Hence, different physical paths (approximately) contribute to different elements of $\mathbf{C}^{\mathrm{a}}(l)$. This means that the angle-time domain channel matrix $\mathbf{C}^{\mathrm{a}}(l)$ lends itself to a physical interpretation. For example, in Fig. 1.4, both path 1 and 2 contribute to the element $c^{\mathrm{a}}_{0,0}(l)$. Path 3 and 4 contribute to the elements $c^{\mathrm{a}}_{3,3}(l)$ and $c^{\mathrm{a}}_{1,2}(l)$, respectively. All the other elements of $\mathbf{C}^{\mathrm{a}}(l)$ approach zero because no paths contribute to these elements.

7.3.2 Angle-Frequency Domain MIMO-OFDM Systems

By analogy to the angle-time domain case, from (7.4), we obtain the angle-frequency domain MIMO-OFDM system equation at the kth subcarrier as

$$\mathbf{y}^{\mathrm{a}}(k,n) = \mathbf{H}^{\mathrm{a}}(k)\ \mathbf{x}^{\mathrm{a}}(k,n) + \boldsymbol{\vartheta}^{\mathrm{a}}(k,n) \tag{7.22}$$

where

$$\begin{aligned} \mathbf{x}^{\mathrm{a}}(k,n) &:= \mathbf{U}_t^H \mathbf{x}(k,n), & (7.23)\\ \mathbf{y}^{\mathrm{a}}(k,n) &:= \mathbf{U}_r^H \mathbf{y}(k,n), & (7.24)\\ \boldsymbol{\vartheta}^{\mathrm{a}}(k,n) &:= \mathbf{U}_r^H \boldsymbol{\vartheta}(k,n), & (7.25)\\ \mathbf{H}^{\mathrm{a}}(k) &:= \mathbf{U}_r^H \mathbf{H}(k)\mathbf{U}_t. & (7.26) \end{aligned}$$

7.4 Pilot Design

Angle-domain channel estimation techniques can use conventional array-domain estimators as the coarse estimators and perform post-processing in the angle domain. This follows three steps: first, performing the coarse channel estimation (*e.g.* the LS technique) in the array domain; then, in the post-processor, transforming the estimated channel from the array domain into the angle-time or angle-frequency domain where our proposed angle-domain channel estimation techniques are performed to refine the coarse channel estimation; finally, transforming back the refined estimated channel into the array-frequency domain. But the materialization as a post-processor is not a requirement for the angle-domain techniques. They can be also directly implemented in the angle domain by first transforming the transmitted and received signals from the array domain into the angle domain. In this section, we introduce the design of suitable pilots to facilitate the direct implementation of angle-domain techniques. The idea of pilot design is to exploit the property of Fourier transform and incorporate the DFT as an integral part of the estimation structure.

For the direct implementation of angle-domain techniques, we may first represent the transmitted and received signals from the array domain into the angle domain. Then, using the angle-domain transmitted and received signals to directly estimate the angle-domain channel coefficients. In such a manner, we could pretransform the transmitted signals with $\mathbf{U}_{\mathrm{t}}$ and received signals with $\mathbf{U}_{\mathrm{r}}^{H}$ either before or after FFT/IFFT block so that all the transmitted signals and received signals are represented in the angle domain. For this reason, we may add a block to perform transformation $\mathbf{U}_{\mathrm{r}}^{H}$ in the estimation block of the receiver. However, the transformation block may not be required in the transmitter because of two reasons. First, adding a block requires the adaptation of the existing transmitter architecture and the increase of hardware requirements. Second, this block might not be used in a data transmission period if the channel is unknown at the transmitter. To have a better understanding of the second statement, consider the case in which some of the angle-time domain beams contain no scatterers. Then, the transmitted data signals in this beam will be completely lost, which is undesirable when the spatial multiplexing gain is of the utmost concern[2]. On the other hand, when this pretransformation block is not incorporated, transmitted signals are spread across all the angle-time domain beams. Then, at least part of the information for each transmitted data signal is retained.

[2]Spatial interleaving may be used to reduce this effect but will further increase the complexity of code design.

From the above, instead of incorporating the pretransformation block $\mathbf{U}_{\mathrm{t}}$ in the transmitter, we may directly design the pilots in the angle domain. This can be achieved by pretransforming the arbitrary existing array-frequency domain pilots (*e.g.* [17]) with the use of (7.23) at each subcarrier. Then, the obtained new pilots are in the angle-frequency domain and can be directly used for transmission. In this case, after transforming the received signals with the $\mathbf{U}_r^H$ block at each subcarrier as shown in (7.24), we may directly estimate the channel coefficients in the angle domain.

Angle-time domain pilots of this type have three advantages. First, the complexity of angle-domain channel estimation techniques is reduced because we can directly estimate the angle-domain channel coefficients without the use of the transformation $\mathbf{U}_t$. Second, this complexity reduction will not result in a performance degradation. Third, the architecture in the transmitter is not affected. This is desirable to apply the proposed angle-domain techniques in current MIMO-OFDM systems.

7.5 Assumptions List

In this section, we list the assumptions made in our angle-domain channel estimation techniques. These techniques will be discussed in the following two chapters. Please note that we do not intend to limit the application of our techniques by using these assumptions. Instead, we want to introduce the techniques more clearly by considering them in a well defined scenario.

Recall that in MIMO-OFDM systems, we classify the signals and channels in two domain types. The first type is represented by either array or angle domain, and the latter one is represented by either time or frequency domain. The assumptions corresponding to the first domain type are as follows:

- All the transmit and receive antennas have the same polarization and radiation patterns, and the geometry of the antenna array is the uniform linear array.

- The spacing between the antennas at the transmitter or receiver is much smaller than the distances between the scatterers and antenna arrays so that the paths from a given scatterer to all the transmit or receive antennas are approximated to be parallel.
- The mutual couplings between adjacent antennas, and those between the antenna and its surroundings, are ignorable.

When some or all of these assumptions are not satisfied, the array response vectors defined in (7.13) and (7.14) should be revised. Then, for our proposed angle-domain channel estimation techniques, only the unitary transformation matrices (*i.e.* $\mathbf{U}_t$ and $\mathbf{U}_r$) are to be modified due to the change of the array response vectors. The assumptions corresponding to the second domain type are as follows:

- The channel is assumed to be linear and time-invariant over a given training period as our main concern is the indoor propagation environment [170, 174].
- The transmitter and receiver are perfectly synchronized. The distortion and phase noise in the analog and radio frequency (RF) parts of the transmitter and receiver are ignorable.
- The length of the cyclic prefix (CP) is shorter than the temporal span of the channel.

When some or all of these assumptions are not satisfied, our angle-domain channel estimation techniques will suffer a performance degradation. Note that to deal with such situations, many techniques have been developed for the SISO-OFDM and array-domain MIMO-OFDM systems in the literature (*e.g.* [18, 95, 113]). These techniques are directly applicable for our angle-domain channel estimation techniques to compensate for the performance degradation in the second domain type, as our techniques are developed to improve the estimation performance in the first domain type, which is independent of the second domain type.

7.6 Conclusions

MIMO-OFDM systems have been shown to achieve spatial diversity and/or space-division multiplexing gain. In this chapter, we have first introduce the conventional MIMO-OFDM system in the array domain. Then, we have represented MIMO-OFDM systems in the angle domain, where the channel model is shown to to lend itself to a simple physical interpretation. The angle-domain represented MIMO-OFDM systems motivate the development of our novel channel estimation techniques to be discussed in the next two chapters. Further, we have designed suitable pilots that facilitate the direct implementation of our proposed angle-domain channel estimation techniques. Finally, we have listed the assumptions to be made in the following two chapters.

Chapter 8

Channel Instantaneous Power Based Angle-Domain Channel Estimation

8.1 Introduction

The previous two chapters describe the pilot-aided channel estimation techniques and angle-domain multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems, respectively. In this and the next chapters, we start to investigate angle-domain pilot-aided channel estimation techniques for MIMO-OFDM systems.

Generally speaking, pilot-aided channel estimation is based on either least squares (LS) [19] or linear minimum mean square error (LMMSE) technique [57,124]. The essential difference between these two types of techniques is that the channel coefficients are treated as deterministic but unknown constants in the former, and as random variables of a stochastic process in the latter. Compared with LS-based techniques, LMMSE-based techniques yield better performance because they additionally exploit and require the prior knowledge of channel correlation. However, the channel correlation is sometimes not *a priori* known, which makes LMMSE-based techniques infeasible. Therefore, in this chapter, we focus on techniques when the prior knowledge of channel correlation is not available to the receiver. Then, In the next chapter, we focus on techniques when the channel correlation or channel power is known to the receiver.

As mentioned in the previous chapter, we classify the signals and channels in two domain types in MIMO-OFDM systems. The first type is represented by either array or angle domain, the latter one is represented by either time or frequency domain. In this chapter, we investigate channel estimation techniques in both the angle-frequency and angle-time domains. In the angle-frequency domain, when some angle-frequency domain bins contain few physical paths due to limited scattering, the corresponding channel coefficients should approach zero. This allows us to choose a suitable threshold for ignoring the small-valued channel taps and retaining only the most significant taps (MST) to reduce the effect of noise on the estimated channel coefficients, thereby improving the performance of channel estimation. We call this technique of retaining channel coefficients of sufficiently large power the MST selection technique. In the angle-time domain, we can also use the MST selection technique. Furthermore, the channel coefficients in the angle-time domain are approximately uncorrelated. Thus, we may use the channel power instead of the channel correlation to approximately perform the angle-time domain LMMSE technique when the signal-to-noise ratio (SNR) is known or reliably estimated. We call the resulting technique the approximated LMMSE (AMMSE) technique. Note that we refer to the *channel power* as the channel average power in this chapter. When the channel power is not available, we can use the *channel instantaneous power* (*i.e.* the instantaneous power of estimated channel coefficients) to estimate the channel power. In such cases, to maintain the estimated channel power positive and estimation reliable, a threshold is required to ignore coefficients with low instantaneous power. Further, we will not consider the AMMSE technique in the angle-frequency domain because the channel correlation in the frequency domain might be too high to be reasonably replaced by the channel power[1].

[1]For example, when the temporal span of the channel is concentrated over a small number of taps and the number of subcarriers is quite large, the channel coefficients at different subcarriers become strongly correlated in the second domain type.

To the best of our knowledge, we are the first to systematically study angle-domain channel estimation techniques for MIMO-OFDM systems [97]. We note that one angle-domain channel estimation technique has been investigated for single-input multiple-output (SIMO)-OFDM systems [128]. In this technique, the angle-time domain bins that correspond to the identical angular lobes (but with different time indices) are grouped into one angle-time domain beam. Examining all the estimated channel coefficients in one beam, only the beams that contain significant peak values along the time axis are identified as the signal beams and thus retained. All the remaining beams are considered as noise (or interference) beams and are ignored. This technique was shown to be effective when the channel power is concentrated in a few beams. However, when the channel power is distributed over all the beams (*e.g.* the channel model E in [61]), this technique may hardly improve the performance. To overcome this problem and investigate the techniques for MIMO-OFDM systems, we do not group the angle-time domain bins into beams. Instead, we filter the noise independently in each angle-time domain bin, based on the fact that the multipath components are approximately disjoint in the angle-time domain. Then, as introduced, we use either MST selection or AMMSE techniques to estimate channels in the angle-time domain. Further, we also investigate the MST selection technique in the angle-frequency domain.

The three angle-domain channel estimation techniques proposed in this chapter have two main advantages. First, the achieved performance gain over the conventional LS technique does not require the prior knowledge of the channel correlation or even the channel power. Therefore, the techniques are applicable to various propagation environments. Second, they are flexible in implementation. They can either use conventional array-domain estimators as the coarse estimators and perform post-processing in the angle domain, or use the specifically designed pilots introduced in the previous chapter for the direct implementation.

The major contributions and results of this chapter are as follows:

1. We systematically develop the channel estimation techniques in both the angle-

time and angle-frequency domains for MIMO-OFDM systems. We find that the proposed techniques perform especially well in the angle-time domain.

2. We develop a unified approach to analyze the performance of MST selection and AMMSE channel estimation techniques in terms of mean square error. Based on this approach, we also develop a simple way to compare the performances of different angle-domain techniques with the help of the first derivative test [11].

3. The performances of all the angle-domain techniques are dependent on the respective thresholds. Nevertheless, we find that setting the threshold to be two times the noise variance is sufficient for the angle-time domain MST selection and AMMSE techniques to yield better performance than the conventional LS technique at all the SNRs for various IEEE 802.11 TGn channel models [61].

4. Of all the proposed angle-domain techniques, both our theoretical analysis and simulation results demonstrate that the angle-time domain AMMSE technique results in the best performance and achieves up to 8 dB performance gain when the mean square error is 10^{-2} compared to the conventional LS technique.

This chapter is organized as follows. Section 8.2 develops three angle-domain techniques to estimate MIMO-OFDM channels. Then, these three techniques are analyzed in Section 8.3. The performances of these techniques are also evaluated by simulation for typical IEEE 802.11 TGn channel models in Section 8.4. Finally, Section 8.5 concludes this chapter.

8.2 Angle-Domain Channel Estimation

As discussed, the physical path within one angle-domain beam has most of its energy within this beam. Thus, the elements within one angle-time domain channel matrix $\mathbf{C}^{\mathrm{a}}(l)$ defined in (7.20) at a given time delay l, or angle-frequency domain channel matrix $\mathbf{H}^{\mathrm{a}}(k)$

defined in (7.26) at the given kth subcarrier, maintain low spatial correlation[2]. Furthermore, various elements in the angle-domain channel matrices tend to approach zero due to limited scattering. Therefore, we may perform noise filtering independently in each angle-time or angle-frequency domain bin to improve the estimation performance. These channel estimation techniques are flexible in implementation. They can either use conventional array-domain estimators as the coarse estimators and perform post-processing in the angle domain, or use the specifically designed pilots introduced in Chapter 7 for the direct implementation. For the sake of a simple description, we concentrate on describing the angle-domain channel estimation techniques in the framework of a post-processor. The direct implementation of angle-domain techniques is straightforwardly extendable, and thus will not be covered in this chapter.

In typical MIMO-OFDM systems, only one transmit antenna sends pilot symbols in a given time or frequency position [17, 174, 198]. This technique can lead to very simple MIMO-OFDM channel estimation because the received signal at a given time or frequency position corresponds to one unique channel coefficient from a given transmit antenna to a given receive antenna. In such cases, the channel estimation for a MIMO-OFDM system with N_t transmit and N_r received antennas becomes the channel estimation for the total $N_t \times N_r$ single input single output (SISO)-OFDM systems. Thus, the well-developed SISO-OFDM channel estimation techniques (*e.g.* [19, 57, 95, 138]) are directly applicable to the estimation of MIMO-OFDM channels. In this chapter, we use the conventional array-frequency domain LS technique [19] to coarsely estimate the array-frequency domain $h_{i_r,i_t}(k)$ because the knowledge of channel correlation is assumed to be not available. For the sake of a simple description, we assume that pilots from different transmit antennas are time orthogonal to each other, and thus only one transmit antenna is used to transmit pilots in each OFDM training symbol period as introduced in [198]. The proposed techniques can be easily extended to other pilot transmission

[2] In the angle domain, there exists an overlapping area between adjacent angle-domain beams. When the paths fall in this overlapping area, there appears spatial correlation between these two beams. However, this situation can be approximately ignored and the angle-domain beams can be approximated as spatially uncorrelated since the overlapping area occupies only a small percentage of total angles.

schemes (*e.g.* [174]), but will not be covered in this chapter.

8.2.1 Angle-Frequency Domain Technique

Recalling from (7.4), the received block of the nth OFDM symbols at the kth subcarrier can be written as

$$\mathbf{y}(k,n) = \mathbf{H}(k)\ \mathbf{x}(k,n) + \boldsymbol{\vartheta}(k,n) \tag{8.1}$$

where $\mathbf{y}(k,n) = [y_0(k,n),\ y_1(k,n),\ \cdots,\ y_{N_r-1}(k,n)]^T$, $\mathbf{x}(k,n) = [x_0(k,n),\ x_1(k,n),\ \cdots,\ x_{N_t-1}(k,n)]^T$, $\boldsymbol{\vartheta}(k,n) = [\vartheta_0(k,n),\ \vartheta_1(k,n),\ \cdots,\ \vartheta_{N_r-1}(k,n)]^T$, $x_{i_t}(k,n)$, $y_{i_r}(k,n)$ and $\vartheta_{i_r}(k,n)$ represent the frequency-array domain transmitted data sample at the (i_t)th transmitter, the received data sample at the (i_r)th receiver, and the additive white Gaussian noise (AWGN) with variance σ_f^2 at the (i_r)th receiver, respectively, at the kth subcarrier in the nth OFDM symbol, $\mathbf{H}(k)$ defined in (7.5) is the channel transfer function (CTF) matrix at the kth subcarrier, N_t and N_r are the numbers of the transmit and receive antennas, respectively. As defined in (7.2), the channel matrix $\mathbf{C}(l)$ for $l = 0, 1, \cdots, N_h - 1$ is an $N_r \times N_t$ matrix, where N_h is the temporal span of the MIMO channel. Then, if the columns of $\mathbf{C}(l)$ and $\mathbf{H}(k)$ are stacked into the vectors $\mathbf{c}(l)$ and $\mathbf{h}(k)$, respectively, recalling from (7.10), we have

$$\underline{\mathbf{h}} = (\mathbf{F} \otimes \mathbf{I}_{N_r \times N_t})\underline{\mathbf{c}} \tag{8.2}$$

where $\underline{\mathbf{h}} = [\mathbf{h}^T(0), \mathbf{h}^T(1), \cdots, \mathbf{h}^T(N_d-1)]^T$, $\underline{\mathbf{c}} = [\mathbf{c}^T(0), \mathbf{c}^T(1), \cdots, \mathbf{c}^T(N_h-1), \mathbf{0}_{1\times(N_d-N_h)}]^T$, $\otimes$ denotes the Kronecker product, $\mathbf{F}$, $\mathbf{I}_N$ and $\mathbf{0}_{N_1\times N_2}$ are the $N_d \times N_d$ unitary Fourier matrix, $N \times N$ identity matrix and $N_1 \times N_2$ zero matrix, respectively, N_d and N_h are the number of total subcarriers, and the temporal span of the channel in the array-time domain, respectively. Then, using the assumption that the pilots from different transmit antennas are time orthogonal to each other, we obtain

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}}\ \underline{\mathbf{h}} + \underline{\boldsymbol{\vartheta}} \tag{8.3}$$

where $\underline{\mathbf{Y}} = [\underline{\ddot{\mathbf{Y}}}^T(n),\ \underline{\ddot{\mathbf{Y}}}^T(n+N_t),\ \cdots,\ \underline{\ddot{\mathbf{Y}}}^T(n+(N_c-1)N_t)]^T$ with $\underline{\ddot{\mathbf{Y}}}(n) = [\mathbf{y}^T(0,n)$, $\mathbf{y}^T(0,n+1), \cdots, \mathbf{y}^T(0,n+N_t-1), \mathbf{y}^T(1,n), \cdots, \mathbf{y}^T(1,n+N_t-1), \cdots, \mathbf{y}^T(N_d-1,n+$

$N_t - 1)]^T$, $\underline{\mathbf{X}} = [\underline{\ddot{\mathbf{X}}}(n), \underline{\ddot{\mathbf{X}}}(n+N_t), \cdots, \underline{\ddot{\mathbf{X}}}(n+(N_c-1)N_t)]^T$ with $\underline{\ddot{\mathbf{X}}}(n) = \mathrm{diag}[\mathbf{x}^T(0,n)$, $\mathbf{x}^T(0,n{+}1)$, $\cdots$, $\mathbf{x}^T(0,n{+}N_t{-}1)$, $\mathbf{x}^T(1,n)$, $\cdots$, $\mathbf{x}^T(1,n{+}N_t{-}1)$, $\cdots$, $\mathbf{x}^T(N_d{-}1,n{+}N_t{-}1)]$, and $\underline{\boldsymbol{\vartheta}} = [\underline{\ddot{\boldsymbol{\vartheta}}}^T(n), \underline{\ddot{\boldsymbol{\vartheta}}}^T(n{+}N_t), \cdots, \underline{\ddot{\boldsymbol{\vartheta}}}^T(n{+}(N_c{-}1)N_t)]^T$ with $\underline{\ddot{\boldsymbol{\vartheta}}}^T(n) = [\boldsymbol{\vartheta}^T(0,n)$, $\boldsymbol{\vartheta}^T(0,n{+}1)$, $\cdots$, $\boldsymbol{\vartheta}^T(0,n{+}N_t{-}1)$, $\boldsymbol{\vartheta}^T(1,n)$, $\cdots$, $\boldsymbol{\vartheta}^T(1,n{+}N_t{-}1)$, $\cdots$, $\boldsymbol{\vartheta}^T(N_d{-}1,n{+}N_t{-}1)]^T$ are the received signal vector, transmitted signal matrix, and noise vector, respectively, and N_c is the number of pilots used for each channel coefficient in the LS channel estimation. Note that $\underline{\mathbf{X}}$ is a block diagonal matrix. Then, the array-frequency domain LS estimator is given by

$$\underline{\mathbf{h}}_{\mathrm{LS}} = \left(\underline{\mathbf{X}}^H \underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}^H \underline{\mathbf{Y}}. \tag{8.4}$$

By rearranging the vector form $\underline{\mathbf{h}}_{\mathrm{LS}}$ into its matrix form, in this first step, we get the coarsely estimated array-frequency domain channel matrix $\tilde{\mathbf{H}}(k)$ for $k = 0, 1, \cdots, N_d - 1$.

In the second step, we transform the estimated array-frequency domain channel matrix from the array-frequency domain $\tilde{\mathbf{H}}(k)$ into the angle-frequency domain $\tilde{\mathbf{H}}^{\mathrm{a}}(k)$ by the use of (7.26) for all the subcarriers under consideration. Let $\tilde{h}^{\mathrm{a}}_{i_r,i_t}(k)$ denote the (i_r, i_t)th element of $\tilde{\mathbf{H}}^{\mathrm{a}}(k)$. Then, the filtered angle-frequency domain channel coefficient is given by comparing the power of $\tilde{h}^{\mathrm{a}}_{i_r,i_t}(k)$ with a threshold η as follows

$$\tilde{h}^{\mathrm{a}}_{i_r,i_t,\mathrm{MST}}(k) = \begin{cases} \tilde{h}^{\mathrm{a}}_{i_r,i_t}(k), & \text{if } |\tilde{h}^{\mathrm{a}}_{i_r,i_t}(k)|^2 \geq \eta, \\ 0, & \text{otherwise.} \end{cases} \tag{8.5}$$

Finally, from (7.26), we have

$$\mathbf{H}(k) = \mathbf{U}_r \mathbf{H}^{\mathrm{a}}(k) \mathbf{U}_t^H \tag{8.6}$$

where $\mathbf{U}_t$, $\mathbf{U}_r$ are the unitary matrices whose columns are the basis vectors in (7.15) and (7.16), respectively. Then, the filtered angle-frequency domain channel matrices are transformed back to the array-frequency domain via (8.6). We call this technique the angle-frequency domain MST selection technique.

To make a fair complexity comparison among all the techniques, we assume that LS estimated array-frequency domain channel coefficients are available beforehand. Then, from (7.26), obtaining the angle-frequency domain $\tilde{\mathbf{H}}^{\mathrm{a}}(k)$ from the array-frequency domain $\tilde{\mathbf{H}}(k)$ requires N_t+N_r complex multiplications for each channel coefficient. Further, transforming the estimated angle-frequency domain into the array-frequency domain also requires N_t+N_r complex multiplications for each channel coefficient. Therefore, the total required number of complex multiplications for each channel coefficient is $2(N_t+N_r)$.

