

OFDM Transmission

over Rapidly Changing Channels

Zijian Tang

OFDM Transmission over Rapidly Changing Channels

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op dinsdag 20 november 2007 om 10.00 uur
door

Zijian Tang

elektrotechnisch ingenieur
geboren te Sjanghai.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. A.-J. van der Veen

Toegevoegd promotor:

Dr. ir. G. Leus

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. A.-J. van der Veen	Technische Universiteit Delft, promotor
Dr. ir. G. Leus	Technische Universiteit Delft, toegevoegd promotor
Prof. dr. ir. P.M. Dewilde	Technische Universiteit Delft
Prof. ir. P. van Genderen	Technische Universiteit Delft
Prof. dr. ir. J.-P. Linnartz	Technische Universiteit Eindhoven
Prof. dr. ir. M. Moonen	Katholieke Universiteit Leuven
Dr. P. Banelli	Università di Perugia

Copyright © 2007 by Zijian Tang

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author.

ISBN 978-90-9022298-1

To my grandmother, my parents, Ying and Byron

Contents

Glossary	vii
1 Introduction	1
1.1 Wireless Fading Channels	2
1.1.1 Mobile Radio Propagation Characteristics	2
1.1.2 Channel Variation due to the Doppler Effect	3
1.1.3 Impulse Response Model of the Channel and Its Approximation	5
1.2 Orthogonal Frequency-Division Multiplexing	6
1.2.1 Overview of OFDM	7
1.2.2 Challenges due to the Fast Fading Channel in OFDM	9
1.3 Thesis Survey and Contributions	11
2 Preliminaries	15
2.1 A Digital Wireless Communication System Architecture	15
2.1.1 Transmitter Structure	16
2.1.2 Receiver Structure	17
2.1.3 Channel and Discrete Baseband Model	18
2.2 Channel Time-Variation and BEM Approximation	20
2.3 Block Transmission Techniques	26
2.3.1 CP-OFDM System Model	26
2.3.2 Other OFDM System Models	30
2.3.3 Single-carrier Transmission	33
3 Channel Equalization in an OFDM System	35
3.1 Introduction	35
3.2 System Model	36

3.3	A Low-Complexity Equalization Scheme	37
3.4	Window Design	40
3.5	Simulation Results	44
3.6	Summary	46
3A	Proof of Proposition 3.1	47
3B	Proof of Theorem 3.1	47
3C	Proof of (3.33)	48
3D	Equivalence of the Proposed Window and the Window of [90]	49
4	Channel Estimation Based on a Single OFDM Symbol	51
4.1	Introduction	51
4.2	System Model	52
4.2.1	OFDM System Model	52
4.2.2	BEM Model in the Presence of a Window	53
4.2.3	OFDM System Model in Light of BEM	54
4.3	Data Model for Channel Estimation	55
4.4	Channel Estimation and B_c Optimization	60
4.4.1	The LMMSE Estimator	60
4.4.2	The Least Squares Estimator	62
4.4.3	An Iterative BLUE	63
4.4.4	Optimization of B_c	64
4.5	Simulations	66
4.6	Summary	72
4A	Derivation of Covariance Matrices $\mathbf{R}_n^{\{O\}}$, \mathbf{R}_d and $\tilde{\mathbf{R}}_d(\mathbf{c})$	73
4B	Cramer-Rao Bound	76
5	Channel Estimation based on Multiple OFDM Symbols	79
5.1	Introduction	79
5.2	System Model and BEM for Multiple OFDM Symbols	81
5.3	Channel Estimators and Pilot Schemes	83
5.3.1	Data Model for Channel Estimation	83
5.3.2	Channel Estimators	85
5.4	Channel Identifiability	86
5.4.1	Rank Condition of $\mathcal{P}(i)$	87
5.4.2	Rank Condition of \mathcal{P}	89
5.4.3	Channel Identifiability at a High SNR	90
5.5	Some Simulation Results and Discussion	92
5.6	Summary	98
5A	Rank Condition of $\mathbf{Z}(i)$	99

Contents

5B	Proof of Lemma 5.1	103
5C	Proof of Lemma 5.2	103
6	Channel Equalization in a Single-Carrier System	105
6.1	Introduction	105
6.2	System Model	107
6.3	FD Equalization Based on the ODM	107
6.3.1	Equalization Scheme	107
6.3.2	Window Design for the ODM	112
6.4	FD Equalization Based on the EDM	115
6.4.1	Equalization Scheme	115
6.4.2	Window Design for the EDM	119
6.5	Numerical results	123
6.6	Summary	126
6A	Proof of Theorem 6.1	128
6B	Proof of Theorem 6.2	129
6C	Proof of Theorem 6.3	131
7	Channel Estimation in a Single-Carrier System	133
7.1	Introduction	133
7.2	System Model	134
7.3	Data Model for Channel Estimation	136
7.4	Channel Estimation	138
7.4.1	The LMMSE Estimator	139
7.4.2	The Least Squares Estimator	140
7.4.3	An Iterative BLUE	140
7.5	Channel Identifiability	141
7.6	Simulation Results	144
7.7	Summary	145
7A	Proof of Theorem 7.1	147
8	MSE-Optimal Training for Time-Selective Channels	151
8.1	Introduction	151
8.2	Data Model for Time-Selective Channel Estimation	152
8.3	Optimization Algorithm	155
8.3.1	Solving the Primal Problem	155
8.3.2	Solving the Master Problem	156
8.3.3	Imposing Symmetric Constraint	158
8.4	Numerical Examples	160
8.5	Summary	163

8A	Proof of Theorem 8.1	163
8B	Proof of Property 8.1	164
8C	Proof of Property 8.2	165
9	Conclusions and Future Work	167
9.1	Conclusions	167
9.2	Future Research	169
	Bibliography	171
	Summary	181
	Samenvatting	183
	Acknowledgment	185
	Curriculum Vitae	187

Glossary

Mathematical Notation

x	scalar x
\mathbf{x}	vector \mathbf{x}
\mathbf{X}	matrix \mathbf{X}
\mathbf{X}^T	transpose of matrix \mathbf{X}
\mathbf{X}^H	Hermitian transpose of matrix \mathbf{X}
\mathbf{X}^*	complex conjugate of matrix \mathbf{X}
\mathbf{X}^{-1}	inverse of matrix \mathbf{X}
\mathbf{X}^\dagger	pseudoinverse of matrix \mathbf{X}
$\text{tr}\{\mathbf{X}\}$	trace of matrix \mathbf{X}
$\ \mathbf{X}\ $	Frobenius norm of matrix \mathbf{X}
$\mathcal{D}\{\mathbf{x}\}$	square diagonal matrix with \mathbf{x} as diagonal
$[\mathbf{X}]_{k,l}$	element on the k th row and l th column of matrix \mathbf{X}
$[\mathbf{X}]_{k:l,:}$	rows k up to l of matrix \mathbf{X}
$[\mathbf{X}]_{:,k:l}$	columns k up to l of matrix \mathbf{X}
$\mathbf{0}_{m \times n}$	$m \times n$ all-zero matrix
$\mathbf{1}_{m \times n}$	$m \times n$ all-one matrix
\mathbf{e}_n	unit vector with a one in the n th entry
\mathbf{I}_N	identity matrix of size N
\mathbf{W}_N	N -point unitary DFT matrix
$\Re\{x\}$	real part of x
$\Im\{x\}$	imaginary part of x
\hat{x}	estimate of x
\mathcal{S}	set \mathcal{S}
$ x $	absolute value of x

$ \mathcal{S} $	cardinality of set \mathcal{S}
$\mathbf{X}^{\{\mathcal{S}_r, \mathcal{S}_c\}}$	intersection of rows and columns of matrix, for which the row indices are collected in set \mathcal{S}_r and the column indices are collected in set \mathcal{S}_c
$\mathbf{X}^{\{\mathcal{S}_r\}}$	rows of matrix \mathbf{X} with row indices collected in set \mathcal{S}_r
$\lfloor x \rfloor$	largest integer smaller or equal to $x \in \mathbb{R}$
$\lceil x \rceil$	smallest integer larger or equal to $x \in \mathbb{R}$
$\mathcal{E}_\theta\{x\}$	expectation of random variable x with respect to θ
$\text{mod}(x, y)$	remainder after dividing $x \in \mathbb{R}$ by $y \in \mathbb{R}$
\mathbb{R}	the set of real numbers
\mathbb{C}	the set of complex numbers
\otimes	Kronecker product
\odot	Hadamard (point-wise) product

Acronyms and Abbreviations

ADSL	Asymmetrical Digital Subscriber Lines
AWGN	Additive White Gaussian Noise
BEM	Basis Expansion Model
BER	Bit Error Rate
BLUE	Best Linear Unbiased Estimator
BPSK	Binary Phase Shift Keying
CE-BEM	Complex Exponential BEM
(C)CE-BEM	Critically-sampled CE-BEM
CP	Cyclic Prefix
CSI	Channel State Information
CRM	Cramer Rao Bound
DC	Direct Current
DFE	Decision Feedback Equalizer
DFT	Discrete Fourier Transformation
DKL-BEM	Discrete Karhuen-Loéve BEM
DMT	Discrete Multi-Tone
DPS-BEM	Discrete Prolate Spheroidal BEM
DVB-T	Digital Video Broadcasting - Terrestrial
FD	Frequency-Domain
FDM	Frequency-Division Multiplexing
FFT	Fast Fourier Transformation
FIM	Fisher Information Matrix
FIR	Finite Impulse Response

GA	Generic Algorithm
GBD	Generalized Benders Decomposition
IBI	Inter-Block Interference
ICI	Inter-Carrier Interference
I/O	Input-Output
IDFT	Inverse Discrete Fourier Transformation
IFFT	Inverse Fast Fourier Transformation
ISI	Inter-Symbol Interference
KSP	Known Symbol Padding
LHS	Left-Hand Side
LMMSE	Linear Minimum Mean Square Error
LOS	Line-Of-Sight
LS	Least Squares
MAC	Multiply and Accumulate operation
MIMO	Multi-Input Multi-Output
MSE	Mean Squared Error
NLMS	Normalized Least Mean Squares
NMSE	Normalized Mean Squared Error
(O)CE-BEM	Oversampled CE-BEM
OFDM	Orthogonal Frequency-Division Multiplexing
PAR	Power-to-Average Ratio
P-BEM	Polynomial BEM
PDF	Probability Distribution Function
QPSK	Quadrature Phase Shift Keying
RHS	Righ-Hand Side
RLS	Recursive Least Squares
SIMO	Single-Input Multi-Output
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SOS	Second-Order Statistics
s.t.	subject to
SVD	Singular Value Decomposition
TD	Time-Domain
TI	Time-Invariant
TV	Time-Varying
VDSL	Very high data rate Digital Subscriber Lines
WIMAX	Worldwide Interoperability Microwave Access
ZF	Zero-Forcing
ZP	Zero-Padding

Chapter 1

Introduction

In the past decades, communication systems have been undergoing a fundamental revolution. One of the most spectacular events is that the traditional wired connection is, to a great extent if not completely, being replaced by the wireless connection at an exponential speed.

Wireless communication systems gained enormous popularity over wired systems mainly because they are in many applications much cheaper to implement. Besides, in places where the environment hampers cable deployment, radio connection remains the only means for communication. Last but not least, wireless communication systems allow for mobile users. On the other side, it is more challenging for a system engineer to maintain reliable communications over wireless (radio signal) channels than over wired channels. The propagation paths are in most situations hostile and unpredictable, and susceptible to even a small change in the environment. Signal processing techniques are therefore playing an extremely important role to overcome these problems.

One of the most profound achievements of signal processing on the physical layer might be orthogonal frequency division multiplexing (OFDM), which will be the major research subject in this thesis. OFDM proved to be a superior solution against multipath propagation, and is adopted in many contemporary communication protocols. Despite its numerous merits, the performance of OFDM is far less satisfactory in a high mobility communication scenario, where the so-called Doppler effect plays a significant role. In that case, traditional techniques, which are successfully utilized for channel estimation or equalization in a static environment, will function incompetently or inflict more hardware expenses. Seeing the rapidly increasing demand for high mobility communications, it is the imperative target of this thesis to provide some effective and yet affordable solutions for these problems.

In the coming sections, we will first give a brief description of the radio characteristics of the channel, with focus on the Doppler effect due to mobility. Afterwards, a quick overview of the OFDM system will be given. At last, we will sketch the scope of this thesis.

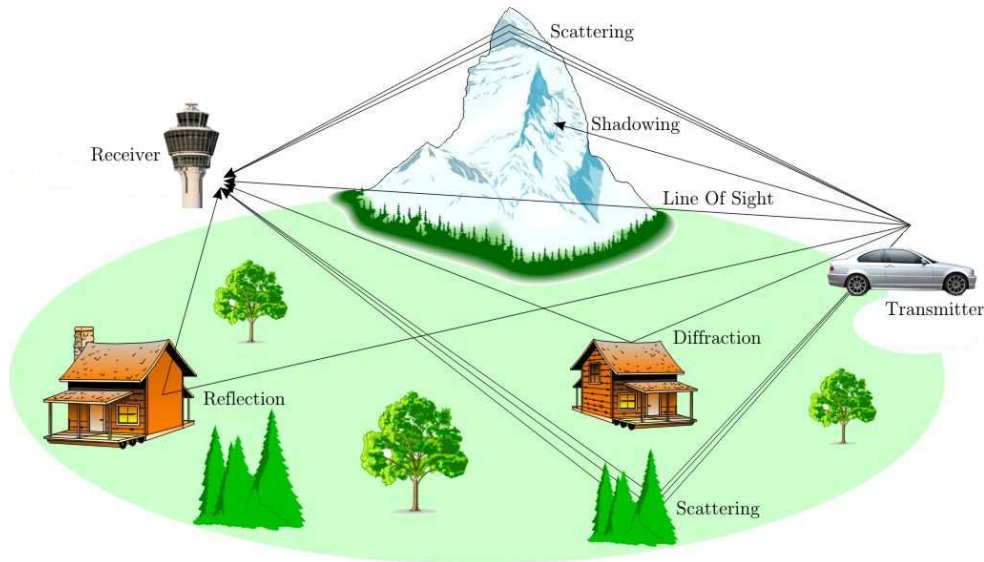


Figure 1.1: A typical transmission scenario.

1.1 Wireless Fading Channels

1.1.1 Mobile Radio Propagation Characteristics

Mobile communication channels are characterized by two types of fading: large-scale fading and small-scale fading. The latter is also simply referred to as ‘fading’ [77].

Large-scale fading represents the average signal power attenuation of the path loss, which is mainly affected by prominent terrain contours, e.g., hills, forests, etc., between the transmitter and receiver. The statistics of large-scale fading provide a means to compute an estimate of path loss as a function of distance. They are in general relatively constant over time.

Small-scale fading is the major research interest in this thesis, which refers to the dramatic changes in signal amplitude and phase over a short period of time. Usually, the same radio wave sent by the transmitter travels in multiple paths, and reaches the receiver at different time with different attenuation. As a result, the received signal is a combination of these so-called multipath signals, which is very sensitive for changes in the environment. For instance, depending on the distribution of the intensity and relative propagation time of the radio waves and the

bandwidth of the transmitted signal, even a slight variation in the spatial positioning between the transmitter and receiver can give rise to interference. Statistically, the envelope of signals experiencing small-scale fading can be described by either the Rayleigh probability density function (PDF) or the Rician PDF [76]. The former is used when there are a large number of multiple reflective paths, and there is no line-of-sight (LOS) component, while the latter is used when there is a dominant non-fading component present, such as an LOS path.

Three basic mechanisms impact the signal propagation in a wireless communication system: reflection, diffraction and scattering. The channel undergoing these physical phenomena manifests itself in the form of multipath. A typical transmission scenario is illustrated in Fig. 1.1.

As suggested in the figure, reflection occurs when a propagating electromagnetic wave impinges upon an object, which has very large dimensions compared to the wavelength of the propagating wave. Reflection occurs from the surface of the earth and from buildings and walls.

Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities. Thanks to diffraction, the radio waves can still get propagated even when an LOS path between the transmitter and the receiver is obstructed by an obstacle. At high frequencies, diffraction, like reflection, depends on the geometry of the object as well as the amplitude, phase and polarization of the incident wave at the point of diffraction.

Scattering occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength, and where the density of such small obstacles is high. Scattered waves are produced by rough surfaces, small objects, or by other irregularities in the channel. In practice, foliage, street signs and lamp posts induce scattering in a mobile communication system.

1.1.2 Channel Variation due to the Doppler Effect

Many physical factors in the radio propagation path influence channel fading. These include:

- Multipath propagation - Due to the presence of reflections, diffraction and scattering in the channel, the transmitted signal often produces multiple versions, which undergo different paths before arriving at the receive antenna. These multiple versions are displaced with respect to one another in time and spatial orientation, and thereby induce fluctuations and/or distortion when added together at the receiver. Another important effect of multipath that constantly plagues communication engineers is that it often takes longer for the baseband signal to reach the receiver: the later version of the previous

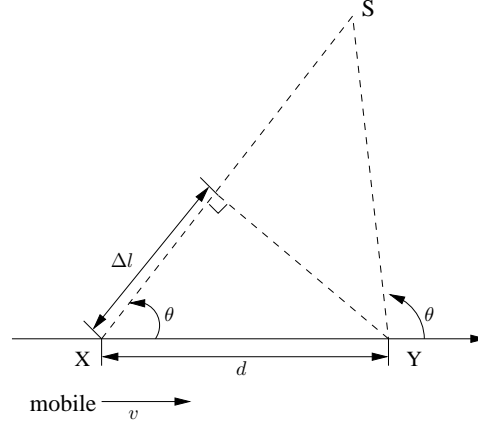


Figure 1.2: The Doppler shift.

symbol is added onto the early version of the next symbol, giving rise to as inter-symbol interference (ISI).

- Velocity of the mobile - The relative motion between the base station and the mobile results in a Doppler frequency shift, which can be positive or negative depending on whether the mobile receiver is moving toward or away from the transmitter. An example is given as follows

Example 1.1. Consider a mobile moving at a constant velocity v , along a path segment of length d between points X and Y . At the same time, it receives signals from the transmitter S , as illustrated in Fig. 1.2.

The difference in path length Δl , traveled by the wave from the base station S to the mobile at points X and Y is

$$\Delta l = d \cos \theta = v \Delta t \cos \theta,$$

where Δt is the time required for the mobile to travel from X to Y , and θ is assumed to be identical at points X and Y since the transmitter is assumed to be far away. The phase change in the received signal due to the difference in path length is

$$\Delta \phi = \frac{2\pi \Delta l f_c}{c} = \frac{2\pi v \Delta t f_c}{c} \cos \theta,$$

with f_c standing for the carrier frequency of the wave, and c the speed of light. The Doppler shift, f_d , defined as the rate of the phase change can be computed as

$$f_d = \frac{\Delta \phi}{2\pi \Delta t} = \frac{v \cos \theta}{c} f_c. \quad (1.1)$$

Due to the Doppler effect, the carrier frequency of the signal is shifted. Besides, if there are multiple Doppler shifts due to multipath, the spectrum of the signal will therefore be expanded. We will see that this presents a serious problem in OFDM.

- Velocity of surrounding objects - If objects in the channel are in motion, they induce a Doppler shift as well. It is worth mentioning that the moving surrounding objects bring a different behavior on the channel variation than a moving mobile. This is reflected in the diverse shapes of the Doppler spectrum. For instance, if only the mobile is in motion, the Doppler spectrum will be bathtub-shaped, while if there is a moving reflector in the environment, and the receiver and transmitter are still, the Doppler spectrum will become bell-shaped [106]. Note that for the same velocity, the maximum Doppler shifts due to the moving surrounding objects can be twice as high as due to the moving receiver (transmitter) alone.
- The bandwidth of transmitted signal - If the transmitted signal bandwidth is greater than the coherence bandwidth of the channel, the received signal will be distorted, but its power will not fluctuate too much. Here, the coherence bandwidth is defined as the range of frequencies over which the channel can be considered 'flat' [77]. Otherwise, if the transmitted signal has a narrow bandwidth compared to the coherence bandwidth of channel, the power of the received signal will change rapidly, but will not be distorted in time.

1.1.3 Impulse Response Model of the Channel and Its Approximation

The small-scale variations of the signal can be directly related to the impulse response of the channel. The channel impulse response is the output of the channel when a very brief signal, an impulse, is transmitted. Although this is impossible in a real system since an impulse has an infinitely short duration, it is a useful concept and contains all information necessary to simulate and analyze any radio transmission through the channel.