8.2.2 Angle-Time Domain Techniques

In many cases, such as the non-line-of-sight (NLOS) scenario, the mean angle of departure (AoD) and angle of arrival (AoA) of the clusters of multipath components tend to be uniformly distributed over all angles [36, 172]. Thus, when the number of clusters is relatively large (*e.g.* the channel model E in [61]), the above angle-frequency domain channel estimation technique may hardly improve over the conventional array-frequency domain LS technique because nearly all the elements of angle-frequency domain channel matrices may not approach zero. As the clusters of multipath components are disjoint in the angle-time domain, we may perform the noise filtering in this domain instead of the angle-frequency domain. For the implementation of angle-time domain channel estimation techniques, we first transform the estimated channel coefficients from the array-frequency domain into the array-time domain by the use of discrete Fourier transform (DFT) [95]. Then, we transform the channel matrices into the angle-time domain by using (7.20). In the angle-time domain, we can select the MST in channel matrices to reduce the effect of noise on the estimates. Now the estimated angle-time domain channel coefficient becomes

$$\tilde{c}^{\mathrm{a}}_{i_r,i_t,\mathrm{MST}}(l) = \begin{cases} \tilde{c}^{\mathrm{a}}_{i_r,i_t}(l), & \text{if } |\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2 \geq \eta, \\ 0, & \text{otherwise}, \end{cases} \tag{8.7}$$

where $\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)$ is the coarsely estimated angle-time domain channel coefficient.

Note that we presume that the channel spatial correlation is not available to the receiver in this chapter. Therefore, the conventional LMMSE technique that utilizes the channel spatial correlation is not applicable here. But as discussed, the channel coefficients in the angle-time domain at a given time are approximately spatially uncorrelated. Therefore, we may use the channel instantaneous power to approximate the channel correlation (Here the channel power is also assumed to be not available). As the approximated channel correlation matrix is a diagonal matrix, the LMMSE technique that jointly filters all the channel coefficients becomes the independent spatial filtering for each channel coefficient. Further, the channel coefficient is uncorrelated with the noise. Therefore, we may estimate the channel instantaneous power for each coefficient as $|\tilde{\mathbf{c}}^{\mathrm{a}}(l)|^2 - \sigma_f^2$[3]. Then, the angle-time domain AMMSE technique is realized as

$$\begin{aligned} &\tilde{c}^{\mathrm{a}}_{i_r,i_t,\mathrm{AMMSE}}(l) \\ &= \begin{cases} \frac{|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2-\sigma_f^2}{|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2}\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l), & \text{if } |\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2 \geq \eta, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{8.8}$$

Here the threshold η is usually chosen to be smaller than σ_f^2. Otherwise, the approximated channel power $(|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2 - \sigma_f^2)$ becomes negative. In comparison with the MST selection technique, the difference is the use of a dynamic multiplication factor instead of a constant one. The factor turns out to be crucial in improving the performance of channel estimation as shown in the following two sections.

After the noise filtering in the angle-time domain, we transform back the estimated channel into the array-time domain, and then into the array-frequency domain. Note that the angle-time domain estimation is well-suited for the sample-spaced channels. The performance will be degraded for the non-sample-spaced channels due to the power leakage [19]. Nevertheless, our results show that the angle-time domain techniques still outperform the array-frequency domain LS technique for typical IEEE 802.11 TGn channel models.

Similar to the angle-frequency domain MST technique, we assume that LS estimated

[3]The noise variance σ_f^2 is assumed to be known here. In practice, it can be estimated during periods when no transmitted signal is detected, or at virtual carriers where no data is transmitted.

array-frequency domain channel coefficients are available beforehand to make a fair complexity comparison. Then, from (7.20), we know that the transformations between the array-time domain and angle-time domain require totally $2(N_t+N_r)$ complex multiplications for each channel coefficient. In addition, the angle-time domain techniques require the transformations between the array-frequency domain and array-time domain. This requires totally $2N_d$ complex multiplications for each channel coefficient. Typically, N_d is a power of 2. Then, using the FFT and IFFT [187] for transformations between the array-time and array-frequency domain, the total complex multiplications required for each channel coefficient is reduced to $\log_2 N_d$ complex multiplications for each channel coefficient. From the above, the total required complex multiplications for each channel coefficient in the angle-time domain MST selection technique is $2(N_t + N_r) + \log_2 N_d$. For the angle-time domain AMMSE technique, an additional complex multiplication is needed. Therefore, the total required complex multiplications for each channel coefficient in the angle-time domain AMMSE technique is $2(N_t+N_r)+1+\log_2 N_d$. Compared with the angle-frequency domain MST technique, these angle-time domain techniques always have higher complexity. Nevertheless, our results show that these angle-time domain techniques always outperform the angle-frequency domain MST technique for all the channel models under consideration.

Note that the angle-time domain MST selection and AMMSE techniques are well-suited for the sample-spaced channels. The performance will be degraded for the non-sample-spaced channels due to the power leakage [19]. Nevertheless, our results show that the angle-time domain techniques still outperform the array-frequency domain LS technique for all the channel models under consideration.

8.3 Performance Analysis

For channel estimation techniques, one of the most important performance measures is the mean square error (MSE), which measures the average mean squared deviation of the estimator from the true value [110]. In this section, we present a unified approach to computing the MSE of the angle-domain (*i.e.* either angle-time or angle-frequency domain) MST selection and AMMSE techniques. Note that the array-time or array-frequency domain is related by an unitary transformation. Thus, of a given estimation technique, the MSE represented in either the array-time or array-frequency domain yields the same result, and is given by

$$\mathrm{MSE} = \frac{1}{N_d N_t N_r} \mathrm{E}\left[||\tilde{\underline{\mathbf{h}}} - \vec{\underline{\mathbf{h}}}||^2 \right] \tag{8.9}$$

where $\vec{\underline{\mathbf{h}}}$ is either $\underline{\mathbf{h}}$ or $\underline{\mathbf{c}}$, and $\tilde{\underline{\mathbf{h}}}$ is the estimated $\vec{\underline{\mathbf{h}}}$.

Here we use the array-frequency domain LS technique to perform the coarse channel estimation. Then, the resulting MSE of the coarse estimation is given by

$$\mathrm{MSE}_{\mathrm{LS}} = \frac{1}{N_d N_t N_r} \mathrm{E}\left[||\underline{\mathbf{h}}_{\mathrm{LS}} - \underline{\mathbf{h}}||^2 \right] = \frac{\sigma_f^2}{N_d N_t N_r} \mathrm{trace}\left\{ \mathrm{E}\left[\left(\underline{\mathbf{X}}^H \underline{\mathbf{X}} \right)^{-1} \right] \right\}. \tag{8.10}$$

In many cases, such as in the IEEE 802.11a standard [6], the powers of all pilots are unity. Therefore, we also assume that $\mathrm{E}[(\underline{\mathbf{X}}^H \underline{\mathbf{X}})^{-1}]$ is the identity matrix. Then, since $\underline{\mathbf{X}}$ is diagonal, (8.10) becomes

$$\mathrm{MSE}_{\mathrm{LS}} = \sigma_f^2. \tag{8.11}$$

Let $\underline{\mathbf{h}}^{\mathrm{a}}$ represent the stacked angle domain channel vector. Then, from (7.26), we obtain

$$\underline{\mathbf{h}}^{\mathrm{a}} = \mathbf{B}\vec{\underline{\mathbf{h}}} \tag{8.12}$$

where

$$\mathbf{B} = \mathbf{I}_{N_d} \otimes \mathbf{U}_t^T \otimes \mathbf{U}_r^H \tag{8.13}$$

is an $(N_t N_r N_d \times N_t N_r N_d)$ matrix, and the superscript 'T' denotes the transpose. It is easily verified that the matrix $\mathbf{B}$ is unitary, *i.e.*

$$\mathbf{B}^H \mathbf{B} = \mathbf{I}_{N_t \times N_r \times N_d}. \tag{8.14}$$

Then, the angle domain MST selection or AMMSE technique is given by

$$\tilde{\underline{\mathbf{h}}}^{\mathrm{a}} = \mathbf{M}\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS}} \tag{8.15}$$

where $\mathbf{M}$ is an $(N_d N_t N_r \times N_d N_t N_r)$ diagonal matrix that represents either the MST selection or the AMMSE process, and $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS}}$ is the stacked LS estimated channel vector in the angle domain. Note that the ith diagonal element of $\mathbf{M}$ (denoted as m_i) is dependent on both the channel coefficient and noise. Denote $\hat{r}_i$ as the instantaneous power of ith element of $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS}}$. In the MST selection techniques, we have

$$m_i = \begin{cases} 1, & \text{if } \hat{r}_i \geq \eta, \\ 0, & \text{otherwise.} \end{cases} \tag{8.16}$$

In the AMMSE technique, from (8.8), we have

$$m_i = \begin{cases} \frac{\hat{r}_i - \sigma_f^2}{\hat{r}_i}, & \text{if } \hat{r}_i \geq \eta, \\ 0, & \text{otherwise,} \end{cases} \tag{8.17}$$

where the threshold η is chosen not to be smaller than σ_f^2.

From (8.9), the MSE of angle-domain techniques is given by

$$\mathrm{MSE} = \frac{1}{N_d N_t N_r} \mathrm{E}\left[||\mathbf{B}^H \mathbf{M} \mathbf{B} \vec{\underline{\mathbf{h}}}_{\mathrm{LS}} - \vec{\underline{\mathbf{h}}}||^2 \right] \tag{8.18}$$

where $\vec{\underline{\mathbf{h}}}_{\mathrm{LS}}$ is the LS estimated $\vec{\underline{\mathbf{h}}}$. Since $\mathbf{M}$ is diagonal and real, using (8.14), we obtain

$$\begin{aligned} \mathrm{MSE} \quad = \quad & \frac{1}{N_d N_t N_r} \mathrm{trace}\Big\{ \mathrm{E}[\mathbf{M}\underline{\mathbf{h}}^{\mathrm{a}}(\underline{\mathbf{h}}^{\mathrm{a}})^H \mathbf{M}^H + \underline{\mathbf{h}}^{\mathrm{a}}(\underline{\mathbf{h}}^{\mathrm{a}})^H \\ & -2\mathbf{M}\underline{\mathbf{h}}^{\mathrm{a}}(\underline{\mathbf{h}}^{\mathrm{a}})^H + \mathbf{M}\underline{\mathbf{v}}^{\mathrm{a}}(\underline{\mathbf{v}}^{\mathrm{a}})^H \mathbf{M}^H] \Big\} \end{aligned} \tag{8.19}$$

where $\underline{\mathbf{v}}^{\mathrm{a}}$ is the angle-domain noise vector. Denoting $\hat{h}_i$, $\hat{v}_i$ as the instantaneous power of ith element of $\underline{\mathbf{h}}^{\mathrm{a}}$ and $\underline{\mathbf{v}}^{\mathrm{a}}$, respectively, we rewrite (8.19) as

$$\mathrm{MSE} = \frac{1}{N_d N_t N_r} \sum_{i=1}^{N_d N_t N_r} \mathrm{E}\left[(m_i - 1)^2 \hat{h}_i + m_i^2 \hat{v}_i \right]. \tag{8.20}$$

More specifically,

$$\begin{aligned} \mathrm{E}\left[(m_i - 1)^2 \hat{h}_i \right] \quad = \quad & \int_0^\infty \hat{h}_i P_h(\hat{h}_i) \int_\eta^\infty (m_i - 1)^2 P_{r|h}(\hat{r}_i|\hat{h}_i) \mathrm{d}\hat{r}_i \mathrm{d}\hat{h}_i \\ & + \int_0^\infty \hat{h}_i P_h(\hat{h}_i) \int_0^\eta P_{r|h}(\hat{r}_i|\hat{h}_i) \mathrm{d}\hat{r}_i \mathrm{d}\hat{h}_i \end{aligned} \tag{8.21}$$

is the MSE part due to the filtered instantaneous power of the ith channel coefficient, and

$$\mathrm{E}\left[m_i^2\hat{v}_i\right] = \int_0^\infty \hat{v}_i P_v(\hat{v}_i) \int_\eta^\infty m_i^2 P_{r|v}(\hat{r}_i|\hat{v}_i)\mathrm{d}\hat{r}_i\mathrm{d}\hat{v}_i, \tag{8.22}$$

is the MSE part due to the unfiltered noise components in the ith estimated channel coefficient, where $P_h(\hat{h}_i)$, $P_v(\hat{v}_i)$, $P_{r|h}(\hat{r}_i|\hat{h}_i)$, and $P_{r|v}(\hat{r}_i|\hat{v}_i)$ are the probability density functions of $\hat{h}_i$, $\hat{v}_i$, conditional probability density functions of $\hat{r}_i$ conditioned on $\hat{h}_i$ and $\hat{v}_i$, respectively. Both the channel and noise have the complex normal distributions. Thus, $P_h(\hat{h}_i)$ and $P_v(\hat{v}_i)$ have the exponential distributions, and $P_{r|h}(\hat{r}_i|\hat{h}_i)$ and $P_{r|v}(\hat{r}_i|\hat{v}_i)$ have the noncentral chi-square distributions [11].

Let σ_i^2 denote the channel power of the ith element of $\underline{\mathbf{h}}^{\mathrm{a}}$.

1. When σ_i^2 is not equal to zero, the MSE of the ith estimated channel coefficient is given by

$$\mathrm{MSE}_i = \mathrm{E}\left[m_i^2\hat{v}_i\right] = \int_\eta^\infty m_i^2 \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \mathrm{d}\hat{v}_i \tag{8.23}$$

where

$$\begin{aligned}\mathrm{E}\left[(m_i-1)^2\hat{h}_i\right] &= \int_0^\infty \frac{\hat{h}_i}{\sigma_i^2} e^{-\frac{\hat{h}_i}{\sigma_i^2}} \int_\eta^\infty (m_i-1)^2 \frac{e^{-\frac{\hat{r}_i+\hat{h}_i}{\sigma_f^2}}}{\sigma_f^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{h}_i}}{\sigma_f^2}\right)\mathrm{d}\hat{r}_i\mathrm{d}\hat{h}_i \\ &\quad + \int_0^\infty \frac{\hat{h}_i}{\sigma_i^2} e^{-\frac{\hat{h}_i}{\sigma_i^2}} \int_0^\eta \frac{e^{-\frac{\hat{r}_i+\hat{h}_i}{\sigma_f^2}}}{\sigma_f^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{h}_i}}{\sigma_f^2}\right)\mathrm{d}\hat{r}_i\mathrm{d}\hat{h}_i,\end{aligned} \tag{8.24}$$

and

$$\mathrm{E}\left[m_i^2\hat{v}_i\right] = \int_0^\infty \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \int_\eta^\infty m_i^2 \frac{e^{-\frac{\hat{r}_i+\hat{v}_i}{\sigma_i^2}}}{\sigma_i^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{v}_i}}{\sigma_i^2}\right)\mathrm{d}\hat{r}_i\mathrm{d}\hat{v}_i, \tag{8.25}$$

and $J_0(z)$ is the modified Bessel function of the first kind and zero order.

2. When σ_i^2 is equal to zero, $\hat{h}_i = 0$ and $\hat{r}_i = \hat{v}_i$, then $\mathrm{E}[(m_i-1)^2\hat{h}_i] = 0$ and the MSE of the ith estimated channel coefficient is simplified as

$$\mathrm{MSE}_i = \mathrm{E}\left[m_i^2\hat{v}_i\right] = \int_\eta^\infty m_i^2 \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \mathrm{d}\hat{v}_i. \tag{8.26}$$

8.3.1 Performance of MST Selection Techniques

In this subsection, we analyze the performance of MST selection techniques and show that the optimal strategy to minimize the MSE for the MST selection techniques is to ignore all the array-frequency domain LS estimated channel coefficients whose corresponding channel powers are smaller than the noise variance, and retain all the remaining estimated channel coefficients.

When σ_i^2 is not equal to zero

From (8.24) and (8.25), the MSE of the ith estimated channel coefficient is given by

$$\begin{aligned}\mathrm{MSE}_i &= \mathrm{E}\left[(m_i-1)^2\hat{h}_i\right]+\mathrm{E}\left[m_i^2\hat{v}_i\right]\\ &= \sigma_f^2+\int_0^\infty \frac{\hat{h}_i}{\sigma_i^2}e^{-\frac{\hat{h}_i}{\sigma_i^2}}\int_0^\eta \frac{e^{-\frac{\hat{r}_i+\hat{h}_i}{\sigma_f^2}}}{\sigma_f^2}J_0\left(\frac{2\sqrt{\hat{r}_i\hat{h}_i}}{\sigma_f^2}\right)\mathrm{d}\hat{r}_i\mathrm{d}\hat{h}_i\\ &\quad -\int_0^\infty \frac{\hat{v}_i}{\sigma_f^2}e^{-\frac{\hat{v}_i}{\sigma_f^2}}\int_0^\eta \frac{e^{-\frac{\hat{r}_i+\hat{v}_i}{\sigma_i^2}}}{\sigma_i^2}J_0\left(\frac{2\sqrt{\hat{r}_i\hat{v}_i}}{\sigma_i^2}\right)\mathrm{d}\hat{r}_i\mathrm{d}\hat{v}_i.\end{aligned}\tag{8.27}$$

Differentiating MSE_i with respect to η, we obtain

$$\frac{\partial \mathrm{MSE}_i}{\partial \eta}=\frac{(\sigma_i^2-\sigma_f^2)e^{-\frac{\eta}{\sigma_i^2+\sigma_f^2}}\eta}{(\sigma_i^2+\sigma_f^2)^2}.\tag{8.28}$$

From (8.28), we find that although the MSE is difficult to calculate analytically, its gradient is surprisingly simple.

- When $\sigma_i^2=\sigma_f^2$, the last two terms in (8.27) become identical. This is reasonable since setting the estimated channel coefficient as zero or retaining the original coarse estimate yields the same MSE. Therefore, MSE_i will always be σ_f^2 no matter what the threshold is chosen. This result can also be verified from (8.28) where $\frac{\partial \mathrm{MSE}_i}{\partial \eta}$ is always zero for all the η.

- When $\sigma_i^2>\sigma_f^2$, from (8.28) and the first derivative test [11], we conclude that MSE_i reaches its extremum when $\eta=0$ or $\eta=\infty$[4]. Since (8.28) is always positive, MSE_i is monotonically increasing and reaches its minimum σ_f^2 at $\eta=0$. This

[4]For $\sigma_i^2=\sigma_f^2$, $\frac{\partial \mathrm{MSE}_i}{\partial \eta}$ is always zero, hence indicating that the MSE is constant as claimed earlier.

result indicates that the LS estimated channel coefficient should be retained when the corresponding channel power is larger than the noise variance. It is reasonable since when we ignore the estimated channel coefficient, the resulting performance loss due to the ignorance of instantaneous power will be larger than the performance gain due to the removal of noise.

- When $\sigma_i^2 < \sigma_f^2$, (8.28) is always negative. Thus, MSE_i reaches its minimum σ_i^2 at $\eta = \infty$. This result indicates that the LS estimated channel coefficient is required to be ignored when the corresponding channel power is smaller than the noise variance. It is reasonable as the performance gain surpasses the performance loss when the corresponding estimated coefficient is ignored.

When σ_i^2 is equal to zero

In this case, (8.26) becomes

$$\text{MSE}_i = \int_{\eta}^{\infty} \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \mathrm{d}\hat{v}_i = e^{-\frac{\eta}{\sigma_f^2}} (\eta + \sigma_f^2). \tag{8.29}$$

Differentiating MSE_i with respect to η, we obtain

$$\frac{\partial \text{MSE}_i}{\partial \eta} = -\frac{\eta e^{-\frac{\eta}{\sigma_f^2}}}{\sigma_f^2} \leq 0. \tag{8.30}$$

Similar to the discussion above, we find that MSE_i reaches its minimum at $\eta = \infty$. This result indicates that the LS estimated channel coefficient is required to be ignored when the corresponding channel power is smaller than the noise variance. It is reasonable since the corresponding estimated channel coefficient only contains the noise, and ignoring this estimated coefficient will result in no performance loss.

From the discussion above, we conclude that the choice of threshold η is dependent on σ_i^2 and σ_f^2 as shown in Table. 8.1 because we need to balance the performance gain and loss when we ignore the estimated channel coefficients. Thus, the optimum threshold should be obtained for each channel coefficient. Since the channel is independent of noise, the average power of the LS estimated channel coefficient is the sum of the channel

Table 8.1: Maximum and minimum MSE_i and the corresponding thresholds for the MST selection techniques.

	$\sigma_i^2 > \sigma_f^2$	$\sigma_i^2 < \sigma_f^2$	$\sigma_i^2 = \sigma_f^2$
Maximum MSE_i	σ_i^2	σ_f^2	σ_f^2
Minimum MSE_i	σ_f^2	σ_i^2	σ_f^2
Threshold for Minimum MSE_i	0	∞	arbitrary

power and noise variance (*i.e.* $\sigma_i^2 + \sigma_f^2$). For this reason, we may set the threshold η to be $2\sigma_f^2$. When the average power of the ith LS estimated channel coefficient exceeds this threshold $2\sigma_f^2$, it means $\sigma_i^2 > \sigma_f^2$. Therefore, we set m_i to be 1 (see (8.16)) to retain the ith LS estimated channel coefficient. This is optimum to minimize the MSE_i as we discussed above. Similarly, this threshold is optimum to the minimization of MSE_i for the case when $\sigma_i^2 < \sigma_f^2$. Therefore, when the average power of the LS estimated channel coefficient is available, we can still obtain the corresponding optimal threshold.

As σ_i^2 is assumed to be not available in this chapter, the average power of the corresponding LS estimated channel coefficient may also not be available. Then, we may use the channel instantaneous power (*i.e.* the instantaneous power of the estimated channel coefficient) to approximate the average power of the LS estimated channel coefficient. Due to the monotone property of MSE_i with the increase of η, Table. 8.1 implies that for the given threshold $\eta = 2\sigma_f^2$, MSE_i is always smaller than σ_f^2 when $\sigma_i^2 < \sigma_f^2$, and larger than σ_f^2 when $\sigma_i^2 < \sigma_f^2$. Therefore, the overall performance of MST selection techniques is dependent on the portion[5] of number of channel coefficients, whose average powers are below the σ_f^2, to the total number of channel coefficients. When a majority of channel coefficients whose average powers are below σ_f^2 (such as in the angle-time domain), the MST selection technique can improve over the array-frequency domain LS technique. This is also verified in our simulation results presented in the next section.

[5] $\eta = 2\sigma_f^2$ may not be the optimum choice but reasonable when no prior information of σ_i^2 is available to the receiver. When additional information (*e.g.* the portion) is available, we may moderately adjust the threshold to improve the estimation performance as shown in the simulation results section.

8.3.2 Performance of AMMSE Technique

As m_i is dependent on the channel instantaneous power, directly analyzing the performance of AMMSE technique results in high computational complexity. Therefore, in this subsection, we only compare the AMMSE technique with the MST selection technique to provide some general understandings on the performance trends.