More specifically, we will model the channel as a time-varying (TV) finite impulse response (FIR) filter. The filtering nature is due to multipath: different signals arriving at the receiver at the same time are mixed together. The filter is assumed to be of finite order because the transmitted signal in practice does not have infinite duration, which means that the channel response ultimately settles to zero. The time-variation is due to the Doppler effect as we discussed in the previous section.

In the sequel, we will use delay spread to characterize the relative delay of the last multipath component arriving at the receiver with respect to the first component. As a dual, we will use Doppler spread to characterize the maximum Doppler shift with respect to the carrier frequency. A combination of different situations of the delay and Doppler spread leads to four types of channels:

- **AWGN channel** - an idealized situation where both the delay and Doppler spread are zero.
- **Time-selective channel** - a channel with zero delay spread and non-zero Doppler spread.
- **Frequency-selective channel** - a channel with non-zero delay spread but zero Doppler spread.
- **Doubly-selective channel** - a channel that is non-zero in both the delay and Doppler spread.

Although the TV FIR filter model provides a quite precise perception of a realistic channel, in practice, it can be still too cumbersome to utilize especially in the context of channel estimation. To this end, a parsimonious model that can capture the channel time-variation proves often useful. For instance, [116, 27] uses a first-order Markovian Model to approximate Rayleigh fading channels. For a block-wise transmission scheme like OFDM as dealt with in this thesis, a block model such as the basis expansion model (BEM) [110] can be more suitable. The BEM fits the channel taps with a superimposition of weighted basis expansion functions. The weighting parameters, referred to as the BEM coefficients in this these, remain constant within the duration of the block, and are usually much fewer in number than the channel taps. Some widely used BEMs will be introduced in the next chapter.

1.2 Orthogonal Frequency-Division Multiplexing

The mobile channel is characterized by multipath, giving rise to ISI at the receiver. To prevent a significant performance degradation, one solution is to deploy a properly designed filter at the receiver that counteracts ISI. On the other hand, from a practical point of view, it is desired that the receiver (in many situations the mobile handheld) be designed as simple as possible, and that the transmitter (in many situations the base station) should cover most of the hardware expenses. Orthogonal frequency-division multiplexing (OFDM) is devised for this purpose.

Due to its robustness against ISI and simplicity in hardware implementation, OFDM has found widely applications in numerous communication systems. Amongst

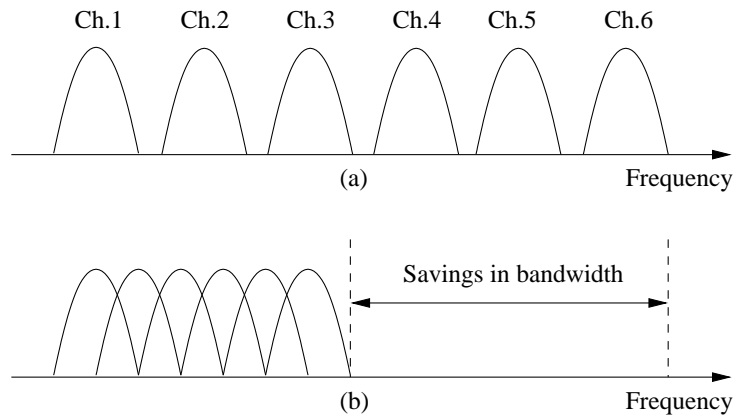


Figure 1.3: Comparison of FDM and OFDM: (a) conventional FDM, and (b) OFDM.

others, these include digital audio broadcasting (DAB), terrestrial digital video broadcasting (DVB-T), wireless LAN systems such as IEEE 802.11 and ETSI HiperLAN/2. In cable access networks, OFDM is also adopted for asymmetrical digital subscriber lines (ADSL) and very high bit rate digital subscriber lines (VDSL). In a future wireless access standard, IEEE 802.16, which is also referred to as worldwide interoperability microwave access (WIMAX), OFDM is adopted in the definition of the physical and medium access control (MAC) layers.

1.2.1 Overview of OFDM

OFDM is a special case of multicarrier transmission, where several signals are transmitted simultaneously at a lower rate over separate channels, or in other words subcarriers. These subcarriers are supposed not to interfere with each other, and thus we are able to recover the transmitted signal individually from each subcarrier. This leverages a great hardware complexity reduction at the receiver.

In addition, OFDM increases the robustness against frequency-selective fading. In a single-carrier system, where the transmitted signals share the same frequency band, just one fade can cause the entire link to fail, but in a multicarrier system, only a small percentage of the subcarriers will be affected. Although this could imply that the signals transmitted through these subcarriers are corrupted, it is still possible to recover them from the other transmitted signal by means of error-correction coding for instance.

The concept of using parallel data transmission can be dated back to the 1960s,

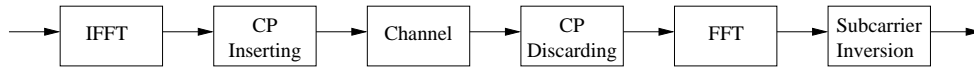


Figure 1.4: Block diagram of OFDM.

when frequency-division multiplexing (FDM) emerged [83]. In the classical FDM scheme, the total frequency band is divided into N non-overlapping frequency subchannels, as shown in the upper panel of Fig. 1.3, where each subchannel is modulated with a separate signal. Spectral redundancy is introduced in FDM to avoid interchannel interference, but leads to spectrum inefficiency. In contrast, the subcarriers in OFDM are allowed to overlap, as shown in the lower panel of Fig. 1.4, thereby making use of the spectrum much more efficiently. In spite of overlapping, the subcarriers in OFDM can still be free from mutual interference. This is achieved by designing the position and shape of the subcarriers properly such that they are mathematically orthogonal to each other.

Fig. 1.4 illustrates the process of an OFDM system. In the figure, (I)FFT stands for (inverse) fast Fourier transform, and CP for cyclic prefix. Inserting a CP is an indispensable step in combating the ISI. It boils down to appending the last L_z symbols of the transmitted stream to the beginning. At the receiver, the CP is discarded by cutting off the first L_z symbols of the received stream. To eliminate ISI completely, it is required that L_z is not shorter than the FIR length of the channel. Besides CP, it is also possible to introduce other forms of transmitter redundancy, e.g., zero padding (ZP) [35, 86, 87, 117] or known symbol padding (KSP) [25, 79].

One of the major disadvantages of OFDM is the high peak-to-average power ratio (PAR) at the transmitter due to the fact that an OFDM symbol is a summation of a large number of modulated subcarriers. Even with a low level modulation such as quaternary phase shift keying (QPSK) on each subcarrier, the PAR can still be far beyond the dynamic range of a practical transmitter power amplifier. A single-carrier system with a frequency-domain equalizer can avoid this problem, and at the same time eliminate ISI without inducing a large complexity increase. We will call such a scheme simply a single-carrier system in the thesis, but bear in mind that its channel equalization is implemented in the frequency domain. The block diagram of such a single-carrier system is depicted in Fig. 1.5, where, compared to Fig. 1.4, one can easily observe that the IFFT is shifted from the transmitter to the receiver. Although this will inevitably increase complexity at the receiver, it is still attractive since the IFFT can nowadays be very efficiently implemented. Other advantages and differences of the single-carrier system with respect to OFDM can be found in [118, 72].

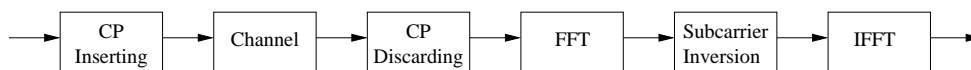


Figure 1.5: Block diagram of the single-carrier system.

1.2.2 Challenges due to the Fast Fading Channel in OFDM

In this thesis, we will primarily consider the following problems due to the fast fading channel, which undermines the reliability of the OFDM transmission.

Intercarrier Interference

In a time-invariant (TI) channel, the signals modulated on different subcarriers will not interfere with each other due to the orthogonality among the subcarriers. In a time-varying (TV) channel, the Doppler spread expands the signal frequency band, and ruins this orthogonality, resulting in inter-carrier interference (ICI).

The existence of ICI prevents the use of a simple equalizer at the receiver. Since each subcarrier is contaminated by the signals from neighboring subcarriers, it is not possible to recover the transmitted signals by simply inverting the channel on each subcarrier. A similar problem arises in the single-carrier system if we want to equalize the channel in the frequency domain. The cost required for a traditional equalizer to annihilate the ICI completely will be at least cubic in the total number of subcarriers, which is too expensive if they are numerous.

Limited Coherence Time

Another important constraint as a consequence of the Doppler spread is the much reduced channel coherence time. In a low-mobility situation, it is often reasonable to assume a constant channel during the transmission of at least one block, which means that the ICI is almost negligible, but this assumption becomes invalid for channels that vary faster.

Besides the increased complexity for equalization, ICI inflicts more difficulties in channel estimation. First, more channel parameters need to be estimated. Because the TV channel has a different realization at every time instance, it is not uncommon that the number of channel unknowns can supersede the number of observation samples. Although this problem can be alleviated by applying some parsimonious model on the channel as we will introduce in the next chapter, TV channel estimation remains far more problematic compared to TI channel estimation. Besides, if we use pilots in OFDM systems for channel estimation, due to ICI, it is practically

impossible to find those observation samples that are solely dependent on pilots and thus not contaminated by unknown data symbols.

Another problem related to the limited coherence time is that it is hard to gather a large amount of observation samples that are relevant to channel estimation. Such a large amount of relevant observation samples are extremely useful to increase the estimation precision. For instance, they are used in subspace-based blind channel estimation to reconstruct a reliable subspace where the channel resides [68, 1, 125]. In TV channels, such methods are more difficult to apply [55].

Inter-symbol Interference

OFDM is devised to combat ISI. In the context of this thesis where the channel equalization is implemented in the frequency domain for both the OFDM and single-carrier systems, ISI can still emerge in disguise of ICI and inter-block interference (IBI). The former is caused by the Doppler spread or frequency offset [66], while the latter often arises when the channel is fairly long, and it is not affordable to accommodate a sufficiently long CP due to bandwidth constraints.

Apart from the above research challenges, there are also other issues that are interesting but are not treated in this thesis.

Diversity due to Time-Selectivity

As a knife has two blades, the Doppler spread can also be exploited to improve the performance of a communication system. With the frequency band expanded, the information of the transmitted signal is spread out, and is therefore less susceptible to failure if the fading only hits part of the frequencies. This phenomenon is well known as diversity. Compared to the diversity offered by space [4], frequency-selectivity [118], or both (see [52] and the references therein), the Doppler diversity is less commonly put in the spotlight except in [85, 60, 61].

MIMO Setup

Multi-input multi-output (MIMO) systems have received enormous attention recently thanks to the feature that they are able to provide a more reliable or higher capacity link under the same total transmit power constraint.

The issue of MIMO can be related to diversity as how to optimally distribute information among multiple transmit antennas. In spite of its robustness, the MIMO

system requires often a high-complexity channel equalizer and estimator. Also, synchronization of the received signals from different transmit antennas is a huge challenge for system designers.

1.3 Thesis Survey and Contributions

The remainder of this thesis is organized as follows.

In Chapter 2, we present the background knowledge that is going to be used throughout the thesis. We first introduce wireless setups and show how the channel models and input-output (I/O) relationships are derived. Subsequently, we introduce the idea of a BEM to approximate the TV channel, and its different realizations. In the last part of Chapter 2, we present different OFDM and single-carrier schemes that will be investigated as well as their data models.

In Chapter 3, we show how to construct a low-complexity channel equalizer for OFDM. We will exploit the fact that under a realistic Doppler spread, most of the ICI stems from the adjacent subcarriers. This implies that the channel matrix in the frequency domain is approximately banded. Utilizing this banded structure, we can lower the complexity of the channel equalizer by employing some numerical techniques. The out-of-band entries that are neglected by the equalizer give rise to a high noise floor in performance. In order to maximally suppress their impact, a receiver window is deployed.

The publication that is associated to this chapter are the following:

- Z. Tang and G. Leus, "A Novel Receiver Architecture for Single-Carrier Transmission over Time-Varying Channels", manuscript submitted to IEEE Journal on Selected Areas in Communications (JSAC), 2007.

Constructing the channel equalizer requires channel state information (CSI), which is attained in this thesis by means of pilot-aided channel estimators. This subject is treated in Chapter 4. Although the pilots can be interleaved with data in the frequency domain, it is impossible to find observation samples that are solely dependent on the pilots and free from the interference induced by the unknown data symbols. It is thus of great significance to select observation samples wisely. It turns out that different types of channel estimators call for different approaches to optimally select the observation samples.

The publications that are associated to this chapter are the following:

- Z. Tang, R. C. Cannizzaro, G. Leus and P. Banelli, "Pilot-Assisted Time-Varying Channel Estimation for OFDM Systems", in *IEEE Transactions on Signal Processing*, Page(s):2226 - 2238, Volume 55, Issue 5, May 2007.

- Z. Tang, R. C. Cannizzaro, G. Leus and P. Banelli, "Pilot-Assisted Time-Varying OFDM Channel Estimation", in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, Page(s):IV-133 - IV-136, Volume 4, May 2006.

The channel estimation in Chapter 4 is based on a single OFDM symbol. Despite the short coherence time of the TV channel, it can be still beneficial to apply channel estimation based on multiple OFDM symbols. An issue then arises as to what the best strategy is to distribute the pilots along the frequency axis as well as the time axis. Diverse pilot placement schemes are studied and compared in Chapter 5. Besides, Chapter 4 does not consider channel identifiability. This issue is also discussed in Chapter 5, which is generalized to a multiple OFDM symbol case.

The publications that are associated to this chapter are the following:

- Z. Tang, G. Leus and P. Banelli, "Time-Varying Channel Estimation - A Block Approach", Chapter in *Wireless Communications over Rapidly Time-Varying Channels* (F. Hlawatsch and G. Matz ed.) approved by Academic Press.
- Z. Tang and G. Leus, "Pilot Schemes for Time-Varying Channel Estimation in OFDM Systems", in *Proceedings of the IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2007)*, June 2007.
- Z. Tang, G. Leus and P. Banelli, "Pilot-Assisted Time-Varying OFDM Channel Estimation Based on Multiple OFDM Symbols", in *Proceedings of the IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2006)*, June 2006.

Chapter 6 discusses how to build a frequency-domain (FD) equalizer for a single-carrier system, which is plagued by both ICI and IBI. In principle, the band assumption with aid of windowing can also be employed here to lower the complexity, as discussed for OFDM in Chapter 3. In addition, we propose in this chapter to introduce some receiver redundancy, which is unique to the single-carrier system and allows the FD equalizer to benefit from a better resolution of the Doppler spread. As a result, the out-of-band interference can be further considerably reduced. Because the redundancy is introduced at the receiver, the data rate is not compromised.

The publications related to this chapter are the following:

- Z. Tang and G. Leus, "A Novel Receiver Architecture for Single-Carrier Transmission over Time-Varying Channels", manuscript submitted to *IEEE Journal on Selected Areas in Communications (JSAC)*, 2007.
- Z. Tang and G. Leus, "Receiver Design for Single-Carrier Transmission over Time-Varying Channels", in *Proceedings of the IEEE International Conference on*

Acoustics, Speech and Signal Processing (ICASSP 2007), Page(s):III-129 - III-132, Volume 3, April 2007.

In Chapter 7, we discuss how to attain the CSI in a single-carrier system. Pilots are interleaved with data symbols in the time domain. In the traditional approaches, because the time-domain (TD) channel matrix is strictly banded due to the FIR assumption, only those observation samples that are exclusively dependent on pilots are selected for channel estimation. However, it is also possible to deliberately introduce some interference by including more observation samples. In this way, the noise can be better averaged out, while the interference can be suppressed by employing appropriate channel estimators. Note that this is especially significant to TV channel estimation, for which usually a much larger number of channel unknowns need to be estimated. Further, channel identifiability for a single-carrier system is discussed in this chapter.

The publication that is associated to this chapter are the following:

- Z. Tang, G. Leus and P. Banelli, "Time-Varying Channel Estimation - A Block Approach", Chapter in *Wireless Communications over Rapidly Time-Varying Channels* (F. Hlawatsch and G. Matz ed.) approved by Academic Press.

Like in OFDM, the pilot structure also plays an important role to the channel estimation performance in a single-carrier system. In Chapter 8, we try to find an optimal pilot structure for a simpler situation, the time-selective channel. We prove that the mean squared error (MSE) of the channel estimator is a convex function of the pilot powers but not in the pilot positions. The problem can be resolved by optimizing the pilot powers and positions iteratively. To expedite the searching time, we propose to impose a symmetric constraint on the pilot structure, which is inspired by the BEM structure that is used to approximate the TV channel.

The publication that is associated to this chapter is the following:

- Z. Tang and G. Leus, "Time-Multiplexed Training for Time-Selective Channels", accepted for publication in *IEEE Signal Processing Letters*, 2007.

Besides the above topics that are presented in this thesis, other contributions have been made in the following publications:

- Z. Tang and G. Leus, "Low-Complexity Equalization of Time-Varying Channels with Precoding", in *IEEE Transactions on Signal Processing*, Page(s):3642 - 3648, Volume 54, Issue 9, September 2006.
- Z. Tang and G. Leus, "A Receiver Architecture for Maximum Diversity Transmissions over Doubly-Selective Channels", in *Proceedings of IEEE Workshop on*

Signal Processing Advances in Wireless Communications (SPAWC 2005), Page(s): 171 - 175, June 2005.

- Z. Tang and G. Leus, "RLS Direct Equalizer Estimation with Assistance of Pilots for Transmissions over Time-Varying Channels", in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, September 2005.

Chapter 2

Preliminaries

Prior to presenting the main contributions of the research, we first introduce in this chapter some background concepts and ideas, which will be used throughout the remainder of the thesis.

In Section 2.1, we first briefly describe how wireless transceivers are built up, and then introduce the mathematical model of the channel, which takes the filtering characteristics of the transmitter and receiver into account. Special attention is paid to the time-varying feature of the channel.

In Section 2.2, we introduce the idea of using parsimonious channel models, namely the basis expansion model (BEM), to approximate TV channels. The modeling performance of various BEMs is compared.

Block transmission techniques and their corresponding data models are presented in Section 2.3, which consists of two parts: the OFDM and single-carrier systems. It is also shown that depending on the specific forms of redundancy at the transmitter, we can obtain different variants of OFDM and single-carrier systems.

2.1 A Digital Wireless Communication System Architecture

In this section, we will take a brief tour going through the major elements of a wireless communication system. This implies that not each step of the transmission/reception will be cast to spotlight. For instance, we will begin our tour directly from the point where the data stream is fed to the transmit filter. Before reaching that point, the original information stream, which is usually in binary form, may have already undergone several steps such as channel encoding, interleaving, and mapping [9]. These steps, despite their significance to the overall system performance, will not be reviewed here in light of their irrelevance to the results that will be presented in this thesis. For the same consideration, we will end our tour at the decimator, thus skipping the subsequent procedures such as channel decoding, de-interleaving and demapping.

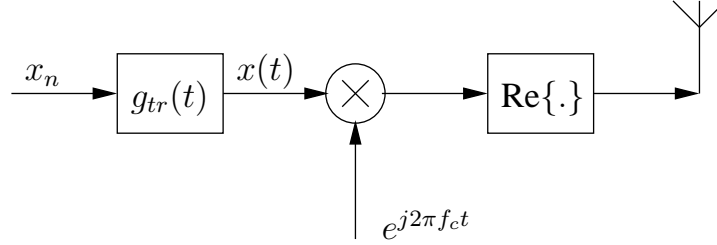


Figure 2.1: A typical transmitter structure.

2.1.1 Transmitter Structure

A typical transmitter structure is illustrated in Fig. 2.1. Let us use x_n to denote the sequence of data symbols that have been mapped to a certain constellation. This (discrete-time) data sequence is fed through a pulse-shaping transmitter filter with impulse response $g_{tr}(t)$, resulting in the (continuous-time) complex baseband signal $x(t)$:

$$x(t) = \sum_{n=-\infty}^{\infty} g_{tr}(t - nT)x_n, \quad (2.1)$$

where $1/T$ stands for the data rate. The bandwidth occupied by $x(t)$ depends on the specific choice of the pulse $g_{tr}(t)$, which is required to have a minimum bandwidth $W = 1/T$ in order to allow for an error-free reconstruction of x_n from the signal $x(t)$ according to the Nyquist criterion. In many practical systems, the raised-cosine pulse is adopted [76]:

$$g_{tr}(t) = \frac{\sin(\pi t/T)}{\pi t/T} \frac{\cos(\alpha\pi t/T)}{1 - (2\alpha t/T)^2}, \quad (2.2)$$

where the parameter α is called the roll-off factor that takes a value between 0 and 1. The bandwidth of $g_{tr}(t)$ is determined by α as $W = (1 + \alpha)/T$. Although a pulse with a small bandwidth is desired from the point of view of bandwidth efficiency, other problems can arise in practice. For instance, the sidelobes of the pulse in the time domain will be higher and attenuate more slowly for a smaller α as exhibited in Fig. 2.2. Correspondingly, the signal power will be further dissipated to neighboring time intervals, which makes the system more susceptible to ISI degradation for just a small timing error.