When σ_i^2 is not equal to zero

From (8.24) and (8.25), the MSE due to the ith estimated channel coefficient is given by

$$\begin{aligned}\mathrm{MSE}_i &= \int_0^\infty \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \int_\eta^\infty \frac{(\hat{r}_i-\sigma_f^2)^2}{\hat{r}_i^2} \frac{e^{-\frac{\hat{r}_i+\hat{v}_i}{\sigma_i^2}}}{\sigma_i^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{v}_i}}{\sigma_i^2}\right) \mathrm{d}\hat{r}_i \mathrm{d}\hat{v}_i \\ &+ \int_0^\infty \frac{\hat{h}_i}{\sigma_i^2} e^{-\frac{\hat{h}_i}{\sigma_i^2}} \int_0^\eta \frac{\sigma_f^4}{\hat{r}_i^2} \frac{e^{-\frac{\hat{r}_i+\hat{h}_i}{\sigma_f^2}}}{\sigma_f^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{h}_i}}{\sigma_f^2}\right) \mathrm{d}\hat{r}_i \mathrm{d}\hat{h}_i \\ &+ \int_0^\infty \frac{\hat{h}_i}{\sigma_i^2} e^{-\frac{\hat{h}_i}{\sigma_i^2}} \int_0^\eta \frac{e^{-\frac{\hat{r}_i+\hat{h}_i}{\sigma_f^2}}}{\sigma_f^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{h}_i}}{\sigma_f^2}\right) \mathrm{d}\hat{r}_i \mathrm{d}\hat{h}_i. \end{aligned} \tag{8.31}$$

Compared to (8.27), the difference of MSE_i, which is contributed by the ith channel coefficient, between the AMMSE technique and MST selection technique is given by

$$\begin{aligned}\mathrm{DIF}_i &= \int_0^\infty \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \int_\eta^\infty \frac{\sigma_f^4}{\hat{r}_i^2} \frac{e^{-\frac{\hat{r}_i+\hat{v}_i}{\sigma_i^2}}}{\sigma_i^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{v}_i}}{\sigma_i^2}\right) \mathrm{d}\hat{r}_i \mathrm{d}\hat{v}_i \\ &+ \int_0^\infty \frac{\hat{h}_i}{\sigma_i^2} e^{-\frac{\hat{h}_i}{\sigma_i^2}} \int_\eta^\infty \frac{\sigma_f^4}{\hat{r}_i^2} \frac{e^{-\frac{\hat{r}_i+\hat{h}_i}{\sigma_f^2}}}{\sigma_f^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{h}_i}}{\sigma_f^2}\right) \mathrm{d}\hat{r}_i \mathrm{d}\hat{h}_i \\ &-2 \int_0^\infty \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \int_\eta^\infty \frac{\sigma_f^2}{\hat{r}_i} \frac{e^{-\frac{\hat{r}_i+\hat{v}_i}{\sigma_i^2}}}{\sigma_i^2} J_0\left(\frac{2\sqrt{\hat{r}_i\hat{v}_i}}{\sigma_i^2}\right) \mathrm{d}\hat{r}_i \mathrm{d}\hat{v}_i. \end{aligned} \tag{8.32}$$

Differentiating DIF_i with respect to η, we obtain

$$\frac{\partial \mathrm{DIF}_i}{\partial \eta} = \frac{\sigma_f^4 e^{-\frac{\eta}{\sigma_i^2+\sigma_f^2}} f(\sigma_i^2, \sigma_f^2, \eta)}{(\sigma_i^2+\sigma_f^2)^3 \eta^2} \tag{8.33}$$

where

$$f(\sigma_i^2, \sigma_f^2, \eta) = 2\sigma_f^2\eta^2 + (\sigma_i^4 + 2\sigma_i^2\sigma_f^2 - \sigma_f^4)\eta - 2\sigma_i^2\sigma_f^2(\sigma_i^2+\sigma_f^2)$$

$$= 2\sigma_f^2\left(\eta - \frac{\sigma_f^4 - 2\sigma_f^2\sigma_i^2 - \sigma_i^4}{4\sigma_f^2}\right)^2 - 2\sigma_i^2\sigma_f^2(\sigma_i^2+\sigma_f^2)$$
$$-\frac{\left(\sigma_f^4 - 2\sigma_f^2\sigma_i^2 - \sigma_i^4\right)^2}{8\sigma_f^2}. \tag{8.34}$$

As the threshold η is not smaller than σ_f^2, which is larger than $(\sigma_f^4 - 2\sigma_f^2\sigma_i^2 - \sigma_i^4)/(4\sigma_f^2)$, (8.34) reaches its minimum 0 when the threshold η is equal to σ_f^2. Therefore, when the threshold η is larger than σ_f^2, (8.34) is larger than 0 (*i.e.* positive) and so does (8.33). Therefore, DIF_i is monotonically increasing with the increase of η and reaches its maximum 0 when $\eta = \infty$. Consequently, DIF_i is always negative, which implies that the AMMSE technique performs better than the MST selection technique when σ_i^2 is not equal to zero.

When σ_i^2 is equal to zero

In this case, (8.26) becomes

$$\mathrm{MSE}_i = \int_\eta^\infty m_i^2 \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \mathrm{d}\hat{v}_i < \int_\eta^\infty \frac{\hat{v}_i}{\sigma_f^2} e^{-\frac{\hat{v}_i}{\sigma_f^2}} \mathrm{d}\hat{v}_i \tag{8.35}$$

because m_i is always smaller than one. This implies that the AMMSE technique performs better than the MST selection technique when σ_i^2 is equal to zero.

Therefore, we conclude that the AMMSE technique always performs better than the MST selection technique. Since the angle-time domain MST selection technique yields better performance than the array-frequency domain LS technique when the threshold is $2\sigma_f^2$ as discussed in the previous subsection, the angle-time domain AMMSE technique should achieve further performance gain.

8.4 Simulation Results

Computer simulations are carried out for the IEEE 802.11 TGn channel models [61,112]. In the simulation, pilots from different transmit antennas are time orthogonal to each

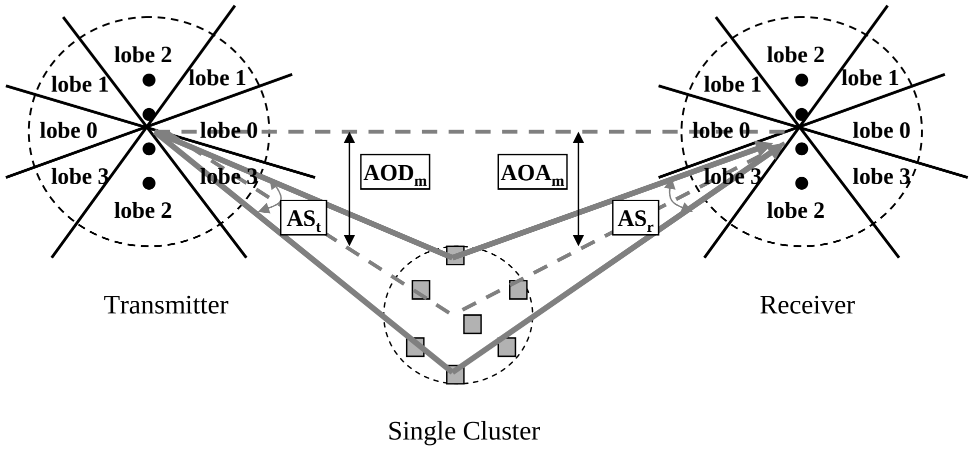

Figure 8.1: Relation of the clustered model and the angle-domain representation.

other, and thus only one transmit antenna is used to transmit pilots in each OFDM training symbol period. Further, N_c pilots are inserted at each subcarrier during the training period. Then, the total number of pilots used is $N_cN_tN_rN_d$. For simplicity, we assign the values of $+1$ and -1 to pilot symbols. We evaluated different channel estimation techniques for the five IEEE 802.11 TGn channel models that represent various indoor environments, such as residential homes and small offices. Model A corresponds to the MIMO flat-fading channel with a single cluster. This simple model serves as the basis to investigate the characteristics of cluster-based channel models in the angle domain. Here $\{\mathrm{AoD_m}, \mathrm{AS}_t\}$ and $\{\mathrm{AoA_m}, \mathrm{AS}_r\}$ refer to the mean angle of clusters, and angular spread of clusters, respectively, for the transmit and receive antennas, respectively, as illustrated in Fig. 8.1. All these four parameters are physically determined for a given propagation scenario and will affect the relative average power for each angle-time domain beam. In the simulation, we assign different values to these parameters for the model A to represent various propagation environments. This model is usually used for stressing the detection performance. It occurs only a small percentage of time in reality for the systems under consideration [61]. The models B, C, D and E correspond to 15ns, 30ns, 50ns and 100ns root-mean-square (rms) delay spread, respectively. All these four models represent non-sample-spaced channels. As we find that the performance trends in different channel estimation techniques are similar in all these four models, we only present the results for model B and model E because of their minimum and maximum degrees of freedom, respectively. In the simulations, we assign the values indicated in appendix C of [61] to $\{\mathrm{AoD_m}, \mathrm{AS}_t\}$ and $\{\mathrm{AoA_m}, \mathrm{AS}_r\}$ to represent typical small environments. We also assume $N_c = 2$, $N_d = 64$, $N_t = N_r = 4$, and the normalized separation between adjacent antennas $\Delta_t = \Delta_r = 0.5$. Further, we assume that the channel power for each link between one transmit and one receive antennas is normalized to one throughout the simulations. Note that both the line-of-sight (LOS) and NLOS scenarios have been considered in the IEEE 802.11 TGn channel models [61, 112]. Compared to the NLOS scenario, the LOS scenario has an additional fixed LOS signal component. This component can be seen as a cluster with zero AS_t and AS_r, and should have similar effect on the performance of angle-domain channel estimation techniques compared with the model A with very small AS_t and AS_r. Therefore, it is unnecessary to separately investigate this LOS component. From the above, we only provide the results for the NLOS scenario in this section as similar conclusions can be drawn from the LOS and NLOS scenarios.

8.4.1 Channel Model A

As discussed, we assign various values to the $\{\mathrm{AoD_m}, \mathrm{AS}_t\}$ and $\{\mathrm{AoA_m}, \mathrm{AS}_r\}$ in channel model A to investigate the performance of angle-domain channel estimation techniques in the presence of a single cluster. As the angular spread is usually not smaller than 40° for typical IEEE 802.11 TGn channel models [61], we consider 40° for both AS_t and AS_r as the worst case in performing angle-domain channel estimations. We also consider the angular spread to be 2°, which is valid for outdoor environments, for both AS_t and AS_r as the best case. Further, the energy of multipath components from 45° leaks into more than one angle-time domain beam, which is undesirable in the angle-domain channel estimations. Thus, we consider 45° to be the $\mathrm{AoD_m}$ and $\mathrm{AoA_m}$ of the cluster for the worse-case consideration. We also consider 0°, from which most of energy of multipath components concentrate on one angle-time domain beam, to be the $\mathrm{AoD_m}$ and $\mathrm{AoA_m}$. In the following figures that represent the channel power for each angle-time domain beam (see Fig. 8.2 and Fig. 8.4), the areas of square and circle are proportional to the angle- and array-domain normalized average power, respectively, with respect to the corresponding maximum average power in the angle domain.

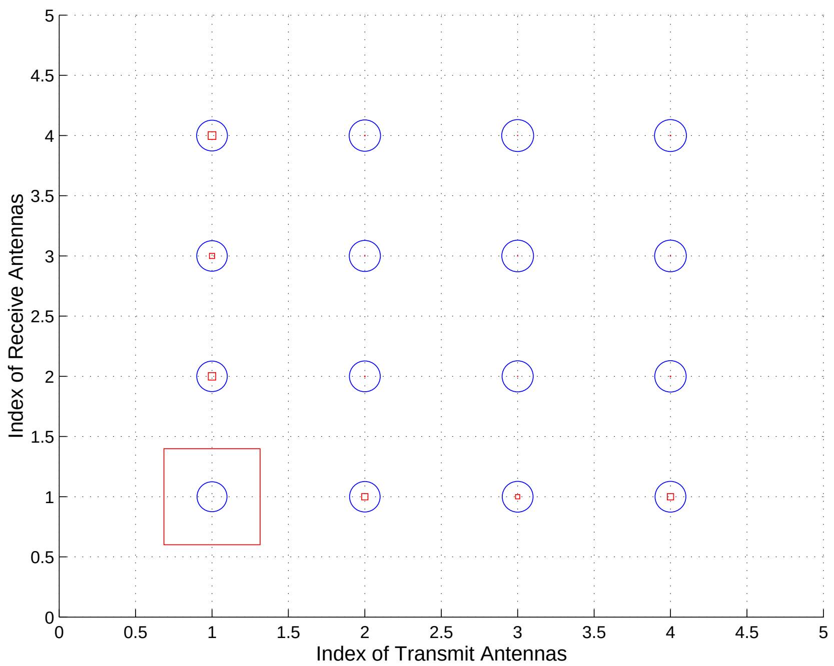

Figure 8.2: The normalized channel power for each angle-time domain beam of model A with $\mathrm{AoA_m} = 0°$, $\mathrm{AS}_t = 2°$, $\mathrm{AoA_m} = 0°$, and $\mathrm{AS}_t = 2°$.

Fig. 8.2 corresponds to the best case where the angular spread is 2°. As expected, we see that angle-domain channel power is concentrated in the (1,1)th angle-time domain beam while the array-domain channel power is uniformly distributed over all the array-time domain beams. These observations indicate the effectiveness of the MST selection and AMMSE techniques. From Fig. 8.3, it is clear that all the angle-domain techniques improve over the array-frequency domain LS technique at all the SNRs under consideration. We also find that the angle-time domain AMMSE technique improves over the angle-time domain MST selection technique by around 3.5 dB because the former is more like a LMMSE technique. Although the MST selection is a nonlinear process, it is interesting to find that the performances of angle-domain techniques are proportional to that of the array-frequency domain LS technique in Fig. 8.3 because the ignored channel instantaneous power is not significant.

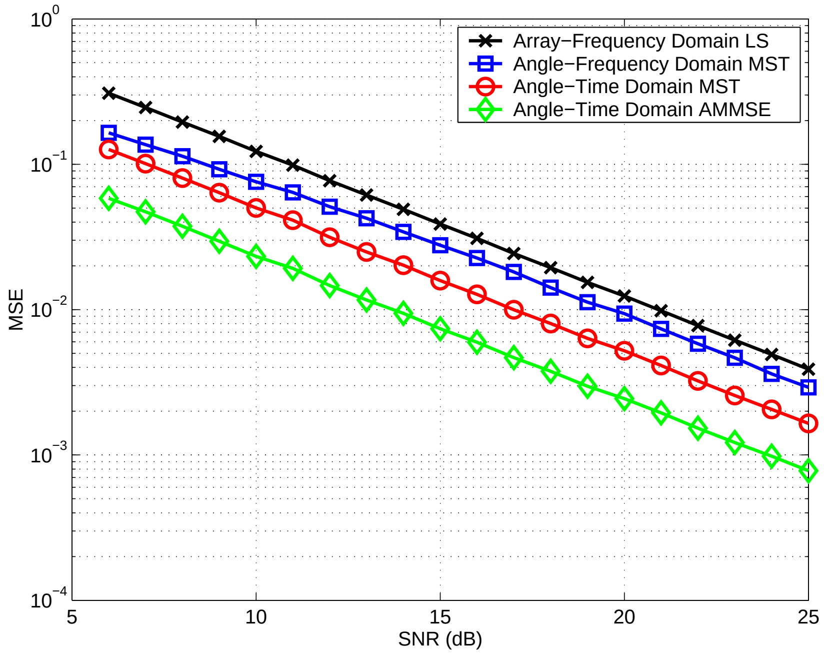

Figure 8.3: Performances of different channel estimation techniques for model A with $\mathrm{AoA_m} = 0°$, $\mathrm{AS}_t = 2°$, $\mathrm{AoA_m} = 0°$, and $\mathrm{AS}_t = 2°$.

We consider the worst case where the angular spread is 40° in Fig. 8.4 and Fig. 8.5. Compared to Fig. 8.2, the angle-domain channel power tends to leak into other angle-time domain beams. Therefore, the channel powers in all angle-time domain beams are relatively high. From Fig. 8.5, we find that all the angle-time domain techniques still outperform the array-frequency domain LS technique for all the SNRs. However, the angle-frequency domain MST selection technique does not perform well because the number of channel coefficients whose channel powers are above the noise variance is relatively large. We also observe that the angle-time domain AMMSE technique achieves the best performance as expected.

8.4.2 Typical Channel Model

We consider typical non-sample-spaced indoor channels in Fig. 8.6 and Fig. 8.7. In such cases, the $\mathrm{AoD_m}$ and $\mathrm{AoA_m}$ of clusters of multipath components tend to be uniformly

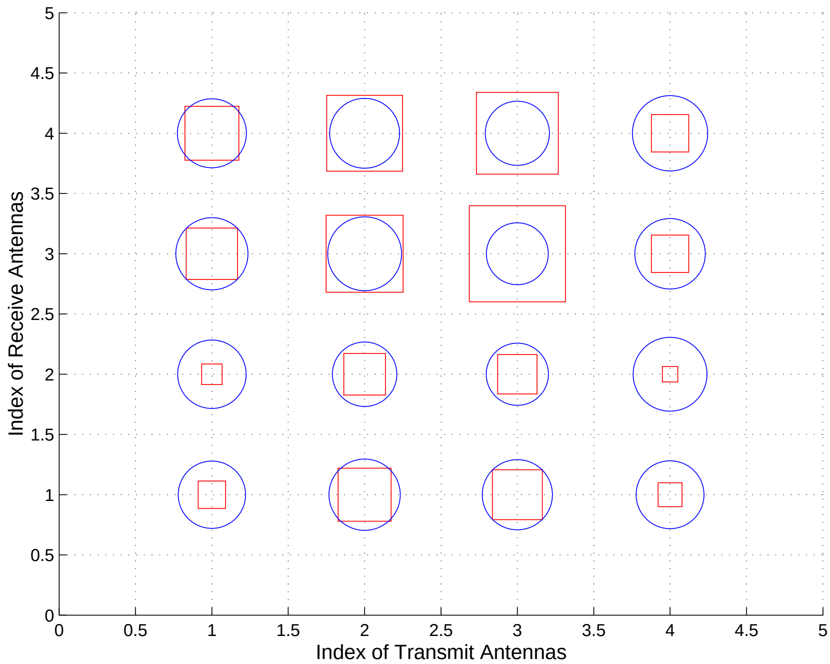

Figure 8.4: The normalized channel power of model A for each angle-time domain beam with $\mathrm{AoA_m} = 45°$, $\mathrm{AS}_t = 40°$, $\mathrm{AoA_m} = 45°$, and $\mathrm{AS}_t = 40°$.

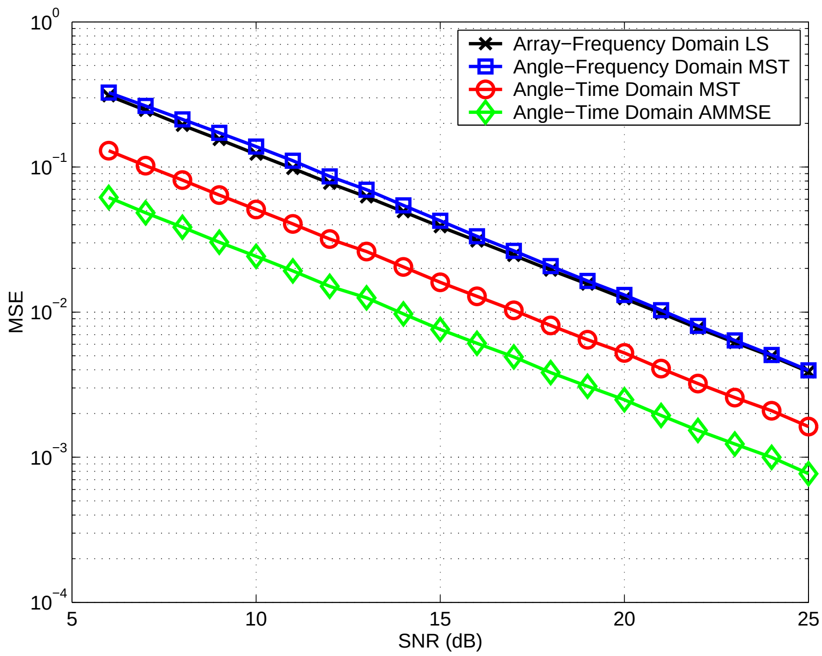

Figure 8.5: Performances of different channel estimation techniques for model A with $\mathrm{AoA_m} = 45°$, $\mathrm{AS}_t = 40°$, $\mathrm{AoA_m} = 45°$, and $\mathrm{AS}_t = 40°$.

distributed over all the angles. Then, the technique proposed in [128] will be the same as the array-frequency LS technique because all the angle-time domain beams are identified as signal beams. For this reason, its simulation results are not presented here. Compared with Fig. 8.3 and Fig. 8.5, we find that unlike in channel A, the achieved performance gain of angle-domain techniques over the array-frequency domain LS technique is not significant at high SNRs in model B and model E. This is because the ignored channel instantaneous powers are always larger in the non-sample-spaced channels. Nevertheless, the angle-time domain AMMSE technique still achieves the best performance for both model B and model E. Therefore, we can choose the angle-time domain AMMSE technique to perform channel estimation in the IEEE 802.11 TGn MIMO-OFDM systems because of its superior performance as well as its robustness.

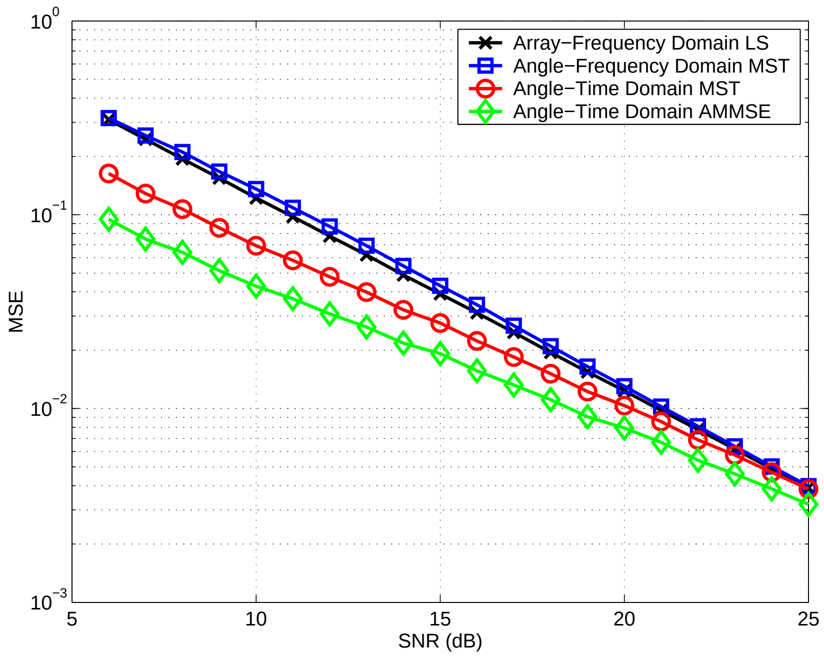

Figure 8.6: Performances of different channel estimation techniques for model B.

As the channel power tends to be concentrated over a short temporal span in the angle-time domain[6], the number of channel coefficients whose corresponding channel

[6]The temporal spans for different angle-time domain beams are not likely to be the same because the first paths falling in each beam may arrive asynchronously.