The complex baseband signal is next modulated on the carrier frequency f_c . In mathematics, it is multiplied by a complex exponential $e^{j2\pi f_c t}$. This multiplication

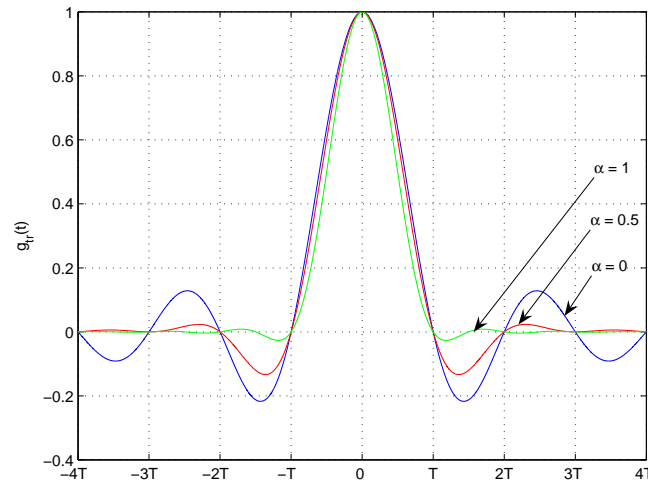


Figure 2.2: Impulse response of the raised-cosine pulse.

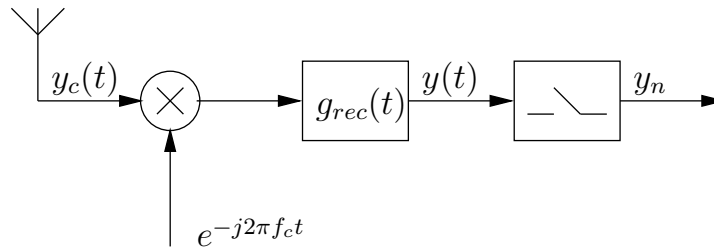


Figure 2.3: A typical receiver structure.

shifts the spectrum of the complex baseband signal from DC to f_c . The real part of the resulting signal is transmitted by the antenna over the air interface.

2.1.2 Receiver Structure

It is the task of the receiver to recover the baseband signal $x(t)$ from the received signal $y(t)$. In general, $y(t)$ contains not only the desired information transmitted at the carrier frequency f_c , but also information from other applications and users

operating at other carrier frequencies. To minimize interference, one solution is presented in Fig. 2.3, where the spectrum of the incoming signal is first shifted to the left by multiplying the received signal with a complex exponential $e^{-j2\pi f_c t}$. By this means, the spectrum of the desired baseband signal will be located in a bandwidth W centered around DC. In order to remove the interference beyond the desired bandwidth, we feed the signal next to a low-pass filter. The resulting signal stream $y(t)$ is related with $y_c(t)$ as

$$y(t) = \int_{-\infty}^{\infty} g_{\text{rec}}(t - \tau) y_c(\tau) e^{-j2\pi f_c \tau} d\tau, \quad (2.3)$$

where $g_{\text{rec}}(t)$ stands for the low-pass filter at the receiver.

2.1.3 Channel and Discrete Baseband Model

Due to the multipath mechanism as described in Chapter 1, radio waves impinge upon the receiving antenna in different angles, and experience different phase shifts and attenuation. Let us group the radio waves that arrive with the same delay in one cluster, with each component of the same cluster characterized by its own complex gain and phase shift [45]. As a result, the physical channel can be modeled as an aggregate of several such clusters:

$$g_{\text{ch}}(t, \tau) = \sum_c \delta(\tau - \tau_c) \sum_w G_{c,w} e^{j2\pi f_{c,w} t}, \quad (2.4)$$

where τ_c stands for the delay of the c th cluster, and $G_{c,w}$ and $f_{c,w}$ represent the complex gain and frequency offset of the w th ray of the c th cluster, respectively.

With this physical channel model, we can associate the transmitted discrete symbol stream x_n in (2.1) with the received signal $y(t)$ in (2.3) as

$$y(t) = \sum_{l=-\infty}^{\infty} h(t, t - lT) x_l + v(t), \quad (2.5)$$

where $v(t)$ represents the additive noise; $h(t, \tau)$ stands for the composite channel that takes the transmitter filter, the physical channel, and the receiver filter into account:

$$h(t, \tau) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{\text{rec}}(s) g_{\text{tr}}(\tau - \theta - s) g_{\text{ch}}(t - s, \theta) ds d\theta. \quad (2.6)$$

By assuming that the physical channel remains static during the span of the receive

filter $g_{\text{rec}}(t)$, we can replace $g_{\text{ch}}(t - s; \theta)$ by $g_{\text{ch}}(t; \theta)$ in the above, which leads to

$$\begin{aligned} h(t, \tau) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g_{\text{rec}}(s) g_{\text{tr}}(\tau - \theta - s) ds \right) g_{\text{ch}}(t, \theta) d\theta, \\ &= \int_{-\infty}^{\infty} \psi(\tau - \theta) g_{\text{ch}}(t, \theta) d\theta, \\ &= \sum_c \psi(\tau - \tau_c) \sum_w G_{c,w} e^{j2\pi f_{c,w} t}, \end{aligned} \quad (2.7)$$

where $\psi(t)$ represents the convolution of $g_{\text{tr}}(t)$ with $g_{\text{rec}}(t)$.

In the sequel, we will simply use “channel” to refer to the joint effect of the transmitter filter, the physical propagation channel and the receiver filter. Suppose that the bandwidth of the channel is smaller than $1/T$, then sampling at the symbol rate T is sufficient to avoid aliasing in terms of the Nyquist criterion (otherwise, we need to increase the sampling rate). In that case, the discrete-time received signal y_n can be defined as

$$y_n = y(nT), \quad (2.8)$$

and the discrete channel input/output (I/O) relationship can be expressed as

$$y_n = \sum_{l=-\infty}^{\infty} h_{n,n-l} x_l + v_n, \quad (2.9)$$

where v_n is similarly defined as y_n . The baseband discrete-time channel is represented by $h_{n,l}$, where the subscript n stands for the time index, and the subscript l corresponds to the delay of the filter:

$$h_{n,l} = \sum_c \psi(lT - \tau_c) \sum_w G_{c,w} e^{j2\pi f_{c,w} nT}. \quad (2.10)$$

In practice, it is more convenient to use $h_{n,l}$ directly as an abstraction of the channel rather than use its cumbersome structure on the right-hand side (RHS) of (2.10). This is especially true seeing that in a realistic communication system, most of the channel power is concentrated within a limited time interval, which implies that the channel has a limited support, say $L + 1$. In addition, by taking the causality of the transmission process into account, we can further simplify the channel to an FIR filter with $h_{n,l} = 0$ if $l < 0$ or $l > L$. As a result, (2.9) can be written as

$$y_n = \sum_{l=0}^L h_{n,l} x_{n-l} + v_n. \quad (2.11)$$

The notation $h_{n,l}$ is typically used to denote doubly-selective channels, which can be viewed as a generalized description of various channel situations. For instance,

time-selective channels correspond to the special case where $L = 0$, resulting in

$$h_{n,l} = h_n, \quad (2.12)$$

while the notation of frequency-selective channels degrades to

$$h_{n,l} = \sum_{m=0}^L \delta_{l-m} h_l, \quad (2.13)$$

where δ_n denotes the Kronecker delta, which equals one if $n = 0$, or zero otherwise.

2.2 Channel Time-Variation and BEM Approximation

It is understood that TV channels occur when the communicating parties and/or the scatterers in the environment are moving. This is reflected in (2.10), where $h_{n,l}$ varies with the time index n due to the non-zero Doppler frequency shifts $f_{c,w}$. Recall from (1.1) that $f_{c,w}$ is linear function of the velocity v , and the carrier frequency f_c . To characterize how fast the channel varies, we use in the sequel the normalized Doppler spread ν_D , which is defined as

$$\nu_D = f_{\max} T \quad (2.14)$$

with f_{\max} denoting the maximum absolute Doppler frequency shift

$$f_{\max} := \max_{c,w} |f_{c,w}|. \quad (2.15)$$

Statistically, TV channels can be described by their Doppler spectrum defined as

$$\mathcal{E}_h \{h_{p,l} h_{p-n,l}^*\} = \sigma_l^2 \gamma_n, \quad (2.16)$$

where

$$\sigma_l^2 := \mathcal{E}_h \{h_{p,l} h_{p,l}^*\}, \quad (2.17)$$

and γ_n characterizes the shape of the Doppler spectrum. For instance, if the Doppler spectrum is bathtub-shaped,

$$\gamma_n = J_0 \left(2\pi \frac{f_c v}{c} T |n| \right), \quad (2.18)$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind. Such a bathtub-shaped Doppler spectrum is better known as Jakes' spectrum. If the Doppler spectrum is bell-shaped [106], then

$$\gamma_n = \int_0^\infty J_0^2 \left(2\pi \frac{f_c v}{c} T |n| \right) p(v) dv, \quad (2.19)$$

where $p(v)$ is the distribution function of the scatterer velocity v . In practice, one might observe a mixture of the above two spectrum. The statistics assumed in these channel models are uniquely determined by physical propagation parameters such as the path delays, frequencies and complex attenuation etc [45]. Despite their accuracy, these statistical models are generally bulky and difficult to handle. Therefore, many existing works resort to a parsimonious channel model such as the basis expansion model (BEM).