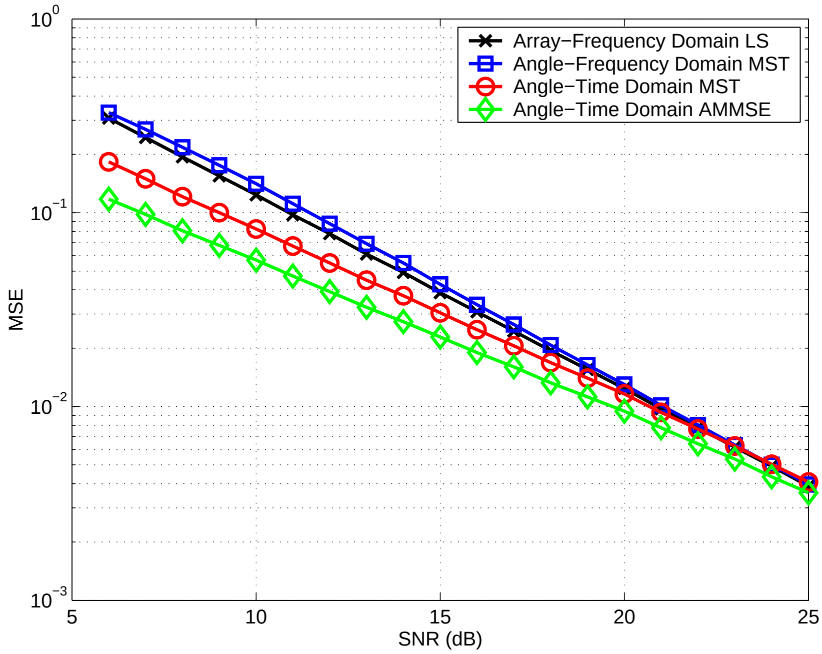

Figure 8.7: Performances of different channel estimation techniques for model E.

powers are below the noise variance is relatively large. For such channel coefficients, the optimum threshold should be infinite as discussed in Section 8.3. Intuitively, increasing the threshold will improve the estimation performance for these channel coefficients. However, this will result in adverse effect on the estimation performance for the channel coefficients whose channel powers exceed σ_f^2. As the number of channel coefficients whose channel powers are below σ_f^2 is larger compared to that of the remaining channel coefficients, it should be interesting to investigate the overall performances of the angle-time domain techniques when the threshold is increased. Fig. 8.8 shows that both the angle-time MST selection and AMMSE techniques are improved in terms of performance when the threshold is increased to $3\sigma_f^2$ at nearly all the SNRs. Note that the selection of threshold is a trade-off between the performance loss due to the ignorance of channel instantaneous power and the performance gain due to the removal of noise. Therefore, we find that further increasing the threshold to $6\sigma_f^2$ degrades the performance at high SNRs because the performance loss will dominate. Note that when the threshold is relatively high, we only retain the channel coefficients whose corresponding $[|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2 - \sigma_f^2]/[|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2]$ shown in (8.8) approaches one. This implies that the performance of the angle-time domain MST selection technique should gradually approach that of the angle-time domain AMMSE technique with the increase of thresholds. This implication is verified in Fig. 8.8. Therefore, we may use the angle-time domain MST selection technique when the threshold is relatively high for typical MIMO-OFDM systems.

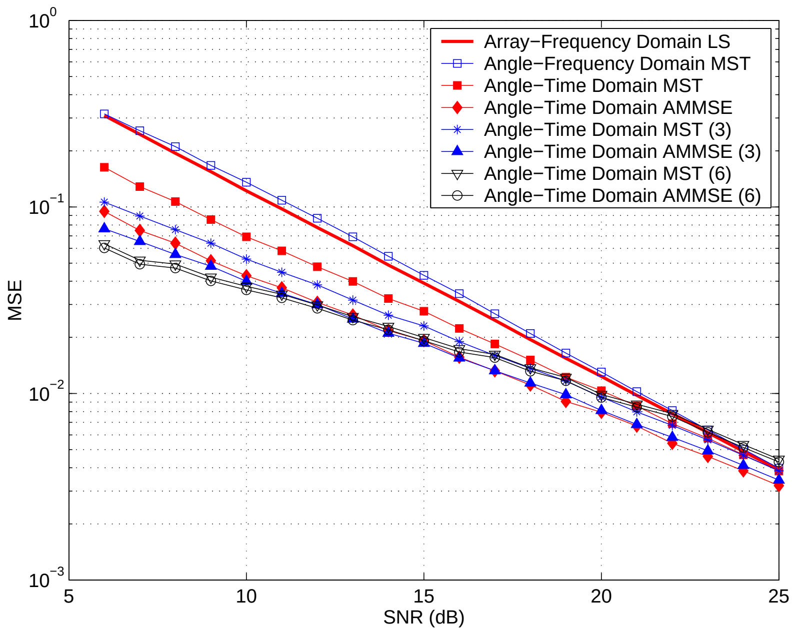

Figure 8.8: Performances of angle-domain channel estimation techniques with different thresholds for model B. The number α shown in the bracket indicates that the threshold is set to $\alpha\sigma_f^2$. Otherwise, the threshold is set to $2\sigma_f^2$.

Note that the noise variance σ_f^2 is required to be priorly known in order to decide the threshold in the above simulations. However, the exact knowledge of σ_f^2 may not always be available. Therefore, for a robust estimator design, we should fix the threshold for a target range of SNRs. As illustrated in Fig. 8.8, increasing the threshold will even improve the performance of angle-time domain techniques especially at low SNRs. Therefore, for a given target SNR range, we may use $2\sigma_f^2$ at the lowest SNR (because of the largest σ_f^2) as the fixed threshold. In Fig. 8.9, we divide the whole SNR range into five disjoint groups. Each corresponds to one target SNR range, within which the threshold is fixed. As Fig. 8.8 indicates that the threshold should be smaller (or at

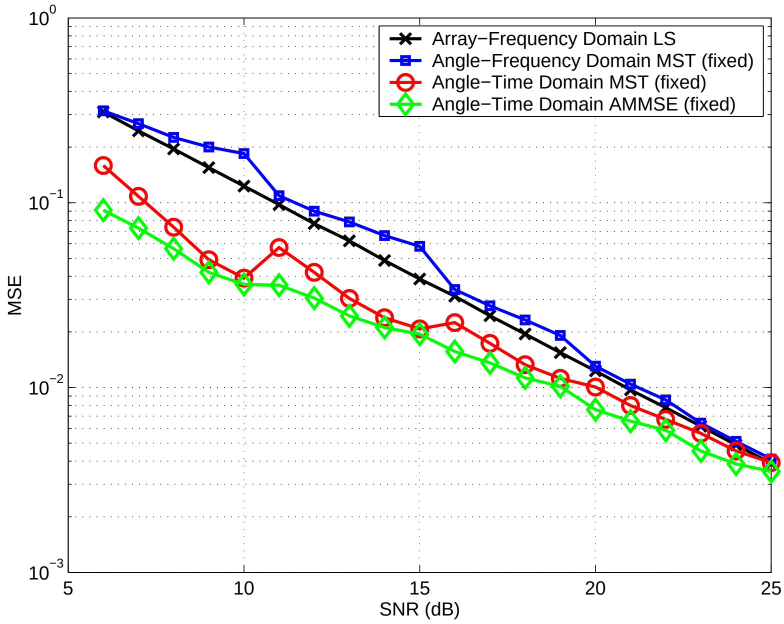

Figure 8.9: Performances of angle-domain channel estimation techniques for model B. Fixed in the bracket indicates that the threshold is fixed for a given SNR range.

most slightly larger) than the corresponding $6\sigma_f^2$ at low SNRs, and smaller than the corresponding $3\sigma_f^2$ at high SNRs. These results imply that we could choose larger α (thus more elements for each group) at low SNRs than at high SNRs. From Fig. 8.9, we observe that the performances of angle-time domain estimation techniques are always better than that of the array-frequency domain LS technique. Furthermore, we find that the performance of the angle-time domain MST selection technique is close to that of the angle-time domain AMMSE technique at the highest SNR in each target range (see the points when SNR = 10 dB, 15 dB, 19 dB, 22 dB and 25 dB in Fig. 8.9). These results are consistent with the observations in Fig. 8.8, which shows that increasing the threshold makes the performance of the angle-time domain MST selection technique approach that of the angle-time domain AMMSE technique. However, the angle-frequency domain MST selection technique may not perform well at the highest SNR for each target range because of the relatively large performance loss due to the channel instantaneous power ignorance. As the angle-time domain AMMSE technique performs best at all the SNRs under consideration, we may conclude that the angle-time domain AMMSE technique is suitable for the typical IEEE 802.11 TGn MIMO-OFDM systems when the target range of SNRs is available.

Note that at higher SNRs, the threshold η approaches zero and thus almost all the channel coefficients will not be filtered. Therefore, the angle-time domain MST selection and angle-frequency domain MST selection techniques will perform more similarly to the array-frequency domain LS technique as SNR increases. In addition, the angle-time AMMSE technique will also perform more similarly to the angle-time domain MST selection technique as SNR increases because the multiplication factor shown in (8.8) approaches to one for each angle-time domain channel coefficient. In summary, all the channel estimation techniques will converge at a relatively high SNR as implied in all the figures shown in this subsection. Thus, the angle-time domain AMMSE technique is more preferable at lower SNRs than at higher SNRs because of the larger achievable performance gain.

8.5 Conclusions

In this chapter, we have proposed the angle-frequency domain MST selection technique, angle-time domain MST selection technique and angle-time domain AMMSE technique for MIMO-OFDM systems. These three techniques do not require the prior knowledge of channel correlation and were shown to be effective when the angular spread of clusters of multipath components is small. More importantly, the angle-time domain techniques can improve over the array-frequency domain LS technique in all the cases considered even when the angular spread is relatively large. Further, both our theoretical analysis and simulation results indicate that the angle-time domain AMMSE technique achieves the best performance and is robust to the choice of threshold and mismatch of operating SNR. Thus, when only the target SNR range is available to the receiver, the angle-time domain AMMSE technique is suitable for the typical IEEE 802.11 TGn MIMO-OFDM systems. In addition, we have also found that with a suitable threshold and known operating SNR, the angle-time domain MST selection technique results in little performance degradation compared to the angle-time domain AMMSE technique. Therefore, in such cases, the angle-time domain MST selection technique may be a potential candidate for the IEEE 802.11 TGn MIMO-OFDM systems because of its lower computational complexity.

Chapter 9

LMMSE-Based Angle-Domain Channel Estimation

9.1 Introduction

Chapter 8 has systematically developed channel instantaneous power based angle-domain channel estimation techniques under the assumption that the channel correlation is not available to the receiver. This chapter investigates angle-domain channel estimation techniques for the situation when the channel correlation or channel power is known to the receiver. Further, various channel estimation techniques for multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems are compared in terms of performance and complexity in this chapter.

As before, we classify the signals and channels in two domain types. The first type is represented by either the array or angle domain, the latter one is represented by either the time or frequency domain. The channel correlations in the two domain types are referred to as the channel spatial correlation and channel frequency correlation, respectively. Additionally, we use the term *channel correlation* for both the channel spatial and frequency correlation. Prior knowledge of these channel correlations can be exploited in channel estimation. As prior knowledge of the channel frequency correlation is equivalent to prior knowledge of the channel power delay profile, we will not consider prior knowledge of the channel power delay proile separately in channel estimation. Hereinafter, we will explicitly state which representation is used for each domain type. For example, the designation angle-time domain means that the angle and time representations are used for the above two domain types, respectively. In this chapter, the group of angle-time domain bins that correspond to the identical angular lobes (but with different time indices) is called the angle-time domain *beam*. Similarly, the group of angle-frequency domain bins that correspond to the identical angular lobes (but with different subcarrier indices) is called the angle-frequency domain beam. Note that when we state the representation for only one domain type, we mean that both representations are applicable for the other domain type unless indicated otherwise. For example, we refer to the angle domain as either the angle-time or angle-frequency domain. Similarly, we refer to the angle-domain beam as either the angle-time or angle-frequency domain beam.

In this chapter, we focus on channel estimation techniques in both the angle-time and angle-frequency domain. Note that we ultimately need the knowledge of array-frequency domain channel coefficients to realize the coherent demodulation. Thus, the estimated channel coefficients in the angle domain will finally be transformed back to the array-frequency domain.

- Techniques in the angle-frequency domain: As the channel coefficients in different angle-frequency domain beams are approximately spatially uncorrelated, we can divide the 2D channel correlation matrix with the size $(N_tN_rN_d \times N_tN_rN_d)$ into (N_tN_r) one-dimensional (1D) channel correlation submatrices each having the size $(N_d \times N_d)$, where N_t and N_d are the number of transmit and receive antennas, respectively, N_d is the number of subcarriers. Each channel correlation submatrix corresponds to one angle-frequency domain beam. Thus, we can use 1D LMMSE channel estimation techniques for each angle-frequency domain beam. We call this technique the quasi 1D (Q1D) LMMSE technique. This division of the 2D channel correlation matrix into a set of 1D channel correlation submatrices will greatly reduce the complexity in channel estimation. More importantly, our simulation results show that this complexity reduction will result in negligible performance degradation for typical channel models. Since the angle-time and angle-frequency domains are related by a unitary transformation, the 2D channel correlation in the angle-time and angle-frequency is interchangeable. Then, the performances of 2D LMMSE based on either angle-time and angle-frequency domain channel correlation are equivalent. Further, the 1D angle-time and angle-frequency channel correlation within the same angle-domain beam is also interchangeable. Then, the performances of Q1D LMMSE based on either angle-time and angle-frequency domain channel correlation are also equivalent. Therefore, we do not consider the 2D LMMSE and Q1D LMMSE techniques in the angle-time domain in this chapter.

- Techniques in the angle-time domain: As the channel coefficients in all the angle-time domain bins are approximately uncorrelated, we may use the channel power instead of the channel correlation to approximately apply the angle-time domain Q1D LMMSE technique. We call the resulting technique the approximated LMMSE (AMMSE) technique. Note that we refer to the *channel power* as the channel average power in this chapter. When the channel power is not available, we can use the *channel instantaneous power* (*i.e.* the instantaneous power of estimated channel coefficients) to estimate the channel power as shown in Chapter 8. In such cases, to maintain the estimated channel power positive and estimation reliable, a threshold is required to ignore the coefficients with low instantaneous power. Note that the AMMSE technique is not considered in the angle-frequency domain because the channel correlation within each angle-frequency domain beam might be too high to be reasonably replaced by the channel power.

To the best of our knowledge, the work of this chapter is the first to systematically investigate LMMSE-based angle-domain channel estimation techniques for MIMO-OFDM systems when the knowledge of either the channel correlation or power is available to the receiver [96]. Angle-domain techniques when both the channel correlation and power are unavailable to the receiver are described in [97]. The angle-domain LMMSE-based channel estimation techniques proposed in this chapter have two main advantages. First, compared to the 2D LMMSE technique, they can achieve significant complexity reduction while still maintaining good estimation performance for typical MIMO-OFDM channel models. Second, they are flexible in implementation in the angle domain as explained in Section 8.1.

The major contributions and results of this chapter are as follows:

1. We systematically develop reduced-complexity LMMSE-based channel estimation techniques in both the angle-time and angle-frequency domain for MIMO-OFDM systems. The choice of LMMSE-based techniques is largely dependent on the extent of channel stochastic information (*e.g.* channel correlation or power) available to the receiver.

2. We analyze the performances and complexity of different channel estimation techniques. Our simulation results show that the Q1D LMMSE technique that is based on the angle-frequency domain correlation can achieve similar performance as the 2D LMMSE technique for typical MIMO-OFDM channel models, but has significantly lower complexity.

3. We find that the AMMSE technique has the lowest complexity of all the LMMSE-based techniques. Moreover, the channel power based AMMSE technique yields significant performance gain while maintaining comparable complexity compared to the LS technique for all the MIMO-OFDM channel models under consideration.

This chapter is organized as follows. In Section 9.2, we introduce and compare various channel estimation techniques to estimate MIMO-OFDM channels. Then, we evaluate the performances of different techniques in Section 9.3 by simulating typical IEEE 802.11 TGn channel models. Finally, we conclude the chapter in Section 9.4.

9.2 Channel Estimation for MIMO-OFDM

For channel estimation techniques, one of the most important performance measures is the mean square error (MSE), which measures the average mean squared deviation of the estimator from the true value [110]. Thus, like the previous chapter, we use the MSE to compare the performance of different techniques in this chapter. Note that the array-time or array-frequency domain is related by a unitary transformation. Thus, of a given estimation technique, the MSE represented in either the array-time or array-frequency domain yields the same result, and is given by

$$\text{MSE} = \frac{1}{N_d N_t N_r} \text{E}\left[||\underline{\tilde{\mathbf{h}}} - \underline{\vec{\mathbf{h}}}||^2 \right] \tag{9.1}$$

where $\underline{\vec{\mathbf{h}}}$ is either $\underline{\mathbf{h}}$ or $\underline{\mathbf{c}}$, and $\underline{\tilde{\mathbf{h}}}$ is the estimated $\underline{\vec{\mathbf{h}}}$.

9.2.1 LS Technique

In MIMO-OFDM systems, there exist four domains, *i.e.* the angle-time, angle-frequency, array-time, and array-frequency domain. As the noise representing in any domain is related by a unitary transformation, the performance of LS technique in any of the four domains should be the same. Further, we ultimately need the knowledge of array-frequency domain channel coefficients to realize the coherent demodulation. Thus, we only illustrate the array-frequency domain LS technique here. Recalling (8.3), we have

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}}\ \underline{\mathbf{h}} + \underline{\boldsymbol{\vartheta}} \tag{9.2}$$

under the assumption that the pilots from different transmit antennas are time orthogonal to each other, where $\underline{\mathbf{Y}} = [\underline{\ddot{\mathbf{Y}}}^T(n),\ \underline{\ddot{\mathbf{Y}}}^T(n+N_t),\ \cdots,\ \underline{\ddot{\mathbf{Y}}}^T(n+(N_c-1)N_t)]^T$ with $\underline{\ddot{\mathbf{Y}}}(n) = [\mathbf{y}^T(0,n),\ \mathbf{y}^T(0,n+1),\ \cdots,\ \mathbf{y}^T(0,n+N_t-1),\ \mathbf{y}^T(1,n),\ \cdots,\ \mathbf{y}^T(1,n+N_t-1),$ $\cdots,\ \mathbf{y}^T(N_d-1,n+N_t-1)]^T$, $\underline{\mathbf{X}} = [\underline{\ddot{\mathbf{X}}}(n), \underline{\ddot{\mathbf{X}}}(n+N_t), \cdots, \underline{\ddot{\mathbf{X}}}(n+(N_c-1)N_t)]^T$ with $\underline{\ddot{\mathbf{X}}}(n)$ $= \text{diag}[\mathbf{x}^T(0,n),\ \mathbf{x}^T(0,n+1),\ \cdots,\ \mathbf{x}^T(0,n+N_t-1),\ \mathbf{x}^T(1,n),\ \cdots,\ \mathbf{x}^T(1,n+N_t-1),\ \cdots,$ $\mathbf{x}^T(N_d-1,n+N_t-1)]$, and $\underline{\boldsymbol{\vartheta}} = [\underline{\ddot{\boldsymbol{\vartheta}}}^T(n), \underline{\ddot{\boldsymbol{\vartheta}}}^T(n+N_t), \cdots, \underline{\ddot{\boldsymbol{\vartheta}}}^T(n+(N_c-1)N_t)]^T$ with $\underline{\ddot{\boldsymbol{\vartheta}}}^T(n) = [\boldsymbol{\vartheta}^T(0,n),\ \boldsymbol{\vartheta}^T(0,n+1),\ \cdots,\ \boldsymbol{\vartheta}^T(0,n+N_t-1),\ \boldsymbol{\vartheta}^T(1,n),\ \cdots,\ \boldsymbol{\vartheta}^T(1,n+N_t-1),$ $\cdots,\ \boldsymbol{\vartheta}^T(N_d-1,n+N_t-1)]^T$ are the received signal vector, transmitted signal matrix, and noise vector, respectively, $\mathbf{y}(k,n) = [y_0(k,n),\ y_1(k,n),\ \cdots,\ y_{N_r-1}(k,n)]^T$, $\mathbf{x}(k,n) =$ $[x_0(k,n),\ x_1(k,n),\ \cdots,\ x_{N_t-1}(k,n)]^T$, $\boldsymbol{\vartheta}(k,n) = [\vartheta_0(k,n),\ \vartheta_1(k,n),\ \cdots,\ \vartheta_{N_r-1}(k,n)]^T$, $x_1(k,n),\ \cdots,\ x_{N_t-1}(k,n)]^T$, $\boldsymbol{\vartheta}(k,n) = [\vartheta_0(k,n),\ \vartheta_1(k,n),\ \cdots,\ \vartheta_{N_r-1}(k,n)]^T$, $x_{i_t}(k,n)$, $y_{i_r}(k,n)$ and $\vartheta_{i_r}(k,n)$ represent the frequency-array domain transmitted data sample at the (i_t)th transmitter, the received data sample at the (i_r)th receiver, and the additive white Gaussian noise (AWGN) with variance σ_f^2 at the (i_r)th receiver, respectively, at the kth subcarrier in the nth OFDM symbol, $\underline{\mathbf{h}} = [\mathbf{h}^T(0), \mathbf{h}^T(1), \cdots, \mathbf{h}^T(N_d - 1)]^T$, $\mathbf{h}(k)$ is the vector that stacks the columns of $\mathbf{H}(k)$, which is the channel transfer function (CTF) matrix at the kth subcarrier defined in (7.5), N_t and N_r are the numbers of the transmit and receive antennas, respectively, and N_c is the number of pilots used for each channel coefficient in the LS channel estimation.

In this case, the array-frequency domain LS estimator is given by

$$\underline{\mathbf{h}}_{\text{LS}} = \left(\underline{\mathbf{X}}^H \underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}^H \underline{\mathbf{Y}}. \tag{9.3}$$

By rearranging the vector form $\underline{\mathbf{h}}_{\text{LS}}$ into its matrix form, we get the LS estimated array-frequency domain channel matrix $\tilde{\mathbf{H}}(k)$ for $k = 0, 1, \cdots, N_d - 1$. Let $\beta = \mathrm{E}(|x_{i_t}(k,n)|^2)E(1/|x_{i_t}(k,n)|^2)/N_c$ and the signal-to-noise ratio (SNR) $= \mathrm{E}(|x_{i_t}(k,n)|^2)/\sigma_f^2$, where $x_{i_t}(k,n)$ is the transmitted signal at the kth subcarrier in the nth OFDM symbol from the i_tth transmit antenna. Then, the mean square error (MSE) per subcarrier is given by

$$\mathrm{MSE}_{\text{LS}} = \frac{\beta}{\mathrm{SNR}}. \tag{9.4}$$

The angle-frequency domain LS estimator is easily obtained by using the linear transformation (*e.g.* from (7.26)) of the obtained array-frequency domain LS estimator, or directly computing (9.3), in which the array-frequency domain transmitted and received signals should be replaced by the angle-frequency domain ones.