The BEM is especially useful to reduce the number of channel parameters in the context of block transmission as we will address in the next section. For a block transmission that spans N symbol times, the whole channel will be uniquely characterized by $N(L + 1)$ parameters with L denoting the channel order because each channel tap has a different realization at each time instance. Suppose \mathbf{h}_l stands for an $N \times 1$ vector that collects the time-variation of the l th channel tap from the 1st until the N th time instance:

$$\mathbf{h}_l := [h_{0,l}, \dots, h_{N-1,l}]^T, \quad (2.20)$$

then each \mathbf{h}_l , for $l = 0, \dots, L$, can be expressed in terms of a BEM as

$$\mathbf{h}_l = \underbrace{[\mathbf{b}_0 \ \dots \ \mathbf{b}_Q]}_{\mathbf{B}} \mathbf{Q} \underbrace{\begin{bmatrix} c_{0,l} \\ \vdots \\ c_{Q,l} \end{bmatrix}}_{\mathbf{c}_l} + \boldsymbol{\delta}_l, \quad (2.21)$$

where the $N \times 1$ vector \mathbf{b}_q is termed as the q th expansion basis; $c_{q,l}$ is the corresponding BEM coefficient. $\boldsymbol{\delta}_l$ stands for the BEM modeling error. Note that the definition above is a bit different than in the literatures due to the presence of a $(Q+1) \times (Q+1)$ matrix \mathbf{Q} . It is added here to make the columns of \mathbf{B} orthonormal to each other*:

$$\mathbf{B}^H \mathbf{B} = \mathbf{I}_{Q+1}. \quad (2.22)$$

Despite \mathbf{Q} , we will follow the tradition in the sequel and call \mathbf{b}_q as the q th basis expansion function. By introducing the BEM in (2.21), we are able to decompose the TV channel into an ideal ‘‘BEM channel’’ and its corresponding error. The BEM channel is comprised of the BEM matrix \mathbf{B} , which can in principle be designed to be independent of the channel, and the BEM coefficients $c_{q,l}$, whose values are determined by the specific channel realizations. Apparently, if $N > Q + 1$, \mathbf{B} will be a ‘‘tall’’ matrix, and the BEM modeling error $\boldsymbol{\delta}_l$ will seldom be zero. However, by

*This step is optional, and basically will not influence the BEM modeling error.

optimally designing the BEM matrix, we can yet reduce the BEM modeling error to a negligible level such that

$$\mathbf{h}_l \approx \mathbf{B}\mathbf{c}_l. \quad (2.23)$$

Note that apart from the BEM, it is also possible to use a Gauss-Markov process to simulate the channel dynamics [27]. Such a model is interesting for sequential time-domain processing. When we deal with block transmission/precoding schemes as in this thesis, it is often more convenient to use a block-based channel model like a BEM.

Discrete Karhunen-Loève BEM (DKL-BEM)

The DKL-BEM is optimal in terms of the mean squared error (MSE) [114, 120, 41, 105] of the BEM. To generate the DKL-BEM, we first construct a kernel matrix \mathbf{C} , which is the normalized covariance matrix of any channel tap:

$$\mathbf{C} = \mathcal{E}_h\{\mathbf{h}_l\mathbf{h}_l^H\}/\sigma_l^2. \quad (2.24)$$

Afterwards, we select \mathbf{b}_q as the q th most significant eigenvector of \mathbf{C} ,

$$\mathbf{C}\mathbf{b}_q = \lambda_q\mathbf{b}_q, \quad (2.25)$$

where λ_q stands for the q th eigenvalue of \mathbf{C} , which are sorted in a descendent order. The DKL-BEM is in essence a reduced-rank decomposition of the Doppler spectrum, which could be the bathtub-shaped or bell-shaped spectrum.

Discrete Prolate Spheroidal BEM (DPS-BEM)

A weakness of the DKL-BEM is that it relies on the knowledge of the channel statistics. If the assumed knowledge deviates from the true scenario, which is highly likely in the case of TV channels, the DKL-BEM will perform sub-optimally. As a compromise, one can derive a BEM that is based on a general approximation for all kinds of channel statistics. For instance, the DPS-BEM corresponds to the DKL-BEM with a rectangular spectrum assumption [122]. To be more specific, the kernel matrix \mathbf{C} for the DPS-BEM is defined as

$$[\mathbf{C}]_{m,n} = \mathcal{E}_h\{h_{m,l}h_{n,l}^*\}/\sigma_l^2 \approx \frac{\sin(2\pi(m-n)\nu_D)}{\pi(m-n)}, \quad (2.26)$$

from which \mathbf{b}_q is retained as the q th most significant eigenvector of \mathbf{C} . Note that by adopting such a kernel matrix, the resulting \mathbf{b}_q is actually a finite-length approximation of an infinite-length Slepian sequence [95] denoted as $\tilde{\mathbf{b}}_q := [\tilde{b}_{-\infty}, \dots, \tilde{b}_{\infty}]$. $\tilde{\mathbf{b}}_q$ enjoys the merit that most of its power is concentrated within the interval $0, \dots, N-1$, and it is strictly band-limited to the maximum Doppler bandwidth ν_D .

Complex Exponential BEM (CE-BEM)

From (2.10), it is clear that the TV channel can also be viewed as a summation of $Q + 1$ paths, each path consisting of a TI component and a TV component in the form of a complex exponential [110, 19]:

$$h_{n,l} = \sum_{q=0}^Q e^{j\omega_q n} c_{q,l}, \quad (2.27)$$

and hence the q th expansion basis admits the expression as

$$\mathbf{b}_q = [1, \dots, e^{j\omega_q n}, \dots, e^{j\omega_q(N-1)}]^T. \quad (2.28)$$

In the above, ω_q stands for the Doppler frequency associated with the q th path. We observe that this CE-BEM falls not in our flavor of BEM designs in that the basis function itself depends on the channel. Although the Doppler frequency is further approximated in [36] as

$$\omega_q \approx 2\pi \frac{\nu_D}{T} \cos(2\pi \frac{q+1}{Q+1}), \quad (2.29)$$

it is still not convenient to use in practice. Besides, (2.27) is based on the assumption that there are a large number of scatterers, which implies that to enable a precise model, the number of paths Q in the CE-BEM must be very high (more than ten). Some more practical variants of the CE-BEM will be discussed as below.

Critically-sampled CE-BEM [(C)CE-BEM]

The (C)CE-BEM uses complex exponentials, whose period equals the block size N , to construct the BEM matrix [39, 60, 59, 55, 47]. By defining its q th expansion basis as

$$\mathbf{b}_q = [1, \dots, e^{j\frac{2\pi}{N}n(q-\frac{Q}{2})}, \dots, e^{j\frac{2\pi}{N}(N-1)(q-\frac{Q}{2})}]^T, \quad (2.30)$$

we understand that the (C)CE-BEM is in essence the truncated Fourier expansion, which picks up the first $Q + 1$ frequency elements in the Doppler spectrum that are located symmetrically around the zero frequency (DC).

Note that it is also possible to interpret the basis expansion functions of the (C)CE-BEM as the most significant eigenvectors of a covariance matrix \mathbf{C} , just like the case of the DKL-BEM and DPS-BEM. In this case, \mathbf{C} must obviously be some circulant matrix, which is not realistic of course. This intuitively explains why the modeling performance of the (C)CE-BEM is inferior to that of the DKL-BEM and the DPS-BEM.

Despite the fact that it is not always a tight model, the (C)CE-BEM is one of the BEMs that were initially studied for modeling simplifications. It gained most of its popularity due to the following property.

Property 2.1. Let \mathbf{E}_q denote a permutation matrix, which is obtained from an identity matrix by shifting its columns circularly q -times to the right; \mathbf{W}_N stands for the N -point DFT matrix $[\mathbf{W}_N]_{m,n} = 1/\sqrt{N}e^{j\frac{2\pi}{N}mn}$, It can be shown that the q th expansion basis vector \mathbf{b}_q of the (C)CE-BEM satisfies

$$\mathbf{W}_N \mathcal{D}\{\mathbf{b}_q\} \mathbf{W}_N^H = \mathbf{E}_{q-\frac{Q}{2}}, \quad (2.31)$$

where $\mathcal{D}\{\mathbf{b}_q\}$ denotes a diagonal matrix with \mathbf{b}_q as its diagonal.

Proof. With the definition in (2.30), \mathbf{b}_q corresponds to one of the rows (columns) of \mathbf{W}_N . Due to its circularity, each row of \mathbf{W}_N , when right multiplied by $\mathcal{D}\{\mathbf{b}_q\}$, becomes just another row, and hence $\mathbf{W}_N \mathcal{D}\{\mathbf{b}_q\}$ is related with \mathbf{W}_N with all the rows permuted,

$$\mathbf{W}_N \mathcal{D}\{\mathbf{b}_q\} = \mathbf{E}_{q-\frac{Q}{2}} \mathbf{W}_N. \quad (2.32)$$

Substituting the above in the left-hand side (LHS) of (2.31), and understanding that \mathbf{W}_N is a unitary matrix, we complete the proof. \square

Oversampled CE-BEM [(O)CE-BEM]

A major weakness of the (C)CE-BEM is its poor modeling performance. As an improvement, the CE-BEM that is used in [107, 53, 21] is obtained by selecting the complex exponentials with a period that is larger than the block size, i.e.,

$$\mathbf{b}_q = [1, \dots, e^{j\frac{2\pi}{K}n(q-\frac{Q}{2})}, \dots, e^{j\frac{2\pi}{K}(N-1)(q-\frac{Q}{2})}]^T, \quad (2.33)$$

with $K > N$. Such a BEM is called as an oversampled CE-BEM in the thesis.

Compared to the (C)CE-BEM, the (O)CE-BEM uses $2\pi(q - \frac{Q}{2})/K$ as a better approximation of ω_q in (2.28), which renders a much tighter fit. However, Property 2.1 loses its validity.

Polynomial BEM (P-BEM)

In the world of numerical analysis, the most conventional approximation method consists of using polynomials [22]. The underlying consideration is that if the channel time-variation is smooth enough (in continuous form), it can be represented by the leading derivative terms using e.g., a Taylor series expansion. This idea is reflected in the BEM design of [10] and its application can be found in [11, 108]. In discrete form, the P-BEM matrix is constructed as

$$\mathbf{b}_q = [(-\frac{N-1}{2})^q, \dots, (\frac{N-1}{2})^q]^T. \quad (2.34)$$

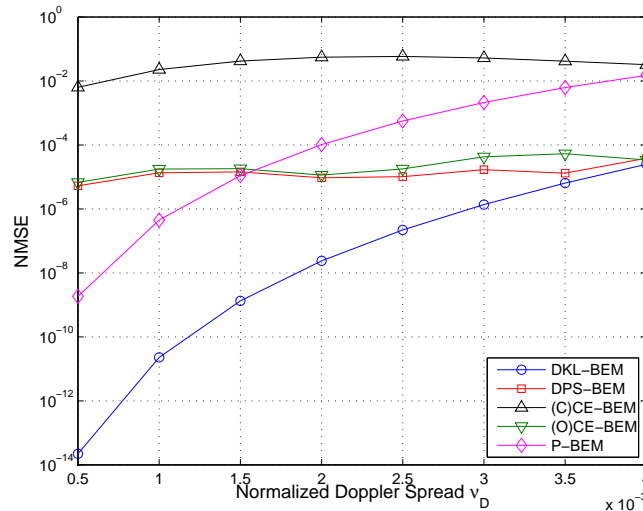


Figure 2.4: BEM Modeling Performance

Like the (O)CE-BEM, which can also be viewed as a special polynomial in trigonometric form, the P-BEM serves purely as an approximation/interpolation method, and does not really admit a physical interpretation related to the Doppler frequency.

Other BEM Options

Other BEM designs than the above are also reported in various works. For instance, [121] uses a spline approach. Besides, it is also possible to combine the above BEMs for different purposes [96, 70, 38].

BEM Modeling Performance Comparison

To conclude this section, we compare in Fig. 2.4 the modeling performance of the DKL-, DPS-, (C)CE-, (O)CE- and P-BEM. For the simulation, we generated TV channels following Jakes' model [124]. The other simulation parameters are $N = 256$, $Q = 4$, and $K = 512$ for the (O)CE-BEM, which is twice the block size. As a performance criterion, we use the normalized mean squared error (NMSE), which is defined as

$$\text{NMSE} = \frac{\mathcal{E}_h\{\|\mathbf{h}_l - \mathbf{B}\mathbf{c}_l\|^2\}}{N\sigma_l^2}. \quad (2.35)$$

It is easy to observe from Fig. 2.4 that the DKL-BEM yields the best modeling performance while the (C)CE-BEM yields the worst. The P-BEM is tight for slowly varying channels but loses track if the channel varies faster. As a good compromise, the DPS-BEM and the (O)CE-BEM render a relatively robust modeling performance.

2.3 Block Transmission Techniques

2.3.1 CP-OFDM System Model

In OFDM transmissions, the data symbol sequence[†] s_n is grouped in blocks of size N , producing

$$\mathbf{s}(i) := [s_{iN}, \dots, s_{(i+1)N-1}]^T \quad (2.36)$$

for the i th block. Prior to transmission, $\mathbf{s}(i)$ is first applied to an IFFT, and afterwards a cyclic-prefix (CP) of length L_z is added. This results in an $(N + L_z) \times 1$ vector

$$\mathbf{x}(i) := [x_{i(N+L_z)}, \dots, x_{(i+1)(N+L_z)-1}]^T, \quad (2.37)$$

which is related to $\mathbf{s}(i)$ as

$$\mathbf{x}(i) := \mathbf{T}_{\text{cp}} \mathbf{W}_N^H \mathbf{s}(i), \quad (2.38)$$

where \mathbf{T}_{cp} denotes the $(N + L_z) \times N$ matrix

$$\mathbf{T}_{\text{cp}} := \begin{bmatrix} \mathbf{0}_{L_z \times (N-L_z)} & \mathbf{I}_{L_z} \\ & \mathbf{I}_N \end{bmatrix}. \quad (2.39)$$

\mathbf{T}_{cp} accounts for the operation of adding the CP, which appends the last L_z symbols of $\mathbf{W}_N^H \mathbf{s}(i)$ to the beginning of it.

At the receiver, the received samples y_n are grouped into blocks of size $N + L_z$, with the i th block

$$\mathbf{y}_t(i) := [y_{i(N+L_z)}, \dots, y_{(i+1)(N+L_z)-1}]^T. \quad (2.40)$$

Conform the FIR assumption of the channel, we can rewrite the channel I/O relationship in (2.11) in a block form as

$$\mathbf{y}_t(i) = \mathbf{H}(i) [\mathbf{x}_{\text{IBI}}^T(i), \mathbf{x}^T(i)]^T + \mathbf{v}_t(i), \quad (2.41)$$

where we use $\mathbf{v}_t(i)$ to denote the noise, which is similarly defined as $\mathbf{y}_t(i)$; further

$$\mathbf{x}_{\text{IBI}}(i) := [x_{i(N+L_z)-L}, \dots, x_{i(N+L_z)-1}]^T \quad (2.42)$$

[†]Usually, the data symbols are the output of an error-control encoder. Despite the significance of channel encoding to OFDM, we will not look at this aspect in the thesis and only concentrate on the uncoded case.

stands for the inter-block interference (IBI), which is comprised of the last L symbols from the previous block $\mathbf{x}(i-1)$; $\mathbf{H}(i)$ represents the convolutive operation of the channel:

$$\mathbf{H}(i) := \begin{bmatrix} h_{i(N+L_z),L} & \cdots & h_{i(N+L_z),0} & & 0 \\ & \ddots & \vdots & \ddots & \\ 0 & & h_{(i+1)(N+L_z)-1,L} & \cdots & h_{(i+1)(N+L_z)-1,0} \end{bmatrix}. \quad (2.43)$$

Let us decompose the matrix $\mathbf{H}(i)$ into two parts, where $\mathbf{H}_t(i)$ consists of the last $N + L_z$ columns of $\mathbf{H}(i)$ corresponding to the present transmitted block $\mathbf{x}(i)$:

$$\mathbf{H}_t(i) := \begin{bmatrix} h_{i(N+L_z),0} & & & & \\ \vdots & \ddots & & & \\ h_{i(N+L_z)+L,L} & \cdots & h_{i(N+L_z)+L,0} & & \\ & \ddots & \vdots & \ddots & \\ & & h_{(i+1)(N+L_z)-1,L} & \cdots & h_{(i+1)(N+L_z)-1,0} \end{bmatrix}, \quad (2.44)$$

and $\mathbf{H}_{\text{IBI}}(i)$ consists of the first L columns of $\mathbf{H}(i)$ corresponding to the IBI $\mathbf{x}_{\text{IBI}}(i)$:

$$\mathbf{H}_{\text{IBI}}(i) := \begin{bmatrix} h_{i(N+L_z),L} & \cdots & h_{i(N+L_z),1} \\ & \ddots & \vdots \\ & & h_{i(N+L_z)+L-1,L} \\ \hline & & \mathbf{0}_{(N+L_z-L) \times L} \end{bmatrix}. \quad (2.45)$$

In this way, (2.41) can be written as

$$\mathbf{y}_t(i) = \mathbf{H}_t(i)\mathbf{x}(i) + \mathbf{H}_{\text{IBI}}(i)\mathbf{x}_{\text{IBI}}(i) + \mathbf{v}_t(i). \quad (2.46)$$

If we assume perfect block and symbol synchronization, a pre-processing will be applied at the receiver, which contains the operation of discarding the CP and applying an FFT. The output of this pre-processing step, which is defined as

$$\mathbf{y}_f(i) := [y_{f,iN}, \cdots, y_{f,(i+1)N-1}]^T, \quad (2.47)$$

will admit the expression

$$\begin{aligned} \mathbf{y}_f(i) &= \mathbf{W}_N \mathbf{R}_{\text{cp}} \mathbf{r}(i) \\ &= \mathbf{W}_N \mathbf{R}_{\text{cp}} \mathbf{H}_t(i)\mathbf{x}(i) + \mathbf{W}_N \mathbf{R}_{\text{cp}} \mathbf{H}_{\text{IBI}}(i)\mathbf{x}_{\text{IBI}}(i) + \mathbf{v}_f(i) \end{aligned} \quad (2.48)$$

where

$$\mathbf{v}_f(i) := \mathbf{W}_N \mathbf{R}_{\text{cp}} \mathbf{v}(i) \quad (2.49)$$

and

$$\mathbf{R}_{\text{cp}} := \left[\mathbf{0}_{N \times L_z} \mid \mathbf{I}_N \right] \quad (2.50)$$

stands for the operation of removing the CP, which discards the first L_z samples of $\mathbf{H}_t(i)\mathbf{x}(i)$ and $\mathbf{H}_{\text{IBI}}(i)\mathbf{x}_{\text{IBI}}(i)$. For the latter, we understand that only the first L elements may not be zero [c.f. (2.45)]. This implies that if the CP is sufficiently long, i.e., $L_z \geq L$, the IBI term will vanish,

$$\mathbf{R}_{\text{cp}}\mathbf{H}_{\text{IBI}}(i)\mathbf{x}_{\text{IBI}}(i) = \mathbf{0}. \quad (2.51)$$

As a result and taking (2.38) into account, we obtain

$$\begin{aligned} \mathbf{y}_f(i) &= \mathbf{W}_N \mathbf{H}_{t,\text{cp}}(i) \mathbf{W}_N^H \mathbf{s}(i) + \mathbf{v}_f(i) \\ &= \mathbf{H}_{f,\text{cp}}(i) \mathbf{s}(i) + \mathbf{v}_f(i), \end{aligned} \quad (2.52)$$

where

$$\mathbf{H}_{t,\text{cp}}(i) := \mathbf{R}_{\text{cp}} \mathbf{H}_t(i) \mathbf{T}_{\text{cp}} \quad (2.53)$$

stands for the time-domain (TD) channel and

$$\mathbf{H}_{f,\text{cp}}(i) := \mathbf{W}_N \mathbf{H}_{t,\text{cp}}(i) \mathbf{W}_N^H \quad (2.54)$$

for the frequency-domain (FD) channel. More specifically, $\mathbf{H}_{t,\text{cp}}(i)$ takes on a pseudo-circulant form as

$$\mathbf{H}_{t,\text{cp}}(i) = \begin{bmatrix} h_{i(N+L_z)+L_z,0} & & h_{i(N+L_z)+L_z,L} & \cdots & h_{i(N+L_z)+L_z,1} \\ \vdots & \ddots & & \ddots & \vdots \\ h_{i(N+L_z)+L_z+L,L} & & & & h_{i(N+L_z)+L_z+L-1,L} \\ & \ddots & & \ddots & \\ & & \ddots & & \\ & & & h_{(i+1)(N+L_z)-1,L} & \cdots & h_{(i+1)(N+L_z)-1,0} \end{bmatrix} \quad (2.55)$$