9.2.2 2D LMMSE Technique

When both the channel spatial and the channel frequency correlations (*i.e.* the 2D channel correlation) are known to the receiver, we can use the 2D LMMSE technique to realize the channel estimation. Note that because of the unitary transformations between the four domains, the 2D channel correlation in any domain is interchangeable and will result in the same performance for the 2D LMMSE technique. Further, we ultimately need the knowledge of array-frequency domain channel to realize the coherent demodulation. Therefore, we only consider the 2D LMMSE channel estimator in the array-frequency domain, which is given by

$$\underline{\mathbf{h}}_{\text{2D LMMSE}} = \mathbf{R}\left[\mathbf{R} + \sigma_f^2\left(\underline{\mathbf{X}}^H\underline{\mathbf{X}}\right)^{-1}\right]^{-1}\underline{\mathbf{h}}_{\text{LS}} \tag{9.5}$$

where $\mathbf{R} = \mathrm{E}\{\underline{\mathbf{h}}\,\underline{\mathbf{h}}^H\}$ is the array-frequency domain 2D channel correlation. (9.5) clearly shows that 2D LMMSE technique can be obtained from the array-frequency domain LS estimator through a linear transformation. Further, the singular value decomposition (SVD) of $\mathbf{R}$ gives

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \tag{9.6}$$

where $\mathbf{U}$ is a unitary matrix containing the eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N_dN_tN_r-1}$ on its diagonal. If $(\underline{\mathbf{X}}^H\underline{\mathbf{X}})^{-1}$ is replaced with its expectation, the average MMSE in the estimator per subcarrier becomes

$$\text{MMSE}_{\text{2D LMMSE}} = \frac{\beta}{N_dN_tN_r\text{SNR}}\sum_{ii=0}^{N_dN_tN_r-1}\frac{\lambda_{ii}}{\lambda_{ii}+\frac{\beta}{\text{SNR}}}. \tag{9.7}$$

Compared with (9.4), (9.7) clearly shows that the 2D LMMSE technique improves over the LS technique in terms of MSE since $\lambda_{ii}/[\lambda_{ii}+\beta/(\text{SNR})]$ is always not larger than 1. However, the complexity of 2D LMMSE technique is too high to allow implementation. For example, when $N_t = 4$, $N_r = 4$ and $N_d = 64$, the total number of channel parameters to be estimated is 1024. Then, the size of 2D channel correlation matrix becomes 1024×1024. This large size makes the matrix inversion shown in (9.5) highly complex (the complexity is $O(1024^{2.376})$ for the Coppersmith-Winograd algorithm [50]), and thus results in considerable complexity for the 2D LMMSE technique since the matrix inversion is needed every time when $\underline{\mathbf{X}}$ changes. When $\underline{\mathbf{X}}$ does not change, the matrix $\mathbf{R}[\mathbf{R}+\sigma_f^2(\underline{\mathbf{X}}^H\underline{\mathbf{X}})^{-1}]^{-1}$ needs to be computed only once. Then, from (9.5), the 2D LMMSE technique still requires $N_tN_rN_d$ complex multiplications for each channel coefficient when $\underline{\mathbf{h}}_{\text{LS}}$ is available. A possible solution to reduce the complexity is the low-rank realization of LMMSE technique by the use of SVD.

9.2.3 2D SVD Technique

In [57], the SVD technique was proposed for the low-rank realization of the LMMSE technique for SISO-OFDM systems. As $\underline{\mathbf{h}}_{\mathrm{LS}}$ is also a vector, the technique proposed in [57] is easily extended to MIMO-OFDM systems by first replacing $(\underline{\mathbf{X}}^H\underline{\mathbf{X}})^{-1}$ with $E[(\underline{\mathbf{X}}^H\underline{\mathbf{X}})^{-1}]$ in (9.5) to avoid matrix inversion each time. Then, applying SVD on $\mathbf{R}$, and ignoring all the eigenvalues except the largest N_h' eigenvalues, the resulting channel estimator becomes

$$\underline{\mathbf{h}}_{\mathrm{2D\ SVD}} = \mathbf{U}_1\mathbf{\Lambda}'\left(\mathbf{\Lambda}' + \sigma_f^2\mathbf{I}\right)^{-1}\mathbf{U}_1^H\underline{\mathbf{h}}_{\mathrm{LS}} \tag{9.8}$$

where $\mathbf{U}_1$ contains the first N_h' columns of $\mathbf{U}$, and $\mathbf{\Lambda}'$ is the upper left $N_h' \times N_h'$ submatrix of $\mathbf{\Lambda}$. Usually, N_h' is chosen such that $\lambda_{N_h'}$ is the first eigenvalue that is smaller than the noise power. It is clear from (9.8) that the 2D SVD technique results in lower computational complexity[1] compared with the 2D LMMSE technique since the matrix inversion to be done is for diagonal matrices and the size is reduced to $\tilde{N}_h \times \tilde{N}_h$. Further, when $\tilde{N}_h - 1$ is chosen such that $\lambda_k = 0$ for $k \geq \tilde{N}_h$, the performance is the same as the 2D LMMSE technique. If $\tilde{N}_h - 1$ is not correctly chosen such that $\lambda_k \neq 0$ when $k \geq \tilde{N}_h$, the extra average MMSE will be

$$\mathrm{MMSE}_{\mathrm{2D\ SVD,\ res}} = \frac{1}{N_dN_tN_r}\sum_{l=\tilde{N}_h}^{N_dN_tN_r-1}\lambda_l. \tag{9.9}$$

As $\mathbf{\Lambda}' + \sigma_f^2\mathbf{I}$ is a diagonal matrix, its inversion will not result in significant complexity. Further, it has been shown in [57] that the 2D SVD technique requires $2\tilde{N}_h$ multiplications for each channel coefficient if the LS estimated channel coefficients are available. Thus, when $2\tilde{N}_h$ is significantly smaller than $N_tN_rN_d$, the complexity reduction is considerable compared to the 2D LMMSE technique.

[1]For the 2D SVD to be computationally simple, the channel correlation matrix $\mathbf{R}$ should be fixed. This is easily satisfied, for example, in indoor channels.

9.2.4 Q1D LMMSE Technique

When N_t and N_r are large, the complexity of above two 2D techniques is still quite high. One possible technique to reduce the complexity is to partition all the channel coefficients to be estimated into a set of blocks with reasonable sizes and perform channel estimation independently in these blocks at the price of a certain performance loss. Recall that unlike in the conventional array domain, the channel coefficients in different angle-frequency domain beams can be approximated as spatially uncorrelated. Thus, we may divide the channel coefficients into N_tN_r blocks, each having size $(N_d \times N_d)$ and corresponding to one angle-frequency domain beam. Then, the corresponding Q1D LMMSE channel estimator is given by

$$\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{Q1D\ LMMSE},i_{tr}} = \mathbf{R}^{\mathrm{a}}_{i_{tr}} \left[\mathbf{R}^{\mathrm{a}}_{i_{tr}} + \sigma_f^2 \left((\underline{\mathbf{X}}^{\mathrm{a}}_{i_{tr}})^H \underline{\mathbf{X}}^{\mathrm{a}}_{i_{tr}}\right)^{-1}\right]^{-1} \underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS},i_{tr}} \tag{9.10}$$

for $i_{tr} = 0, 1, \cdots, N_tN_r - 1$, where $\mathbf{R}^{\mathrm{a}}_{i_{tr}}$, $\underline{\mathbf{X}}^{\mathrm{a}}_{i_{tr}}$, and $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS},i_{tr}}$ are the angle-frequency domain channel correlation matrix, transmitted signal matrix, LS estimated angle-frequency domain channel vector, respectively, for the (i_{tr})th block. $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS},i_{tr}}$ can be obtained from $\underline{\mathbf{h}}_{\mathrm{LS}}$ by the use of (7.26). Then, again from (7.26), we can finally obtain the corresponding array-frequency domain channel estimator when all the $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{Q1D\ LMMSE},i_{tr}}$ become available. As the size of channel correlation matrix is reduced to $(N_d \times N_d)$, the complexity of matrix inversion is significantly reduced. However, this complexity reduction may result in performance degradation because of the ignorance of channel correlation between each angle-frequency domain beam. Let $\bar{\mathbf{R}}$ represent the revised version of $\mathbf{R}$ such that $\bar{\mathbf{R}}$ ignores the channel correlation between each angle-frequency domain beam. Then, the extra average MMSE is given by

$$\begin{aligned}&\mathrm{MMSE}_{\mathrm{Q1D\ LMMSE,\ res}}\\ =\ &\frac{1}{N_dN_tN_r}\mathrm{trace}\Big[(\bar{\mathbf{S}}\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}\bar{\mathbf{S}} + \mathbf{\Lambda}\mathbf{M}_2^{-1}\mathbf{\Lambda} - \bar{\mathbf{S}}\mathbf{M}_1^{-1}\mathbf{\Lambda} - \mathbf{\Lambda}\mathbf{M}_1^{-1}\bar{\mathbf{S}})\Big]\end{aligned} \tag{9.11}$$

where

$$\mathbf{M}_1 = \bar{\mathbf{S}} + \frac{\beta}{\mathrm{SNR}}\mathbf{I},\ \mathbf{M}_2 = \mathbf{\Lambda} + \frac{\beta}{\mathrm{SNR}}\mathbf{I},$$

and

$$\bar{\mathbf{S}} = \mathbf{U}^H\bar{\mathbf{R}}\mathbf{U}.$$

(9.11) clearly shows that the mismatch of channel correlation may result in extra MMSE. Nevertheless, as shown in our simulation results, this mismatch will not result in significant performance degradation for typical MIMO-OFDM channel models.

Further, when $\underline{\mathbf{X}}$ does not change and $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS},i_{tr}}$ is available, from (9.10), obtaining the Q1D LMMSE angle-frequency domain channel estimator requires only N_d complex multiplications for each channel coefficient because the size of $\mathbf{R}^{\mathrm{a}}_{i_{tr}}[\mathbf{R}^{\mathrm{a}}_{i_{tr}}+\sigma_f^2((\underline{\mathbf{X}}^{\mathrm{a}}_{i_{tr}})^H\underline{\mathbf{X}}^{\mathrm{a}}_{i_{tr}})^{-1}]^{-1}$ becomes $N_d \times N_d$. To make a fair complexity comparison among all the techniques, we assume that LS estimated array-frequency domain channel coefficients are available beforehand. Then, obtaining the angle-frequency domain $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{LS},i_{tr}}$ from the array-frequency domain $\underline{\mathbf{h}}_{\mathrm{LS}}$ requires N_t+N_r complex multiplications for each channel coefficient. Further, transforming the estimated angle-frequency domain $\underline{\mathbf{h}}^{\mathrm{a}}_{\mathrm{Q1D\ LMMSE},i_{tr}}$ into the array-frequency domain also requires N_t+N_r complex multiplications for each channel coefficient. Therefore, the total required number of complex multiplications for each channel coefficient is $N_d+2(N_t+N_r)$. This will result in significant complexity reduction for large N_t and N_r compared to the 2D LMMSE technique.

9.2.5 Channel Power Based AMMSE Technique

In the time domain, the channel coefficients in different time indices are uncorrelated for the sample-spaced channels. Adding that the channel coefficients are approximately spatially uncorrelated in different angle-time domain beams, the channel correlation can be approximated as the channel power in the angle-time domain. In this case, the approximated channel correlation matrix becomes a diagonal matrix. Thus, the joint filtering of all the channel coefficients in the LMMSE technique becomes independent filtering for each channel coefficient, which significantly reduces the computational complexity. When the channel power is known to the receiver, the angle-time domain AMMSE channel estimator is given by

$$\tilde{c}^{\mathrm{a}}_{i_r,i_t,\mathrm{AMMSE}}(l) = \frac{\sigma^2_{i_{rtl}}}{\sigma^2_{i_{rtl}}+\frac{\beta}{\mathrm{SNR}}}\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l) \tag{9.12}$$

where $\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)$ is the LS estimated angle-time domain channel coefficient of the (i_r, i_t)th angle-time domain beam at the delay l, and $\sigma^2_{i_r t l}$ is the power of the angle-time domain channel coefficient $c^{\mathrm{a}}_{i_r,i_t}(l)$. Let $\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)$ (for $l = 0, 1, \cdots, N_d$) form the channel vector $\tilde{\mathbf{c}}^{\mathrm{a}}_{i_r,i_t}$, which consists of all the channel coefficients of the (i_r, i_t)th angle-time domain beam. Then, $\tilde{\mathbf{c}}^{\mathrm{a}}_{i_r,i_t}$ can be obtained by

$$\tilde{\mathbf{c}}^{\mathrm{a}}_{i_r,i_t} = \mathbf{F}\tilde{\mathbf{h}}^{\mathrm{a}}_{i_r,i_t} \tag{9.13}$$

where $\mathbf{F}$ is the $N_d \times N_d$ unitary Fourier matrix, and $\tilde{\mathbf{h}}^{\mathrm{a}}_{i_r,i_t}$ is the channel vector for the (i_r, i_t)th angle-frequency domain beam; $\tilde{\mathbf{h}}^{\mathrm{a}}_{i_r,i_t}$ can be obtained from $\underline{\mathbf{h}}_{\mathrm{LS}}$ via (7.26). Compared to the Q1D LMMSE technique, this channel power based AMMSE technique should have wider application because only the channel power is required to be *a priori* known. Further, the complexity of AMMSE technique is much lower. The penalty paid is a slight performance degradation due to the further mismatch of channel correlation. The AMMSE technique is well suited for clustered-based channel models, and will not result in any performance degradation compared to the LS technique for non-clustered-based channel models. Therefore, it is applicable in any MIMO-OFDM system.

Next, we begin to analyze the performance of channel power based AMMSE technique. Here we use the LS technique to coarsely estimate the angle-time domain channel vector. Let $\underline{\mathbf{c}}^{\mathrm{a}}$ represent the stacked actual angle-time domain channel vector, then from (7.20), we obtain

$$\underline{\mathbf{c}}^{\mathrm{a}} = \mathbf{B}\underline{\mathbf{c}} \tag{9.14}$$

where

$$\mathbf{B} = \mathbf{I}_{N_d} \otimes \mathbf{U}_t^T \otimes \mathbf{U}_r^H \tag{9.15}$$

is an $(N_t N_r N_d \times N_t N_r N_d)$ matrix and the superscript 'T' denotes the transpose. It is easily verified that matrix $\mathbf{B}$ is unitary, *i.e.*

$$\mathbf{B}^H\mathbf{B} = \mathbf{I}_{N_t \times N_r \times N_d}. \tag{9.16}$$

Then, the AMMSE channel estimator is given by

$$\underline{\mathbf{c}}^{\mathrm{a}}_{\mathrm{AMMSE}} = \mathbf{M}\underline{\mathbf{c}}^{\mathrm{a}}_{\mathrm{LS}} \tag{9.17}$$

where $\underline{\mathbf{c}}_{\text{LS}}^{\text{a}}$ is the stacked LS estimated angle-time domain channel vector, and $\mathbf{M}$ is a diagonal $(N_tN_rN_d \times N_tN_rN_d)$ matrix whose ith diagonal element m_i is dependent on the channel power and is given by $\sigma_i^2/(\sigma_i^2 + \beta/\text{SNR})$, where σ_i^2 is the power of the ith angle-time domain channel coefficient. Clearly, $\mathbf{M}$ represents the AMMSE process. Since $\mathbf{M}$ is diagonal and real, we have

$$\mathbf{M}^H = \mathbf{M}. \tag{9.18}$$

From (9.1), the MSE of channel power based AMMSE technique is given by

$$\text{MSE}_{\text{AMMSE}} = \frac{1}{N_dN_tN_r}\text{E}\left[||\mathbf{B}^H\mathbf{M}\mathbf{B}\underline{\mathbf{c}}_{\text{LS}} - \underline{\mathbf{c}}||^2\right]. \tag{9.19}$$

Note that $\mathbf{M}$ is dependent on not $\underline{\mathbf{c}}$ but the channel power. Thus, $\mathbf{M}$ is uncorrelated to $\underline{\mathbf{c}}$. Further, $\underline{\mathbf{X}}$ is uncorrelated to $\underline{\mathbf{c}}$. Then, using (9.16) and (9.18), we obtain

$$\text{MSE}_{\text{AMMSE}} = \frac{1}{N_dN_tN_r}\text{trace}\Big\{\frac{\beta}{\text{SNR}}\mathbf{M}\mathbf{B}\mathbf{B}^H\mathbf{M} + \mathbf{R}^{\text{a}} - 2\mathbf{M}\mathbf{R}^{\text{a}} + \mathbf{M}\mathbf{R}^{\text{a}}\mathbf{M}\Big\} \tag{9.20}$$

where

$$\mathbf{R}^{\text{a}} = \text{E}[\mathbf{B}\underline{\mathbf{c}}\ \underline{\mathbf{c}}^H\mathbf{B}^H] \tag{9.21}$$

is the angle-time domain channel correlation matrix. Further, (9.20) can be rewritten as

$$\text{MSE}_{\text{AMMSE}} = \frac{\beta}{N_dN_tN_r\text{SNR}}\sum_{i=0}^{N_dN_tN_r-1}\frac{\sigma_i^2}{\sigma_i^2 + \frac{\beta}{\text{SNR}}}. \tag{9.22}$$

As $\sigma_i^2/(\sigma_i^2 + \beta/\text{SNR})$ is always smaller than 1, the channel power based AMMSE technique always yields better performance compared to the LS technique. Moreover, this technique should achieve the same performance as the Q1D LMMSE technique in the sample-spaced channel, where the channel correlation is equivalent to the channel power within each angle-time domain beam. But the performance of this AMMSE technique will be degraded for non-sample-spaced channels, where the channel coefficients in different time indices are correlated due to the power leakage effect (see Chapter 6), as shown in the next section. Nevertheless, it still significantly outperforms the conventional array-frequency domain LS technique at all the SNRs under consideration.

As the channel power based AMMSE technique filters each LS estimated channel coefficients independently, it only requires 1 multiplication for each channel coefficient when $\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)$ is known. If only LS estimated array-frequency domain channel coefficients are available beforehand, then from the previous subsection we know that the transformations between the array-frequency domain and angle-frequency domain require totally $2(N_t+N_r)$ complex multiplications for each channel coefficient. In addition, the transformations between the angle-frequency domain and angle-time domain require totally $2N_d$ complex multiplications for each channel coefficient by examining (9.13). Typically, N_d is the power of 2. Then, using the FFT and IFFT [187] for transformations between the angle-time and angle-frequency domain, the total complex multiplications required for each channel coefficient is reduced to $\log_2 N_d$ complex multiplications for each channel coefficient. From the above, the total required complex multiplications for each channel coefficient is $\log_2 N_d + 2(N_t + N_r) + 1$. In view that the Q1D LMMSE technique requires additionally complexity of $O(N_d^{2.376})$ for the matrix inversion shown in (9.10), the complexity of the channel power based AMMSE technique is always lower compared to the previous techniques.

9.2.6 Channel Instantaneous Power Based AMMSE Technique

When even the channel power is not available, we can still realize the AMMSE technique by using the channel instantaneous power to approximate the channel power as shown in [97]. In this technique, we still use (9.17) to get the estimated angle-time domain channel vector but m_i is revised as

$$m_i = \begin{cases} \frac{|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2 - \frac{\beta}{\mathrm{SNR}}}{|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2}, & \text{if } |\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2 \geq \eta, \\ 0, & \text{otherwise.} \end{cases} \tag{9.23}$$

Here the threshold η is chosen not to be smaller than β/SNR. Otherwise, the approximated channel power ($|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2 - \beta/\mathrm{SNR}$) becomes negative. (9.23) clearly shows that m_i is dependent on $|\tilde{c}^{\mathrm{a}}_{i_r,i_t}(l)|^2$ and η, which makes the MSE computation quite difficult. The complicated expression to compute the MSE can be found in Chapter 8, and will not be covered in this chapter. From Chapter 8, we know that when the channel power is completely unknown, setting the threshold to be $2\beta/\mathrm{SNR}$ is sufficient for the channel instantaneous power based AMMSE technique to improve over the LS technique in terms of performance. When additional information (*e.g.* the portion of the number of channel coefficients, whose average powers are below the β/SNR, to the total number of channel coefficients) is available, the threshold may be moderately adjusted to achieve further performance gain. Same as the channel power based AMMSE technique, the channel instantaneous power based AMMSE requires only $\log_2 N_d + 2(N_t + N_r) + 1$ complex multiplications for each channel coefficient when only LS estimated array-frequency domain channel coefficients are available beforehand.

Table 9.1: Required complex multiplications per channel coefficient for different channel estimation techniques.

	General Case	Typical Case
2D LMMSE	$N_t N_r N_d$	1024
2D SVD	$2N'_h$	400
Q1D LMMSE	$N_d + 2(N_t + N_r)$	80
AMMSE (Average)	$\log_2 N_d + 2(N_t + N_r) + 1$	23
AMMSE (Instant)	$\log_2 N_d + 2(N_t + N_r) + 1$	23

Table 9.1 lists the required complex multiplications for each channel coefficient in all the channel estimation techniques introduced for both the general case and one typical case under the assumption that LS estimated array-frequency domain channel coefficients are available beforehand. We use the 'AMMSE (Average)' and 'AMMSE (Instant)' to represent the channel power based AMMSE technique and channel instantaneous power based AMMSE technique, respectively, in the table. In the typical case, we set $N_t = N_r = 4$, $N_d = 64$. Additionally, we consider $N'_h = 200$, which is a typical value obtained by simulating the IEEE 802.11 TGn channel model E in [61] such that only the largest N'_h eigenvalues of the array-frequency domain 2D channel correlation $\mathbf{R}$ are larger than the noise power. As shown in the typical case, all the angle-domain

techniques require much less complex multiplications for each channel coefficient and thus result in significant complexity reduction compared to the 2D LMMSE and SVD techniques.

9.3 Simulation Results

Simulation environment was introduced in Section 8.4 and will not be repeated here. Then, in this section, we use the 'LS', '2D LMMSE', 'Q1D LMMSE', 'AMMSE (Average)' and 'AMMSE (Instant)' to represent the LS technique, 2D LMMSE technique, Q1D LMMSE technique, channel power based AMMSE technique and channel instantaneous power based AMMSE technique, respectively. Further, when ignoring the eigenvalues smaller than β/SNR, our simulation results show that the 2D SVD technique performs similarly as the 2D LMMSE technique. As the complexity of both techniques is relatively large compared to our proposed angle-domain techniques, we will not rigorously differentiate these two techniques in this section and the results of the 2D SVD technique will not be presented.

9.3.1 Channel Model A

The same as Subsection 8.4.1, we consider the AS to be 40° for both the AS_t and AS_r as the worst case in performing channel estimation in the angle domain. We also consider the AS to be 2° for both the AS_t and AS_r as the best case, which is valid for outdoor environments. Further, we consider 45° to be the AoD_m and AoA_m of the cluster for the worse-case consideration. We also consider 0°, from which most of energy of multipath components concentrate on one angle-time domain beam, to be the AoD_m and AoA_m. For the channel model A, the channel correlation is equivalent to the channel power in each angle-time domain beam. Therefore, the 'Q1D LMMSE' achieves the same performance as the 'AMMSE (Average)'. In this sense, we will not differentiate these two techniques in this subsection.

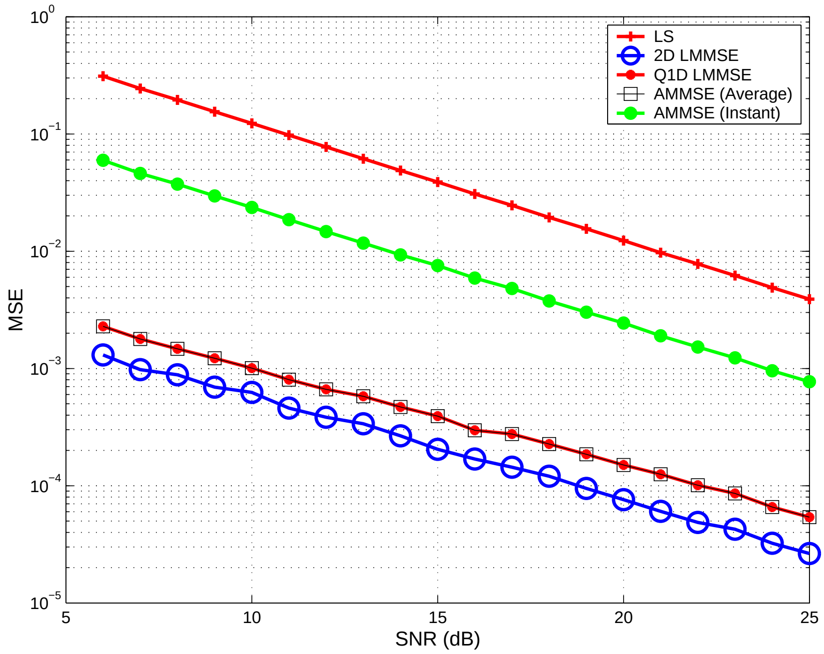

Figure 9.1: Performances of different channel estimation techniques for the model A with $\mathrm{AoD_m} = 0°$, $\mathrm{AS}_t = 2°$, $\mathrm{AoA_m} = 0°$, and $\mathrm{AS}_r = 2°$.

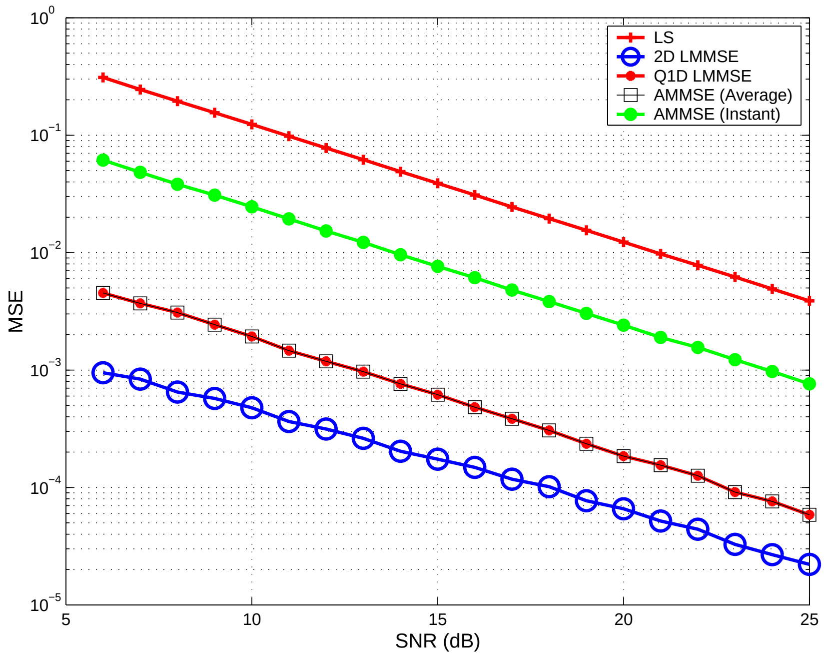

Figure 9.2: Performances of different channel estimation techniques for the model A with $\mathrm{AoD_m} = 45°$, $\mathrm{AS}_t = 2°$, $\mathrm{AoA_m} = 45°$, and $\mathrm{AS}_r = 2°$.

Fig. 9.1 and Fig. 9.2 correspond to the cases when the angular spread is 2°. In the first case when the $\mathrm{AoA_m}$ and $\mathrm{AoD_m}$ are both equal to 0°, each physical path approximately corresponds to one angle-time domain beam. Thus, channel coefficients in different angle-time domain beams are approximated spatially uncorrelated. Therefore, compared to the '2D LMMSE', the 'Q1D LMMSE' does not degrade the performance significantly as shown in Fig. 9.1. In the second case when the $\mathrm{AoA_m}$ and $\mathrm{AoD_m}$ are both equal to 45°, the energy of multipath components leaks into more than one angle-time domain beams. Then, the spatial correlation between channel coefficients in different angle-time domain beams becomes much larger compared to the first case. The ignorance of this spatial correlation will result in more performance degradation for the 'Q1D LMMSE' in the second case (7 dB when the MSE is 10^{-3}) compared to the first case (3 dB when the MSE is 10^{-3}). This result indicates that the performance of the 'Q1D LMMSE' is highly dependent on the $\mathrm{AoA_m}$ and $\mathrm{AoD_m}$ when the angular spread is small. We also find that although the 'AMMSE (Instant)' cannot achieve similar performance as the '2D LMMSE', it still achieves 7 dB performance gain when the MSE is 10^{-2} compared to the 'LS'. This result indicates that we can still use the 'AMMSE (Instant)' to improve the performance of the LS technique when the channel power is not available to the receiver.

Fig. 9.3 and Fig. 9.4 correspond to the cases when the angular spread is 40°. At first glance, it seems surprising that the 'Q1D LMMSE' yileds similar performance as the '2D LMMSE' in both figures. Although not shown here, our further investigations show that similar results are obtained by computing (9.7) and (9.22). These results support the observations in Fig. 9.3 and Fig. 9.4. The essential reason for these results is that the angular spread is relatively large, which makes the first largest 16 (channel model A only consists of 4×4 nonzero angle-time domain coefficients, *i.e.* 1/64 of the total coefficients) eigenvalues of $\mathbf{R}$ much larger than all the β/SNR under consideration. Then, of all the terms $\lambda_{ii}/[\lambda_{ii} + \beta/(\mathrm{SNR})]$ (for $ii = 0, 1, \cdots, 1023$), 1/64 of the total terms approach one and the rest is zero. Similarly, the first largest 16 diagonal elements of $\mathbf{R}^{\mathrm{a}}$ are much larger than all the β/SNR under consideration. Then, 1/64 of all the terms

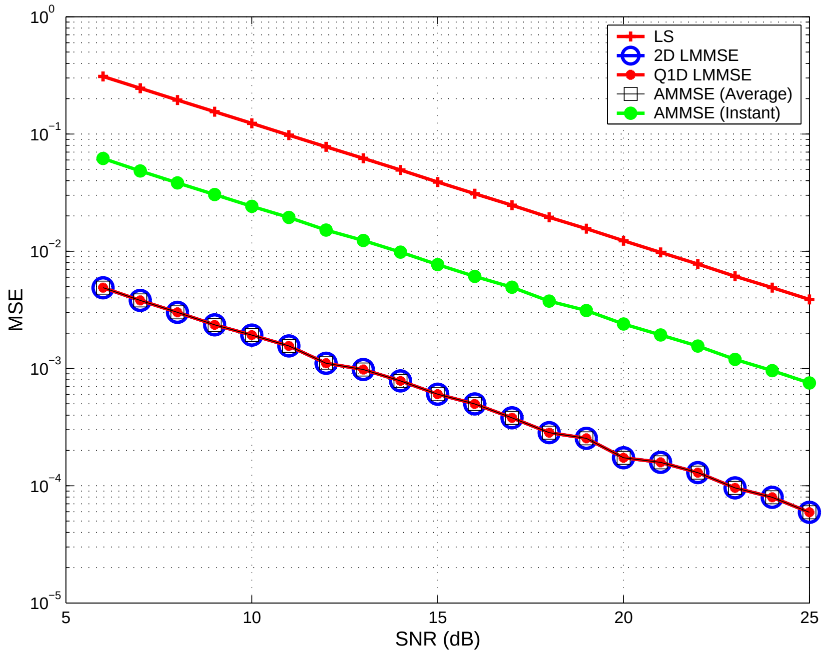

Figure 9.3: Performances of different channel estimation techniques for the model A with $\mathrm{AoD_m} = 0°$, $\mathrm{AS}_t = 40°$, $\mathrm{AoA_m} = 0°$, and $\mathrm{AS}_r = 40°$.

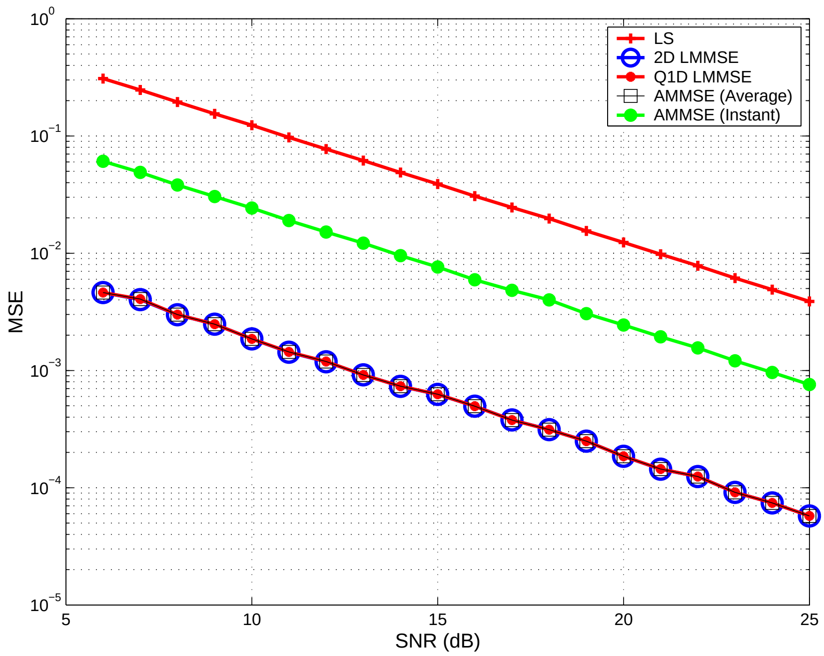

Figure 9.4: Performances of different channel estimation techniques for the model A with $\mathrm{AoD_m} = 45°$, $\mathrm{AS}_t = 40°$, $\mathrm{AoA_m} = 45°$, and $\mathrm{AS}_r = 40°$.

$\sigma_i^2/(\sigma_i^2 + \beta/\text{SNR})$ (for $i = 0, 1, \cdots, 1023$) approach one and the rest are zero. Therefore, the MSEs of both the '2D LMMSE' and 'Q1D LMMSE' approach 1/64 of that of the 'LS' as shown in both Fig. 9.3 and Fig. 9.4. Note that it is the fact that the number of large values of σ_i^2 and λ_{ii} is the same, and similarly that the number of zero values of σ_i^2 and λ_{ii} is the same, which results in the similar performances of the '2D LMMSE' and 'Q1D LMMSE' even though the nonzero values of σ_i^2 and λ_{ii} do not quite coincide with each other. This is unlike in the cases when the angular spread is $2°$ as shown in Fig. 9.1 and Fig. 9.2, where the 'Q1D LMMSE' performs worse than the '2D LMMSE' due to the small mismatch between the values of σ_i^2 and λ_{ii}. Further, we find that the performance gains achieved by the '2D LMMSE' in Fig. 9.3 and Fig. 9.4 are not significant as those shown in Fig. 9.1 and Fig. 9.2. This is because when the angular spread is small, not all the 1/64 of the total terms for the $\lambda_{ii}/[\lambda_{ii} + \beta/(\text{SNR})]$ (for $ii = 0, 1, \cdots, 1023$) approach one. Again, we find that the 'AMMSE (Instant)' achieves better performance compared to the 'LS'.

From the above, we may conclude that the performance of angle-domain techniques are highly dependent on the angular spread, AoA_m and AoD_m. When the angular spread is small, the performance degradation of the 'Q1D LMMSE' compared to the '2D LMMSE' is dependent on the AoA_m and AoD_m. When the angular spread is large, the 'Q1D LMMSE' performs similarly as the '2D LMMSE', and is robust to the change of AoA_m and AoD_m.

9.3.2 Typical Channel Models

We consider typical non-sample-spaced indoor channels in Fig. 9.5 and Fig. 9.6. From these two figures, we have five important findings. First, the 'Q1D LMMSE' performs similarly as the '2D LMMSE' in both Fig. 9.5 and Fig. 9.6 at all the SNRs under consideration. This is because the angular spread for each cluster at both model B and E is relatively large. Then, consistent with the results in the model A, it is not surprising that the 'Q1D LMMSE' performs similarly as the '2D LMMSE'. Therefore, when the channel correlation is available to the receiver, the ‘Q1D LMMSE’ can replace the ‘2D LMMSE’ because of its much lower complexity. Second, unlike in the model A, the ‘AMMSE (Average)’ performs worse than the ‘Q1D LMMSE’ because the channel power is no longer equivalent to the channel correlation in each angle-time domain beam for the non-sample-spaced channels. As shown, the adverse effects of this nonequivalence become more obvious at high SNRs. Third, unlike in the model A, the performance improvement of the ‘AMMSE (Average)’ over the ‘AMMSE (Instant)’ is more significant at low SNRs than at high SNRs. This is because at lower SNRs, it is less accurate to estimate the channel power from the channel instantaneous power in typical models compared to in the model A. Fourth, the ‘AMMSE (Instant)’ always performs better than the LS technique. As the term $\sigma_i^2/(\sigma_i^2 + \beta/\text{SNR})$ shown in (9.22) gradually approaches one with the increase of SNR, the MSE of ‘AMMSE (Average)’ gradually approaches that of ‘LS’ as observed. Nevertheless, when the channel power is known to the receiver, the ‘AMMSE (Average)’ can still replace the ‘LS’ because it achieves better performance for all the SNRs under consideration while requiring comparable complexity for implementation. Fifth, the ‘AMMSE (Instant)’ still outperforms the ‘LS’ at all the SNRs. Additionally, its performance gain achieved at low SNRs is more significant than that at high SNRs because the error floor due to the ignorance of the channel power gradually becomes less obvious at lower SNRs. As the ‘AMMSE (Instant)’ requires no knowledge of channel stochastic information and has simple complexity, we can always use it to replace ‘LS’ for MIMO-OFDM systems.

9.4 Conclusions

In this chapter, we have investigated various angle-domain channel estimation techniques for MIMO-OFDM systems. All the angle-domain techniques are flexible in implementation and were found to perform better than the conventional LS technique for all the MIMO-OFDM channels under consideration. The applicability of angle-domain techniques is dependent on the channel stochastic information available to the receiver. For the situation when no stochastic information is available, we have used the AMMSE

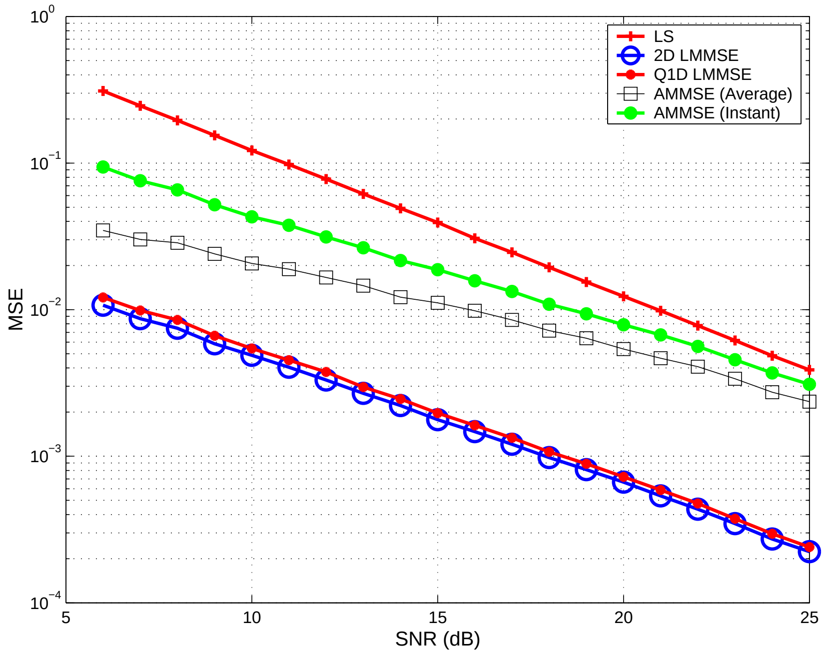

Figure 9.5: Performances of different channel estimation techniques for the model B.

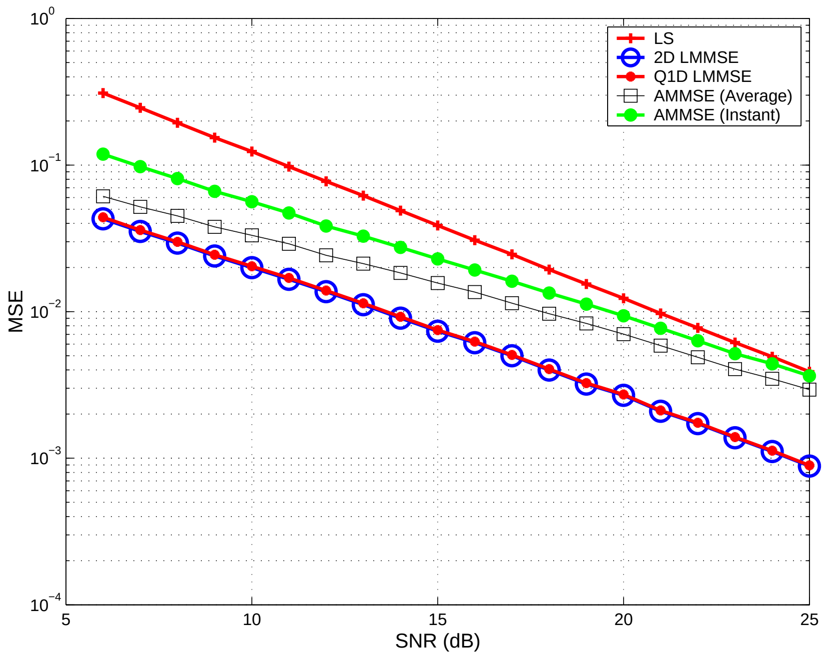

Figure 9.6: Performances of different channel estimation techniques for the model E.

technique based on the angle-time domain channel instantaneous power to improve over the LS technique in terms of performance. For the situation when the channel power is known, we have used the AMMSE technique based on the angle-time domain channel power. For the situation when the channel correlation is known to the receiver, the choice of angle-domain techniques is then a trade-off between the performance and complexity. If the performance is of the utmost concern, we may choose the Q1D LMMSE technique because it can perform similarly as the 2D LMMSE technique with significantly lower complexity for typical MIMO-OFDM models. If the computational complexity is of the utmost concern, we may choose the channel power based ALMMSE technique because it can achieve significant performance gain while maintaining comparable complexity compared to the LS technique for all the MIMO-OFDM channel models under consideration.

Chapter 10

Conclusions and Future Work

We explore reduced-complexity signal processing techniques to facilitate high-speed and high-quality data reception in multiple-input multiple-output storage and wireless communication systems in this thesis. As suggested in the title, this thesis consists of two parts. In Part I of the thesis, we investigate the reduced-complexity detection techniques under the assumption that the channel is *a priori* known. The techniques proposed in this part facilitate the high-speed implementation of the two-dimensional optical storage (TwoDOS) system, which may push the development of the 4th generation optical storage system. Moreover, although the techniques are developed for the TwoDOS system whose bit-cells are arranged in a hexagonal structure, most techniques are applicable to any multi-track data storage system with square or rectangular bit-cells. In Part II of the thesis, we study channel estimation techniques for multiple-input multiple-output systems where prior knowledge of the channel is not available. These channel estimation techniques perform noise filtering in the angle domain, where the channel model lends itself to a simple physical interpretation. To the best of our knowledge, we are the first to systematically investigate these angle-domain channel estimation techniques. Though the techniques in this part are developed for multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems, they are applicable to other multiple-input multiple-output wireless communication systems as well.

10.1 Reduced-Complexity Detection Techniques

10.1.1 Conclusions of Part I

In Part I of the thesis, the two-dimensional (2D) Viterbi detector (VD), which is the optimal 2D detector in the presence of additive white Gaussian noise, serves as the benchmark in terms of performance. We develop techniques to reduce the complexity of the 2D VD in both the temporal and spatial dimensions, and develop a 2D target optimization technique to compensate for the detection performance loss due to the complexity reduction in both dimensions. The techniques proposed in Part I ensure the high-speed implementation of the two-dimensional optical storage (TwoDOS) system. Further, our proposed generalized 2D VD called the FDTS-DF/VD provides flexibility to design multiple-input multiple-output systems with a trade-off between performance and complexity. We also show that by judiciously choosing the target and number of tracks under consideration in FDTS/DF-VD, we can develop a reduced-complexity 2D Viterbi-like detector that facilitates the high-speed TwoDOS implementation without paying a large penalty in detection performance. Moreover, though the techniques are developed for the TwoDOS system whose bit-cells are arranged in a hexagonal structure, most techniques are applicable to any multi-track data storage system with square or rectangular bit-cells.

In order to reduce the complexity of the 2D VD in the temporal dimension as well as maintain reliable detection, we propose two techniques to reduce the complexity of 2D VD by means of shortening the channel memory. One technique concerns the 2D minimum mean square error (MMSE) equalizer for a given 2D partial response (PR) target. The equalizer developed in Chapter 3 is a 2D MMSE equalizer that is more general than that of in [156] since it can deal with correlated data, colored additive noise, domain bloom, and transition jitter. The other technique is to jointly design the equalizer and targets based on MMSE approach to improve the performance of 2D VD. Instead of directly imposing a constraint on the 2D PR target to avoid the trivial solution, we propose a novel technique which converts the 2D target design problem into a 1D problem. This technique is not developed for TwoDOS only, it is applicable to any multiple-input multiple-output system that requires a good target to ensure good detection performance. The results show that the target that has the smallest noise correlation effect leads to the best detection performance for the 2D VD. Note that this conclusion is obtained in the 2D case and can be seen as a generalized conclusion of [137], which is obtained in the one-dimensional (1D) case. Moreover, it is observed that for symmetric targets, the one that has the largest normalized central tap results in the best detection performance of the 2D VD. This observation reconfirms that our proposed 2D monic constraint is a reasonable target constraint since it causes the energy to be concentrated near the normalized central tap and accordingly results in the largest normalized central tap compared with other target constraints. In order to evaluate different targets, we simplify the conventional analytical technique by taking a smaller number of bits into account, which was shown to be still effective for analyzing the detection performance of the 2D VD. This simplification is reasonable since only the bits that have little contribution to the residual intersymbol interference (ISI) or intertrack interference (ITI) are ignored, and will thus result in minor inaccuracies in analyzing the detection performance of the 2D VD. Moreover, the threshold chosen in this reduced-complexity analytical technique is variable. This variable threshold provides great flexibility to deal with the performance-complexity trade-off, which is application-dependent. Furthermore, there are no assumptions about the target. Therefore, the reduced-complexity analytical technique is a general technique that can be used to evaluate any target under any channel condition.

For the purpose of reducing the complexity of the 2D VD in the spatial dimension as well as achieving reliable detection, a generalized 2D VD called the FDTS/DF-VD is proposed. It provides the flexibility to design multiple-input multiple-output systems with trade-off between performance and complexity. The conventional full-fledged 2D VD, QR detector, and our proposed quasi-1D VD can all be seen as special cases of this detector. The basic idea of the detector is to divide the full-fledged 2D VD into a set of sub-2D VDs, each dealing with a smaller number of tracks. The result shows that the dominant factor that degrades the performance of the detector is the reduction in the effective target energy available for detection. Therefore, the choice of the number of tracks considered in each sub-2D VD, which determines the effective target energy, is quite important. In this research, we find that setting the number of tracks to three permits acceptable complexity and performance in the TwoDOS system. As the number of tracks is variable in the design of detectors, the proposed detector can provide flexibility to deal with the performance-complexity trade-off. Thus, it is applicable to other systems that have different constraints on performance and complexity. We also propose a causal ITI target that has been shown to facilitate the design of reduced-complexity 2D Viterbi-like detectors. This can be explained as follows. The causal ITI target makes some inputs before the 2D VD suffer from only ISI, unlike the traditional targets that make all the inputs suffer from both ISI and ITI. Thus, we can detect those inputs that suffer from only ISI without using any technique to estimate the ITI resulting from the spatial dimension. The detections of these target-ITI-free inputs are immune to detrimental effects due to the spatial error propagation compared with conventional 2D targets, and can then aid the detection of other inputs that are constrained to suffer only causal ITI. Further, by putting some additional but reasonable constraints on the causal ITI target, we propose two new targets specifically for the FDTS/DF. They can result in reduction in the latency due to the spatial error propagation for the inputs that are constrained to suffer causal ITI, and can lead to lower complexity in the implementation of FDTS/DF-VD as well as the process of target design. The results suggest that by judiciously choosing the target and number of tracks under consideration in FDTS/DF-VD, we can develop a reduced-complexity 2D Viterbi-like detector that facilitates the high-speed TwoDOS implementation without paying a large penalty in detection performance. Note that there are no assumptions about the casual ITI. Therefore, the concept of causal ITI is applicable to any 2D system. Further, constraining the target to be causal intersymbol interference can be extended to multidimensional systems as well.

Since most of the techniques proposed in this study do not assume any specific constraints of the TwoDOS system, they are applicable to any multiple-input multiple-output system. For instance, the novel technique which converts the 2D target design problem into a 1D problem can be used for any 2D target design. The performance evaluation technique and FDTS/DF-VD both possess one parameter that is adjustable so that both can meet different performance and complexity requirements of various multiple-input multiple-output systems.

10.1.2 Future Work

At the end of this thesis, we would like to present a number of recommendations and open issues to stimulate future research based on our work on detection techniques.

- Unlike the TwoDOS system, many multiple-input multiple-output systems are not limited to binary signals. Therefore, effective techniques (*e.g.* set partitioning [63]) should be investigated to reduce the complexity of our proposed FDTS/DF and RFDTS/DF for multilevel signals.

- Our techniques for the TwoDOS system assume that the system has an equal number of inputs and outputs. However, in many multiple-input multiple-output wireless communication systems, the number of transmit and receive antennas can be different. It has been shown that, for a given number of transmit antennas, the detection performance improves with the increase of the number of the receive antennas due to the receiver diversity. For this reason, the number of transmit antennas is usually larger than that of the receive antennas. Thus, the application of our proposal techniques to these nonsymmetric transmission systems needs to be investigated.

- The 2D equalizer before the 2D detector usually colors the additive white Gaussian noise (AWGN). Therefore, some effective techniques, *e.g.* the noise-predictive maximum likelihood (NPML) technique that embeds a noise prediction process into the branch metric computation of a Viterbi-like detector [40,41,47], should be investigated to whiten the colored noise and to enhance the detection performance.

- As our proposed FDTS/DF and RFDTS/DF divide the conventional full-fledged 2D VD into a set of sub-2D VDs, it is interesting to see whether designing a set of sub-2D equalizers and targets specifically for each sub-2D VD is suitable for the FDTS/DF and RFDTS/DF. Further, as most of the detection errors occur from the lower tracks in the FDTS/DF and RFDTS/DF, it is possible to design track-dependent 2D codes that have stronger error correction capability but relatively lower code rate for lower tracks, and relatively weaker error correction capability but higher code rate for upper tracks.

- Since 2D detectors serve as a basis for the investigation of reduced-complexity higher-dimensional detectors that may emerge in the future, our techniques could be applicable to multidimensional (nD) detectors as well. In such cases, we can first design an nD causal intersymbol interferece target so that part of inputs before the nD VD suffers from only ISI. Then, we can detect the corresponding bits from these inputs that suffer from only ISI. Similar to the FDTS/DF or RFDTS/DF, the resulting detected bits are used to facilitate the data recovery from all the other inputs dimension by dimension.

10.2 Reduced-Complexity Channel Estimation Techniques

10.2.1 Conclusions of Part II

In Part II of the thesis, we develop several reduced-complexity, suboptimal, approximated linear MMSE (LMMSE) channel estimation techniques in the angle domain, where the channel model lends itself to a simple physical interpretation. All the angle-domain techniques proposed are flexible in implementation. They can either use conventional array-domain estimators as the coarse estimators and perform post-processing in the angle domain, or use the specifically designed pilots introduced in Chapter 7 for a direct implementation. The applicability of these angle-domain techniques is highly dependent on the channel stochastic information (*e.g.* channel power or correlation) available to the receiver. To the best of our knowledge, we are the first to systematically investigate these angle-domain channel estimation techniques. Though the techniques are developed for MIMO-OFDM systems, they are applicable to other multiple-input multiple-output wireless communication systems as well.

For the case where no channel stochastic information is available to the receiver, we develop the angle-frequency domain most significant taps (MST) selection technique, the angle-time domain MST selection technique, and the angle-time domain approximated MMSE (AMMSE) technique. These three techniques are shown to be effective when the angular spread of clusters of multipath components is small. More importantly, the angle-time domain techniques can improve over the array-frequency domain LS technique in all the situations under consideration even when the angular spread is relatively large. Further, both our theoretical analysis and simulation results indicate that the angle-time domain AMMSE technique achieves the best performance, and is robust to the choice of the threshold and the mismatch between the actual signal-to-noise ratio (SNR) and the assumed SNR. Thus, when only the target SNR range is available to the receiver, the angle-time domain AMMSE technique is suitable for typical IEEE 802.11 TGn MIMO-OFDM systems. In addition, we also find that with a suitable threshold and known operating SNR, the angle-time domain MST selection technique results in little performance degradation compared to the angle-time domain AMMSE technique. Therefore, in such cases, the angle-time domain MST selection technique may be a potential candidate for the IEEE 802.11 TGn MIMO-OFDM systems because of its lower computational complexity.

The above angle-time domain AMMSE technique utilizes the channel instantaneous power (*i.e.* the instantaneous power of estimated channel coefficients) and is thus referred to as the channel instantaneous power based AMMSE technique. For the case where the channel power is known, we develop the AMMSE technique that is based on the angle-time domain channel power. This channel power based AMMSE technique has similar complexity to the channel instantaneous power based AMMSE technique but provides better performance. For the case where the channel correlation is known to the receiver, we develop the quasi 1D (Q1D) LMMSE technique that can further increase the performance. Our simulation results show that the Q1D LMMSE technique can even perform similarly as the 2D LMMSE technique but with significantly lower complexity.

The channel estimation techniques developed in this thesis are investigated in both the angle-time and angle-frequency domains, which at first glance seems uniquely applicable to MIMO-OFDM systems. However, the angle-domain channel estimation techniques are applicable to many other multiple-input multiple-output wireless communication systems because the angle-domain representations can be used to characterize channels in those wireless multiple-input multiple-output (MIMO) system. For example, the angle-time domain in MIMO-OFDM systems can be seen as an extension of the angle domain, which corresponds to MIMO flat fading channels. Thus, the channel instantaneous power based AMMSE technique, channel power based AMMSE technique, and Q1D LMMSE technique are all applicable in MIMO flat fading channels. Further, the estimated channel coefficients in the angle domain may serve as the basis for the future coding and detection techniques developed in the angle domain.

10.2.2 Future Work

At the end of this thesis, we would like to present a number of recommendations and open issues to stimulate future research based on our work on angle-domain channel estimation techniques.

- We assume the channel to be time-invariant over a given training period as our main concern in this thesis is the indoor propagation environment. But channel estimation for rapidly time-varying MIMO-OFDM systems is also a challenging task because the channel is required to be reestimated frequently. In [35], the channel matrix is found to be factored into time-dependent and time-independent components under the assumption that the scatters are stationary. Thus, we may estimate the time-independent components first and keep tracking the channel by reestimating the time-dependent components, which are dependent on the velocity of the mobile station, the number and angles of departure (AoD) of resolvable scatters. In the angle-domain representation, the number and AoD of scatters are kept constant. Thus, we only need to track the channel by estimating the velocity, which is a scalar. This will significantly reduce the number of coefficients to be reestimated in time-varying MIMO-OFDM systems and could be investigated in our future research.

- We may also investigate channel estimation for "keyhole" (or "pinhole") channels, which exhibit low spatial correlations at both transmit and receive antennas but still have low-rank properties [34, 72]. This situation may be imaged by placing a screen, with a small keyhole punched through it, to separate the scatters from the ones around transmit antennas and receive antennas. As the keyhole is the only way for microwaves to pass through, it will relay the captured signals from the scatters around transmit antennas to the scatters around receive antennas. In the angle domain, the channel matrix at a given delay is an outer product of two independent angle-domain channel vectors. We can thus estimate the channel vectors instead of directly estimating the channel matrix itself. As a result, the number of parameters to be estimated is significantly reduced, which is especially useful for systems with large number of transmit and receive antennas.

- We assume that all subcarriers are used for data or pilot transmission in MIMO-OFDM systems and use the discrete Fourier transform (DFT) or inverse DFT (IDFT) to transform the channel between angle-time and angle-frequency domain. However, in many practical MIMO-OFDM systems, some subcarriers are not used for data or pilot transmission. These unused subcarriers are referred to as the virtual carriers (VCs) [121]. In such MIMO-OFDM systems with VCs, the IDFT operation, which transforms the estimated channel coefficients from the angle-frequency domain into the angle-time domain, is not applicable because it requires that the discrete samples to be processed are equally spaced. But the discrete samples available are the estimated channel coefficients at pilot subcarriers, which are usually not equally spaced. Therefore, some effective techniques, *e.g.* by first estimating the channel coefficients at VCs [95], should be investigated to ultimately improve the estimation performance for the channel coefficients at pilot subcarriers.

- We show that channel estimation techniques developed in the angle domain are effective for multiple-input multiple-output systems. Due to the physical meaning of the angle-domain representation of systems, future work may focus on the development of coding and signal processing in the angle domain. For example, in spatially correlated MIMO channels, some angle-domain beams may contain insufficient information for data recovery. Based on this, we may develop some reduced-complexity detection techniques for such MIMO systems.

- We assume that the antenna array configuration is the uniform linear array. However, 2D and three-dimensional (3D) arrays are also used in practice to form better beamforming patterns compared with the uniform linear array. Besides, the mutual coupling between antenna elements in an array may not always be ignorable in practice. All these cases will affect the array response vectors defined in (7.13) and (7.14) and thus the transformation between the array domain into the angle domain. In the future, we could extend our work to investigate angle-domain channel estimation techniques for such cases by reconstructing the transformation matrix. In addition, we could redesign pilots with the method introduced in Section 7.4 for the direct implementation of such angle-domain techniques.

- We do not consider strong interference conditions, which are likely to exist in MIMO-OFDM systems. For example, in the 2.4 GHz band, systems may suffer from many interferences coming from (other) wireless local area network (WLAN) systems, microwave ovens, Bluetooth devices, *etc.* This might happen only at selected subcarriers and/or for a limited duration. Therefore, we could consider some effective techniques (*e.g.* Huber's M estimator [99], error correction codes, 2D and 3D antenna arrays) to deal with these strong interferences in conjunction with angle-domain channel estimation techniques.

Bibliography

[1] ETSI ETS 300 744, "Digital video broadcasting (DVB); framing structure channel coding and modulation for digital terrestrial television (DVB-T)," Nov. 1996.

[2] ETSI ETS 300 401, "Radio broadcasting systems; digital audio broadcasting (DAB) to mobile, portable and fixed receivers," May 1997.

[3] ETSI TS 101 475, "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer," Apr. 2000.

[4] The Eureka 147 Consortium [Online]. Available: http://www.worlddab.org/eureka.aspx.

[5] IEEE Std 802.11-1997, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications," June 1997.

[6] IEEE Std 802.11a-1999, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer in the 5 GHz band," Sept. 1999.

[7] IEEE Std 802.11b-1999, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer extension in the 2.4 GHz band," Sept. 1999.

[8] IEEE Std 802.11g-2003, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, Amendment 4: further higher data rate extension in the 2.4 GHz band," Sept. 2003.

[9] Multimedia Mobile Access Communication Systems [Online]. Available: http://www.arib.or.jp/mmac/e/.

[10] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, no. 5, pp. 779-785, May 1970.

[11] M. Abramowitz (Editor) and I. A. Stegun (Editor), *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Dover Publications, Mineola, NY, 1972.

[12] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data-rate wireless communication over wideband channels," in *Proc. Intl. Conf. IEEE Veh. Technol.*, Ottawa, Canada, May 1998, pp. 2232-2236.

[13] S. Alamouti, "A simple transmit diversity technique for wireless communications,"*IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[14] H. Ali, J. H. Manton, and Y. Hua, "A SOS subspace method for blind channel identification and equalization in bandwidth efficient OFDM systems based on receive antenna diversity," in *Proc. IEEE Workshop on Statistical Signal Process.*, Singapore, Aug. 2001, pp. 401-404.

[15] M. E. Austin, *Decision-feedback equalization for digital communication over dispersive channels,* Tech. Rep. 437, MIT Lincoln Lab, Lexington, MA, Aug. 1967.

[16] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-carrier digital communications theory and applications of OFDM*, 2nd ed., Springer Science+Business Media, Inc., New York, NY, 2004.

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

[18] I. Barhumi, G. Leus, and M. Moonen, "Time-domain channel shortening and equalization of OFDM over doubly-selective channels," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Montreal, Canada, May 2004, pp. 801-804.

[19] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson, "On channel estimation in OFDM systems," in *Proc. IEEE Intl. Veh. Technol. Conf.*, Chicago, IL, July 1995, pp. 815-819.

[20] M. Berkoff, "Waveform compression in NRZI magnetic recording," *Proc. IEEE*, vol. 52, pp. 1271-1272, Oct. 1964.

[21] C. A. Belfiore and J. H. Park, "Decision feedback equalization," *Proc. IEEE*, vol. 67, no. 8, pp. 1143-1156, Aug. 1979.

[22] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications," *IEEE Trans. Autom. Control*, vol. AC-25, no. 3, pp. 385-399, June 1980.

[23] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, vol. COM-32, no. 8, pp. 871-883, Aug. 1984.

[24] J. W. M. Bergmans, *Digital baseband transmission and recording*, Kluwer Academic Publishers, Boston, MA, 1996.

[25] J. W. M. Bergmans, W. M. J. Coene, R. Otte, and S. Bramwell, "Transition detector for CD and DVD,"*IEEE Trans. Consum. Electron.*, vol. 46, no. 1, pp. 16-19, Feb. 2000.

[26] J. A. C, Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5-14, May 1990.

[27] L. Boden, *Mastering CD-rom technology*, John Wiley & Sons, Inc., New York, NY, 1995.

[28] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K. S. Immink, *Principles of optical disc systems*, Adam Hilger Ltd, Bristol, UK, 1985.

[29] J. Braat, "Optically read disks with increased information density," *Appl. Opt.*, vol. 22, no. 14, pp. 2196-2201, July 1983.

[30] X. Cai and A. N. Akansu, "A subspace method for blind channel identification in OFDM systems," in *Proc. IEEE Intl. Conf. Commun. (ICC)*, New Orleans, LA, June 2000, pp. 929-933.

[31] K. S. Chan and F. Lim, "Analysis and optimized equalizer design for threshold-like detectors for $d = 1$ optical recording channels," *Proc. SPIE*, vol. 5380, pp. 535-543, Sept. 2004.

[32] R. W. Chang, "Orthogonal frequency division multiplexing," U.S. Patent 3 488 445, filed No. 14, 1966, issued Jan. 6, 1970.

[33] X. Chen, K. M. Chugg, and M. A. Neifeld, "Near-optimum parallel distributed data detection for page-oriented optical memories," *IEEE J. Sel. Topics Quantum Electron.*, vol. 4, no. 5, pp. 866-879, Sept./Oct. 1998.

[34] D. Chizhik, G. J. Foschini, and R. A. Valenzuela, "Capacities of multi-element transmit and receive antennas Correlations and keyholes," *Electron. Lett.*, vol. 36, no. 13, pp. 1099-1100, June 2000.

[35] D. Chizhik, "Slowing the time-fluctuating MIMO channel by beam forming," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1554-1565, Sept. 2004.

[36] C.-C. Chong, C.-M. Tan, D. I. Laurenson, S. McLaughlin, M. A. Beach, and A. R. Nix, "A new statistical wideband spatio-temporal channel model for 5-GHz band WLAN systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 139-150, Feb. 2003.

[37] P. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications,"*IEEE J. Sel. Areas Commun.*, vol. 9, no. 6, pp. 895-908, Aug. 1991.

[38] P. S. Chow, J. C. Tu, and J. M. Cioffi, "Performance evaluation of a multichannel transceiver system for ADSL and VHDSL services,"*IEEE J. Sel. Areas Commun.*, vol. 9, no. 6, pp. 909-919, Aug. 1991.

[39] R. D. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Schott, "A PRML system for digital magnetic recording,"*IEEE J. Sel. Areas Commun.*, vol. 10, no. 1, pp. 38-56, Jan. 1992.

[40] R. D. Cideciyan and W. Hirt, "Improving performance of PRML EPRML through noise prediction," *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 3968-3970, Sept. 1996.

[41] R. D. Cideciyan, J. D. Coker, E. Eleftheriou, and R. L. Galbraith, "Noise predictive maximum likelihood detection combined with parity-based post-processing," *IEEE Trans. Magn.*, vol. 37, no. 2, pp. 714-720, Mar. 2001.

[42] L. J. Cimini, Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, no. 7, pp. 665-675, July 1985.

[43] W. Coene, H. Pozidis, M. van Dijk, J. Kahlman, R. van Woudenberg, and B. Stek, "Channel coding and signal processing for optical recording systems beyond DVD," *IEEE Trans. Magn.*, vol. 37, no. 2, pp. 682-688, Mar. 2001.

[44] W. Coene, "Two dimensional optical storage," in *Proc. Intl. Conf. Optical Data Storage (ODS)*, Vancouver, Canada, May 2003, pp. 90-92.

[45] W. M. J. Coene, "Nonlinear signal-processing modal for scalar diffraction in optical recording," *Appl. Optics*, vol. 42, no. 32, pp. 6525-6535, Nov. 2003.

[46] W. M. J. Coene, "Partial response maximum likelihood (PRML) bit detection apparatus," U.S. Patent 6 580 766, June 17, 2003.

[47] J. D. Coker, E. Eleftheriou, R. L. Galbraith, and W. Hirt, "Noise-predictive maximum likelihood (NPML) detection," *IEEE Trans. Magn.*, vol. 34, no. 1, pp. 110-116, Jan. 1998.

[48] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223-229, Sept. 2002.

[49] K. Compaan and P. Kramer, "The Philips VLP system," *Philips. Tech. Rev.*, vol. 33, no. 7, pp. 178-180, 1973.

[50] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *J. Symbolic Comput.*, vol. 9, no. 3, pp. 251-280, Mar. 1990.

[51] M. O. Damen, K. A.-Meraim, and S. Burykh, "Iterative QR Detection for BLAST," *Wireless Pers. Commun.*, vol. 19, no. 3, pp. 179-191, Dec. 2001.

[52] L. Deneire, P. Vandenameele, L. van der Perre, B. Gyselinckx, and M. Engels, "A low-complexity ML channel estimator for OFDM," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 135-140, Feb. 2003.

[53] Z. Ding and Y. Li, *Blind equalization and identification*, Marcel Dekker, New York, NY, 2000.

[54] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, New York, NY, 1981, pp. 565-608.

[55] D. E. Dudgeon and R. M. Mersereau, *Multidimensional digital signal processing*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1984.

[56] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "Analysis of DFT-based channel estimators for OFDM," Research Rep., Lulea University of Technology, Sept. 1996.

[57] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931-939, July 1998.

[58] G. J. van den Enden, "Transmission system and recording system having a simplified symbol detector," European Patent EP0885499, Dec. 23, 1998.

[59] M. Enescu, T. Roman, and V. Koivunen, "Channel estimation and tracking in spatially correlated MIMO OFDM systems," in *Proc. IEEE Workshop on Statistical Signal Process.*, St. Louis, MO, Sept 2003, pp. 347-350.

[60] M. Engels (Editor), *Wireless OFDM systems: how to make them work?*, Kluwer Academic Publishers, Boston, MA, 2002.

[61] V. Erceg *et al.*, (2004, May), IEEE 802.11 document 03/940r4 "TGn Channel Models," [Online]. Available: ftp://ieee:wireless@ftp.802wire-lessworld.com/11/03/11-03-0940-04-000n-tgn-channel-models.doc

[62] R. B. Ert, P. Cardieri, K. W. Sowerby, T. S. Rapport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10-22, Feb. 1998.

[63] V. M. Eyuboğlu, and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. COM-36, no. 1, pp. 13-20, Jan. 1988.

[64] D. D. Falconer and J. F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, pp. 1541-1562, Nov. 1973.

[65] W. A. C. Fernando, and R. M. A. P. Rajatheva, "Performance of COFDM for LEO satellite channels in global mobile communications," in *Proc. IEEE Intl. Veh. Technol. Conf. (VTC)*, Ottawa, Canada, May 1998, pp. 1503-1507.

[66] W. A. C. Fernando and R. M. A. P. Rajatheva, "Performance of turbo and trellis coded OFDM for LEO satellite channels in global mobile communications," in *Proc. IEEE Intl. Conf. Commun. (ICC))*, Atlanta, GA, June 1998, pp. 412-416.

[67] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 363-378, May 1972.

[68] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311-335, May 1998.

[69] P. K. Frenger and N. A. B. Svensson, "Decision-directed coherent detection in multicarrier systems on Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, no. 2, pp. 490-498, Mar. 1999.

[70] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from high-order moments," *IEEE Trans. Autom. Control*, vol. AC-35, no. 1, pp. 27-35, Jan. 1990.

[71] A. Gabor, "Adaptive coding for self-clocking recording,"*IEEE Trans. Electron. Comput.*, vol. EC-16, pp. 866-868, Dec. 1967.

[72] D. Gesbert, H. Bölcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926-1934, Dec. 2002.

[73] D. Gesbert, M. Shafi, D.-S. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems,"*IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281-302, Apr. 2003.

[74] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. ASSP-37, no. 3, pp. 360-377, Mar. 1989.

[75] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1867-1875, Nov. 1980.

[76] S. Gopalaswamy and B. V. K. V. Kumar, "Multi-channel readback model for an optical tape system," in *Proc. IEEE Intl. Conf. Global Telecommun. (GLOBECOM)*, San Francisco, CA, Nov. 1994, pp. 1472-1476.

[77] S. Gopalaswamy, "Multi-channel signal processing for high track density optical recording systems," Ph. D. Thesis, Carnegie-Mellon University, May 1996.

[78] S. Gopalaswamy and B. V. K. V. Kumar, "Multichannel decision feedback equalizer for high track density in optical recording," *Opt. Eng.*, vol. 35, no. 8, pp. 2386-2395, Aug. 1996.

[79] S. Gopalaswamy, B. F. Boroujeny, and W. Y. H, Wilson, "Simple detection technique for d=2 modulated optical recording channels," in *Proc. IEEE Intl. Conf. Commun. (ICC)*, Vancouver, Canada, June 1999, pp. 1967-1971.

[80] F. Gross, *Smart antennas for wireless communications with MATLAB*, McGraw-Hill, New York, NY, 2005.

[81] A. D. Hallen and C. Heegard, “Delayed decision-feedback sequence estimation,” *IEEE Trans. Commun.*, vol. COM-37, no. 5, pp. 428-436, May. 1989.

[82] E. W. Hansen, “Fast Hankel transform algorithm,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 3, pp. 666-671, June 1985.

[83] S. Haykins (Editor), *Blind deconvolution*, Prentice Hall, Englewood Cliffs, NJ, 1994.

[84] J. F. Heanue, K. Gürkan, and L. Hesselink, “Decision feedback Viterbi detector for page-access optical memories,” U.S. Patent 5 740 184, Apr. 14, 1998.

[85] R. W. Heath, Jr. and G. B. Giannakis, “Exploiting input cyclostationarity for blind channel identification in OFDM systems,” *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 848-856, Mar. 1999.

[86] C. K. Ho, B. F.-Boroujeny, and F. Chin, “Added pilot semi-blind channel estimation scheme for OFDM in fading channels,” in *Proc. IEEE Intl. Conf. Global Telecommun. (GLOBECOM)*, San Antonio, TX, Nov 2001, pp. 3075-3079.

[87] P. Hoeher, “TCM on frequency-selective land-mobile fading channels,” in *Proc. Intl. 5th Tirrenia Workshop on Digital Commun.*, Tirrenia, Italy, Sept. 1991, pp. 317-328.

[88] P. Hoeher, S, Kaiser, and P. Robertson, “Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering,” in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Munich, Germany, Apr. 1997, pp. 1845-1848.

[89] P. Hoeher, S, Kaiser, and P. Robertson, “Two-dimensional pilot-symbol-aided channel estimation,” in *IEEE Intl. Symp. Inf. Theory (ISIT)*, Ulm, Germany, June-July 1997, pp. 123.

[90] P. Hoeher, S, Kaiser, and P. Robertson, “Pilot-symbol-aided channel estimation in time and frequency,” in *Proc. IEEE Intl. Conf. Global Telecommun. (GLOBECOM), Commun. Theory Mini-Conf.*, Phoenix, AZ, Nov. 1997, pp. 90-96.

[91] J. J. L. Horikx, K. Asato, N. Eguchi, and J. L. Bakx, “System parameters and detection margins of the multimedia compact disc system,” in *Proc. Intl. Conf. Optical Data Storage (ODS)*, San Diego, CA, July 1995, pp. 40-45.

[92] M.-H. Hsieh and C.-H. Wei, "Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Trans. Consum. Electron.*, vol. 44, no. 1, pp. 217-225, Feb. 1998.

[93] L. Huang, G. Mathew, and T. C. Chong, "Reduced Complexity Viterbi Detection for Two-Dimensional Optical Recording," *IEEE Trans. Consum. Electron.*, vol. 51, no. 1, pp. 123-129, Feb. 2005.

[94] L. Huang, G. Mathew, and T. C. Chong, "Channel modeling and target design for two-dimensional optical storage systems," *IEEE Trans. Magn.*, vol. 41, no. 8, pp. 2414-2424, Aug. 2005.

[95] L. Huang, G. Mathew, and J. W. M. Bergmans, "Pilot-aided channel estimation for systems with virtual carriers," in *Proc. Intl. Conf. Commun (ICC)*, Istanbul, Turkey, June 2006, pp. 3070-3075.

[96] L. Huang, J. W. M. Bergmans, and F. M. J. Willems, "Low-complexity LMMSE-based MIMO-OFDM channel estimation via angle-domain processing," accepted in *IEEE Trans. Signal Process.*, Mar 2007.

[97] L. Huang, C. K. Ho, J. W. M. Bergmans, and F. M. J. Willems, "Pilot-aided angle-domain channel estimation techniques for MIMO-OFDM systems," accepted in *IEEE Trans. Veh. Technol.*, June 2007.

[98] X. Huang, H. Bi, and S. Yu, "A new subspace algorithm of blind channel estimation for OFDM systems," in *Proc. IEEE Intl. Symp. Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Beijing, China, Sept. 2003, pp. 1105-1108.

[99] P. J. Huber, *Robust statistics*, Wiley, New York, NY, 1981.

[100] M. Ibnkahla (Editor), *Signal processing for mobile communications handbook*, CRC Press, Boca Raton, FL, 2005.

[101] K. A. S. Immink, "EFMPlus: the coding format of the multimedia compact disc,"*IEEE Trans. Consum. Electron.*, vol. 41, no. 3, pp. 491-497, Aug. 1995.

[102] K. A. S. Immink, *Codes for mass data storage systems*, Shannon Foundation Publishers, Eindhoven, the Netherlands, 1999.

[103] A. H. J. Immink, W. M. J. Coene, A. M. van der Lee, C. Busch, A. P. Hekstra, J. W. M. Bergmans, J. Riani, S. J. L. v. Beneden, and T. Conway, "Signal processing and coding for two-dimensional optical storage," in *Proc. IEEE Intl. Conf. Global Telecommun. (GLOBECOM)*, San Francisco, CA, Dec 2003, pp. 3904-3908.

[104] K. C. Indukumar, "A novel post-processing scheme for threshold detection on $d = 2$ optical channels," in *Proc. IEEE Intl. Conf. Global Telecommun. (GLOBECOM)*, San Antonio, Texas, Nov. 2001, pp. 1479-1483.

[105] T. Kaiser, "When will smart antennas be ready for the market? Part I," *IEEE Signal Process. Mag.*, vol. 22, no. 2, pp. 87-92, Mar. 2005.

[106] T. Kaiser, "When will smart antennas be ready for the market? Part II - results," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 174-176, Nov. 2005.

[107] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2406-2413, July 1994.

[108] T. Kato, S. Taira, T. Maeda, Y. Katayama, and T. Nishiya, "Two-dimensional run-length-limited code and partial response maximum likelihood system with multi-track recording," in *Joint Intl. Symp. Optical Memory (ISOM) and Optical Data Storage (ODS)*, Waikoloa, Hawaii , July 2002, pp. 51-53.

[109] A. Kavčić and J. M. F. Moura, "Correlation-sensitive adaptive sequence detection," *IEEE Trans. Magn.*, vol. 34, no. 3, pp. 763-771, May 1998.

[110] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice Hall, Englewood Cliffs, NY, 1993.

[111] J. G. Kenney, L. R. Carley, and R. W. Wood, "Multi-level decision feedback equalization for saturation recording," *IEEE Trans. Magn.*, vol. 29, no. 3, pp. 2160-2171, July 1993.

[112] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation,"*IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211-1226, Aug. 2002.

[113] J. G. Kim and J. T. Lim, "MAP based channel estimation for MIMO-OFDM over fast rayleigh fading channels," accepted in *IEEE Trans. Veh. Technol.*, Aug. 2007.

[114] B. M. King and M. A. Neifield, "Parallel detection algorithm for page-oriented optical memories," *Appl. Opt.*, vol. 37, no. 26, pp. 6275-6298, Sept. 1998.

[115] L. Knockaert, "Fast Hankel transform by fast sine and cosine transforms: the Mellin connection," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1695-1701, Jun. 2000.

[116] E. Kurtas, J. G. Proakis, and M. Salehi, "Reduced complexity maximum likelihood sequence estimation for multitrack high-density magnetic recording channel," *IEEE Trans. Magn.*, vol. 35, no. 4, pp. 2187-2193, July 1999.

[117] M. Kuzaka, "An overview of Blu-ray disc application format for HD-movie distribution," in *Proc. Intl. Conf. Optical Data Storage (ODS)*, Monterey, CA, Apr. 2004, pp. 7-9.

[118] I. Lee and J. M. Cioffi, "Equalized maximum likelihood receiver with a unit energy constraint," *IEEE Trans. Magn.*, vol. 33, no. 1, pp. 855-862, Jan. 1997.

[119] E. A. Lee and D. G. Messerschmitt, *Digital communication*, 2nd ed., Kluwer Academic Publishers, Boston, 1999.

[120] J. Lee, J. W. Lee, E. J. Ryu, E. S. Cho, M. Konakov, J. S. Shim, and H. S. Park, "Advanced PRML data detector for high density recording," in *Proc. Intl. Conf. Optical Data Storage (ODS)*, Monterey, CA, Apr. 2004, pp. 234-236.

[121] C. Li and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 141-150, Jan. 2003.

[122] L. Li, "Adaptive signal processing for equalization design in hard disk drives in the presence of media noise," B. Eng. Thesis, National University of Singapore, Apr. 2004.

[123] X. Li and X. Cao, "Low complexity signal detection algorithm for MIMO-OFDM systems," *Electron. Lett.*, vol. 41, no. 2, pp. 287-288, Jan. 2005.

[124] Y. (G.) Li, L. J. Cimini, Jr., and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902-915, July 1998.

[125] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207-1215, July 2000.

[126] Y. (G.) Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for wireless communications: signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1471-1477, Sept. 2002.

[127] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: from cyclostationarity to subspaces," *Signal Process.*, vol. 50, no. 1-2, pp. 83-99, Apr. 1996.

[128] B. Liu, R. Jin, and Y. Fan, "Modified pre-FFT OFDM adaptive antenna array with beam-space channel estimation," *Electron. Lett.*, vol. 40, no. 5, pp. 287-288, Mar. 2004.

[129] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering and array processing*, McGraw-Hill, New York, NY, 2000.

[130] G. Mathew, B. F. Boroujeny, and R. W. Wood, "Design of multilevel decision feedback equalizers," *IEEE Trans. Magn.*, vol. 33, no. 6, pp. 4528-4542, Nov. 1997.

[131] S. W. McLaughlin, "Shedding light on the future of sp for optical recording," *IEEE Signal Process. Mag.*, vol. 15, no. 4, pp. 83-94, July 1998.

[132] X. Mestre, J. R. Fonollosa, and A. P. Zamora, "Capacity of MIMO channels: asymptotic evaluation under correlated fading," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 829-838, June 2003.

[133] H. Miao and M. J. Juntti, "Spatial signature and channel estimation for wireless MIMO-OFDM systems with spatial correlation," in *Proc. IEEE Workshop on Signal Process. Adv. in Wireless Commun.*, Lisbon, Portugal, July 2004, pp. 522-526.

[134] M. Minsky, "Microscopy apparatus," U.S. Patent 3 013 467, Dec. 19, 1961.

[135] A. Moinian, L. Fagoonee, B. Honary, and W. Coene, "Linear channel model for multilevel two dimensional optical storage," in *Proc. 7th Intl. Symp. Commun. Theory and Applicat.*, Ambleside, Lake District, U.K., July 2003, pp. 352-356.

[136] J. Moon and L. R. Carley, "Performance comparison of detection methods in magnetic recording," *IEEE Trans. Magn.*, vol. 26, no. 6, pp. 3155-3172, Nov. 1990.

[137] J. Moon and W. Zeng, "Equalization for maximum likelihood detectors," *IEEE Trans. Magn.*, vol. 31, no. 2, pp. 1083-1088, Mar. 1995.

[138] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3065-3073, Dec. 2001.

[139] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 516-525, Feb. 1995.

[140] B. Muquet and M. de Courville, "Blind and semi-blind channel identification methods using second order statistics," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Phoenix, AZ, Mar. 1999, pp. 2745-2748.

[141] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699-1712, July 2002.

[142] T. Nakagawa, H. Ino, and Y. Shimpuku, "A simple detection method for RLL codes (run detector)," *IEEE Trans. Magn.*, vol. 33, no. 5, pp. 3262-3264, Sept. 1997.

[143] T. Narahara, S. Kobayashi, M. Hattori, Y. Shimpuku, G. v. d. Enden, J. kahlman, M. v. Dijk, and R. v. Wouden, "Optical disc system for digital video recording," *Jpn. J. Appl. Phys.*, pt. 1, vol. 39, no. 2B, pp. 912-919, Feb. 2000.

[144] M. C. Necker and G. L. Stüber, "Totally blind channel estimation for OFDM over fast varying mobile channels," in *Proc. IEEE Intl. Conf. Commun. (ICC)*, New York, NY, May 2002, pp. 421-425.

[145] R. van Nee, G. Awater, M. Morikura, H. Takanashi, and M. Webster, "New high rate wireless LAN standards," *IEEE Commun. Mag.*, vol. 37, no. 12, pp. 82-88, Dec. 1999.

[146] C. L. Nikias and A. P. Petropulu, *Higher-order spectral analysis: a nonlinear signal processing framework*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[147] R. Nilsson, O. Edfors, M. Sandell, and P. O. Börjesson, "An analysis of two-dimensional pilot-symbol assisted modulation for OFDM," in *Proc. IEEE Intl. Conf. Pers. Wireless Commun. (ICPWC)*, Mumbai (Bombay), India, Dec. 1997, pp. 17-19.

[148] T. Ojanperä, and R. van Nee, "An overview of air interface multiple access for IMT-2000/UMTS," *IEEE Commun. Mag.*, vol. 36, no. 9, pp. 82-86, 91-95, Sept. 1998.

[149] D. Pal, "Fractionally spaced semi-blind equalization of wireless channels," in *Proc. Intl. Asilomar Conf. Signals, Systems and Comput.*, Pacific Grove, CA, Oct. 1992, pp. 642-645.

[150] A. Papoulis, *Systems and transforms with application in optics*, McGraw-Hill, New York, 1968.

[151] J. B. H. Peek, "Communications aspects of the compact disk digital audio system," *IEEE Commun. Mag.*, vol. 23, no. 2, pp. 7-15, Feb. 1985.

[152] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a 'stop and go' decision directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, no. 9, pp. 877-887, Sept. 1987.

[153] H. Pozidis and W. M. J. Coene, "A low-complexity and high-speed detection method for high-density $d = 1$ optical recording," *IEEE Trans. Consum. Electron.*, vol. 46, no. 3, pp. 562-571, Aug. 2000.

[154] J. G. Proakis, *Digital communications*, 4th ed., McGraw-Hill, New York, NY, 2001.

[155] U. Reimers, "Digital video broadcasting," *IEEE Commun. Mag.*, vol. 36, no. 6, pp. 104-110, June 1998.

[156] J. Riani, J. W. M. Bergmans, S. J. L. v. Beneden, W. M. J. Coene, and A. H. J. Immink, "Equalization and target response optimization for high density two-dimensional optical storage," in *Proc. 24th Symp. Inf. Theory in the Benelux*, Veldhoven, May 2003, pp. 141-148.

[157] P. Rice, A. Macovski, E. D. Jones, H. Frohbach, R. W. Crews, and A. W. Noon, "An experimental television recording and playback system using photographic discs," *J. Soc. Motion, Pict. Telev. Eng.*, vol. 79, no. 11, pp. 997-1002, Nov. 1970.

[158] J. Rinne and M. Renfors, "Pilot spacing in orthogonal frequency division multiplexing systems on practical channels," *IEEE Trans. Consum. Electron.*, vol. 42, no. 4, pp. 959-962, Nov. 1996.

[159] S. Roy and C. Li, "A subspace blind channel estimation method for OFDM systems without cyclic prefix," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 572-579, Oct. 2002.

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

[161] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Trans. Commun.*, vol. COM-23, no. 6, pp. 679-682, June 1975.

[162] A. M. Sayeed, “Deconstructing multiantenna fading channels,” *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563-2579, Oct. 2002.

[163] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, “A fourth-generation MIMO-OFDM broadband wireless system: design, performance, and field trial results,” *IEEE Commun. Mag.*, vol. 40, no. 9, pp. 143-149, Sept. 2002.

[164] M. Sandell and O. Edfors, “A comparative study of pilot-based channel estimators for wireless OFDM,” Research Rep., Lulea University of Technology, Sept. 1996.

[165] O. Shalvi and E. Weinstein, “Super-exponential methods for blind deconvolution,” *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 504-519, Mar. 1993.

[166] H. Shin, G. J. Foschini, M. J. Gans, and J. M. Kahn, “Fading correlation and its effect on the capacity of multielement antenna systems,” *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502-513, Mar. 2000.

[167] H. Shin and J. H. Lee, “Capacity of multiple-antenna fading channels spatial fading correlation, double scattering, and keyhole,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2636-2647, Oct. 2003.

[168] M. Shin, H. Lee, and C. Lee, “Enhanced channel-estimation technique for MIMO-OFDM systems,” *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 261-265, Jan. 2004.

[169] N. Singla, J. A. O’Sullivan, R. S. Indeck, and Y. Wu, “Iterative decoding and equalization for 2-D recording Channels,” *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2328-2330, Sept. 2002.

[170] C. Snow, L. Lampe, and R. Schober, “Error rate analysis for coded multicarrier systems over quasi-static fading channels,” in *Proc. IEEE Intl. Conf. Global Telecommun. (GLOBECOM)*, San Francisco, CA, Nov. 2006, pp. 1-5.

[171] E. Soljanin, “Coding for improving noise immunity in multi-track, multi-head recording systems,” Ph. D. Thesis, Texas A&M University, Dec. 1994.

[172] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, "Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 347-360, Mar. 2000.

[173] G. L. Stüber, J. R. Barry, S. W. Mclaughlin, Y. (G.) Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271-294, Feb. 2004.

[174] S. Sun, I. Wiemer, C. K. Ho, and T. T. Tjhung, "Training sequence assisted channel estimation for MIMO OFDM," in *IEEE Wireless Commun. and Netw. Conf. (WCNC)*, New Orleans, LA, Mar. 2003, pp. 38-43.

[175] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585-595, Nov. 1999.

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

[177] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998.

[178] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. Intl. Asilomar Conf. Signals, Systems and Comput.*, Pacific Grove, CA, Nov. 1991, pp. 856-860.

[179] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340-349, Mar. 1994.

[180] L. Tong and S. Perreau, " Multichannel blind identification: from subspace to maximum likelihood methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951-1968, Oct. 1998.

[181] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. ASSP-31, no. 2, pp. 459-472, Apr. 1983.

[182] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, Cambridge, UK, 2005.

[183] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proc. IEEE Intl. Veh. Technol. Conf. (VTC)*, Phoenix, AZ, May 1997, pp. 1639-1643.

[184] J. Tugnait, "Identification of linear stochastic systems via second- and fourth-order cumulant matching," *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 393-407, May 1987.

[185] P. Vandenameele, L. van der Perre, M. Engels, B. Gyselinckx, and H. J. de Man,"A combined OFDM/SDMA approach,"*IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2312-2321, Nov. 2000.

[186] P. A. Voois and J. M. Cioffi, "Multichannel signal processing for multiple-head digital magnetic recording," *IEEE Trans. Magn.*, vol. 30, no. 6, pp. 5100-5114, Nov. 1994.

[187] J. S. Walker, *Fast Fourier transforms, Second Edition*, CRC Press, Boca Raton, FL, 1996, Chap. 3.

[188] J. J. Wang, "Development of timing recovery techniques for digital recording channels," Ph. D. Thesis, National University of Singapore, 2002.

[189] H. Wang, Y. Lin, and B. Chen, "Data-efficient blind OFDM channel estimation using receiver diversity," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2613-2623, Oct. 2003.

[190] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. COM-19, no. 5, pp. 628-634, Oct. 1971.

[191] S. K. Wilson, "Digital audio broadcasting in a fading and dispersive channel," Ph. D. Thesis, Stanford University, Aug. 1994.

[192] Y. Wu, J. A. O'Sullivan, N. Singla, and R. S. Indeck, "Iterative detection and decoding for separable two-dimensional intersymbol interference," *IEEE Trans. Magn.*, vol. 39, no. 4, pp. 2115-2120, July 2003.

[193] D. Wübben, R. Böhnke, V. Kühn, and k. D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," in *Proc. IEEE Intl. Veh. Technol. Conf. (VTC)*, Orlando, FL, May 2003, pp. 508-512.

[194] D. Wübben and K. D. Kammeyer, "Low complexity successive interference cancellation for per-antenna-coded MIMO-OFDM schemes by applying parallel-SQRD," in *Proc. IEEE Intl. Veh. Technol. Conf. (VTC)*, Melbourne, Australia, May 2006.

[195] B. Yang, Z. Cao, and K. B. Letaief, "Analysis of low-complexity windowed DFT-based MMSE channel estimator for OFDM systems," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1977-1987, Nov. 2001.

[196] Y.-H. Yeh and S. G. Chen, "Efficient channel estimation based on discrete cosine transform," in *Proc. IEEE Intl. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Hong Kong, Apr. 2003, pp. 676-679.

[197] Y.-H. Yeh and S. G. Chen, "DCT-based channel estimation for OFDM systems," in *Proc. IEEE Intl. Conf. Commun. (ICC)*, Paris, France, June 2004, pp. 2442-2446.

[198] A. v. Zelst and T. C. W. Schenk, "Implementation of MIMO OFDM-based wireless LAN system," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 483-494, Feb. 2004.

[199] W. Zeng and J. Moon, "Modified Viterbi algorithm for a jitter-dominant $1 - D^2$ channel," *IEEE Trans. Magn.*, vol. 28, no. 5, pp. 2895-2897, Sept. 1992.

[200] H. Zhang, Y. (G.) Li, A. Reid, and J. Terry, "Channel estimation for MIMO OFDM in correlated fading channel," in *Proc. IEEE Intl. Conf. Commun. (ICC)*, Seoul, Korea, May 2005, pp. 2626-2630.

[201] J.-K. Zhang, A. Kavčić, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 154-172, Jan. 2005.

[202] Y. Zhao and A. Huang, "A novel channel estimation method for OFDM mobile communication systems based on pilot signals and transform-domain processing," in *Proc. IEEE Intl. Veh. Technol. Conf. (VTC)*, Phoenix, AZ, May 1997, pp. 2089-2093.

[203] W. Y. Zou and Y. Wu, "COFDM: an overview," *IEEE Trans. Broadcast.*, vol. 41, no. 1, pp. 1-8, Mar. 1995.

Author's Publications

1. L. Huang, C. K. Ho, J. W. M. Bergmans, and F. M. J. Willems, "Pilot-aided angle-domain channel estimation techniques for MIMO-OFDM systems," accepted in *IEEE Trans. Veh. Technol.*, June 2007.

2. L. Huang, J. W. M. Bergmans, and F. M. J. Willems, "Low-complexity LMMSE-based MIMO-OFDM channel estimation via angle-domain processing," accepted in *IEEE Trans. Signal Process.*, Mar. 2007.

3. L. Huang, J. W. M. Bergmans, and F. M. J. Willems, "Low-complexity LMMSE-based MIMO-OFDM channel estimation via angle-domain processing," accepted in *IEEE Trans. Signal Process.*, Mar. 2007.

4. L. Huang, G. Mathew, and J. W. M. Bergmans, "LMMSE-based angle-domain MIMO-OFDM channel estimation," in *IEEE BENELUX Signal Process. Symp. and DSP Valley's Annual Research & Technol. Symp. (SPS-DARTS)*, Antwerp, Belgium, Mar. 2007, pp. 199-204.

5. L. Huang, G. Mathew, and T. C. Chong, "Channel modeling and target design for two-dimensional optical storage systems," *IEEE Trans. Magn.*, vol. 41, no. 8, pp. 2414-2424, Aug. 2005.

6. L. Huang, G. Mathew, and T. C. Chong, "Reduced complexity Viterbi detection for two-dimensional optical recording," *IEEE Trans. Consum. Electron.*, vol. 51, no. 1, pp. 123-129, Feb. 2005.

7. L. Huang, G. Mathew, and T. C. Chong, "Reduced complexity two-dimensional Viterbi-like detectors," in *Proc. Intl. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, Jan. 2005, pp. 47-48.

8. L. Huang and T. C. Chong, "Novel target design algorithm for two-dimensional optical storage (TwoDOS)," *Proc. SPIE*, vol. 5380, pp. 105-115, Sept. 2004.

9. L. Huang and T. C. Chong, "Reduced-complexity target design for two-dimensional optical storage (TwoDOS)," in *Proc. Intl. Conf. Optical Data Storage (ODS)*,

Monterey, CA, Apr. 2004, pp. 46-48.

Curriculum Vitae

Li Huang was born in Wuhan, China, on July 23, 1980. He received his B.Eng degree in Electronics and Information Engineering in 2002 from Huazhong University of Science & Technology, Wuhan, China. After 2002, he was initially enrolled in a M. Eng program in National University of Singapore (NUS) under the Data Storage Institute (DSI) Scholarship, and was later selected to be part of the joint Ph.D. program between NUS and Technical University Eindhoven (TU/e), the Netherlands, in 2004 under the Design Technology Institute (DTI) Scholarship.

During his stay in DSI from 2002 to 2005, he focused on the development of reduced-complexity signal processing techniques for data storage systems. In view of the significance of his contributions to the two-dimensional optical storage (TwoDOS) system, he was awarded the Most Outstanding Student Award of the year by DSI in 2005. Since his attachment in Institute for Infocomm Research (I2R) between August and December in 2004, he has been involved in developing channel estimation techniques for wireless communication systems. From 2005, he continued his Ph.D. research with the focus on reduced-complexity signal processing techniques for wireless communication systems in the Signal Processing Systems Group in the Department of Electrical Engineering of TU/e.