If the channel is static, or only slowly changing such that the time-variation of the channel within the OFDM symbol can be neglected, $\mathbf{H}_{t,\text{cp}}(i)$ will become truly circulant. In other words, the entries of $\mathbf{H}_{t,\text{cp}}(i)$ are constant along each diagonal (in a circulant sense). In the frequency domain, this results in a diagonal $\mathbf{H}_{f,\text{cp}}(i)$:

$$\mathbf{H}_{f,\text{cp}}(i) = \mathcal{D}\{\sqrt{N}\mathbf{W}_N[h_0, \dots, h_L]^T\}, \quad (2.56)$$

where we have dropped the time index n in the subscript of the channel tap $h_{n,l}$ [c.f. (2.13)]. The columns (rows) of the diagonal $\mathbf{H}_{f,\text{cp}}(i)$, which are better known as the

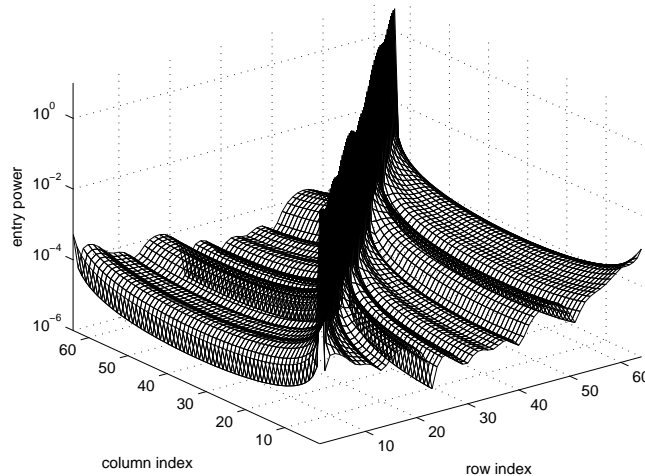


Figure 2.5: Entry powers of $\mathbf{H}_{f,cp}(i)$ for $N = 64$ and $\nu_D = 10^{-5}$

subcarriers of the OFDM symbol, are obviously orthogonal to each other. Fig. 2.6 shows such an example, where the x -axis and y -axis correspond to the rows and columns of $\mathbf{H}_{f,cp}(i)$, respectively, and the z -axis corresponds to the entry power.

However, when the channel is varying faster, the circularity of $\mathbf{H}_{f,cp}(i)$ is destroyed, which is therefore not diagonalizable by the (I)FFT operations. In principle, $\mathbf{H}_{f,cp}(i)$ becomes a full matrix. The non-zero off-diagonal entries inflict inter-carrier interference (ICI). As we understand from the previous chapter, the Doppler shifts due to mobility will cause a frequency expansion of the signal bandwidth. This effect is reflected in the FD channel matrix as the power of $\mathbf{H}_{f,cp}(i)$ will not be concentrated on the diagonal alone, but spread out. Fortunately, for realistic Doppler spreads, it is reasonable to assume the following.

Assumption 2.1. *The frequency-domain channel matrix $\mathbf{H}_{f,cp}(i)$ is roughly banded[‡], which means that the channel power is mainly located in the entries close to the main diagonal, and reduces gradually in an anti-diagonal direction.*

This assumption is advocated in [96, 14], and can be observed from Fig. 2.6, which is obtained for $\nu_D = 0.004$. Of course, if the channel varies even faster, the

[‡]Strictly speaking, we actually mean “circularly-banded” here. However, we will use the term “banded” in the sequel for the sake of brevity.

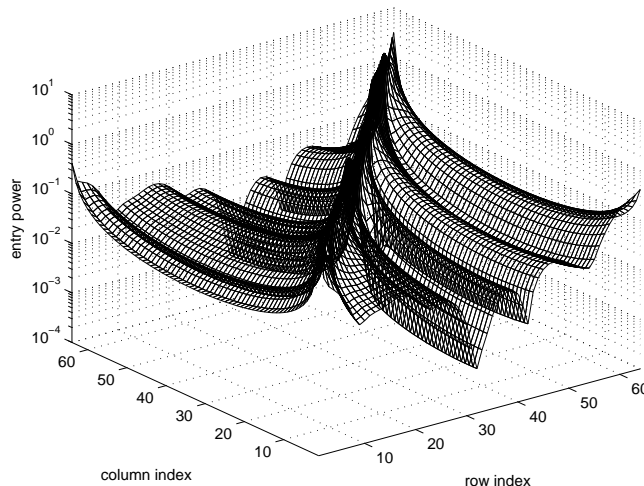


Figure 2.6: Entry powers of $\mathbf{H}_{f,cp}(i)$ for $N = 64$ and $\nu_D = 0.004$

channel power will be distributed further away from the diagonal, and thus the band assumption will be more critical. In such a case, we need to enlarge the assumed bandwidth to account for more off-diagonals. Throughout the thesis, we will exploit this band property of the FD channel matrix, but the thereby induced band approximation error will also be taken into account.

2.3.2 Other OFDM System Models

In the previous section, transmitter redundancy in the form of CP is introduced to counteract the IBI. As a matter of fact, other transmitter redundancy form are also applicable. We will briefly explain them next.

ZP-OFDM

ZP-OFDM is characterized by a slightly different system model than CP-OFDM. We will illustrate this by assuming an adequately long guard interval $L_z = L$. The i th data block $\mathbf{s}(i)$ undergoes first an IFFT and next a precoder \mathbf{T}_{zp} , which leads to an $(N + L) \times 1$ vector $\mathbf{x}(i)$,

$$\mathbf{x}(i) = \mathbf{T}_{zp} \mathbf{W}_N^H \mathbf{s}(i), \quad (2.57)$$

where the precoder

$$\mathbf{T}_{\text{zp}} := \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{L \times N} \end{bmatrix} \quad (2.58)$$

is used to append L zeros (guard interval) at the end of the vector $\mathbf{W}_N^H \mathbf{s}(i)$. With $\mathbf{x}(i)$ transmitted, we obtain at the receiver $\mathbf{y}(i)$, which has the same I/O relationship as defined in (2.46). A two-step pre-processing is applied at the receiver, which is comprised of a decoder \mathbf{R}_{zp} and an FFT, leading to the received block in the frequency domain $\mathbf{y}_{f,\text{zp}}(i)$

$$\mathbf{y}_{f,\text{zp}}(i) = \mathbf{W}_N \mathbf{R}_{\text{zp}} \mathbf{H}_t(i) \mathbf{x}(i) + \mathbf{W}_N \mathbf{R}_{\text{zp}} \mathbf{H}_{\text{IBI}}(i) \mathbf{x}_{\text{IBI}}(i) + \mathbf{v}_{f,\text{zp}}(i), \quad (2.59)$$

where the noise term here is defined as $\mathbf{v}_{f,\text{zp}}(i) := \mathbf{W}_N \mathbf{R}_{\text{zp}} \mathbf{v}(i)$, and the decoder \mathbf{R}_{zp} is defined as

$$\mathbf{R}_{\text{zp}} := \left[\mathbf{I}_N \mid \frac{\mathbf{I}_L}{\mathbf{0}_{(N-L) \times L}} \right] \quad (2.60)$$

which adds the last L entries to the first L entries of the vector $\mathbf{H}(i) \mathbf{x}(i)$.

Because the precoder is applied to each transmitted block, this means that $\mathbf{x}_{\text{IBI}}(i) = \mathbf{0}$. As a result, and if we take (2.57) into account, (2.59) can be rewritten as

$$\begin{aligned} \mathbf{y}_{f,\text{zp}}(i) &= \mathbf{W}_N \mathbf{H}_{t,\text{zp}}(i) \mathbf{W}_N^H \mathbf{s}(i) + \mathbf{v}_f(i) \\ &= \mathbf{H}_{f,\text{zp}}(i) \mathbf{s}(i) + \mathbf{v}_f(i) \end{aligned} \quad (2.61)$$

where

$$\mathbf{H}_{t,\text{zp}}(i) := \mathbf{R}_{\text{zp}} \mathbf{H}_t(i) \mathbf{T}_{\text{zp}} \quad (2.62)$$

stands for the TD channel matrix, and $\mathbf{H}_{f,\text{zp}}(i)$ for its FD counterpart

$$\mathbf{H}_{f,\text{zp}}(i) := \mathbf{W}_N \mathbf{H}_{t,\text{zp}}(i) \mathbf{W}_N^H. \quad (2.63)$$

For the channel matrix $\mathbf{H}(i)$ defined in (2.44), we can easily show that

$$\mathbf{H}_{t,\text{zp}}(i) = \begin{bmatrix} h_{i(N+L_z)+L_z,0} & & h_{k(N+L_z)+L_z,L} & \cdots & h_{i(N+L_z)+L_z,1} \\ \vdots & \ddots & & \ddots & \vdots \\ h_{i(N+L_z)+L_z+L,L} & & & & h_{i(N+L_z)+L_z+L-1,L} \\ & \ddots & & & \\ & & \ddots & & \\ & & & h_{(i+1)(N+L_z)-1,L} & \cdots & h_{(i+1)(N+L_z)-1,0} \end{bmatrix} \quad (2.64)$$

which is exactly the same as the channel matrix of CP-OFDM $\mathbf{H}_{t,cp}(i)$ that is defined in (2.55).

It is noteworthy that there is another way to implement ZP-OFDM by employing an $(N + L)$ -point FFT directly at the receiver. By this means, we will obtain a larger FD channel with $N + L$ subcarriers, whose orthogonality remains in TI channels. This practice is especially significant to diversity gain. Recall in the CP-OFDM, it is possible that a channel zero falls on one of the subcarriers. As a result, the data symbol transmitted at that subcarrier will not be detectable if no channel coding is applied leading to a diversity loss. Using ZP instead of CP and employing a larger-scaled FFT can avoid this problem [35, 86, 87, 117].

NZP-OFDM

In ZP-OFDM, zeros are used as guard interval. Actually, this can be generalized to the case where some pre-defined symbols are inserted between the transmitted blocks, which is also referred to as the known symbol padding (KSP) in [25, 79]. In this case, the k th transmitted block $\mathbf{x}(i)$ can be expressed as

$$\mathbf{x}(i) = \mathbf{P} \begin{bmatrix} \mathbf{W}_N^H \mathbf{s}(i) \\ \mathbf{t}(i) \end{bmatrix} \quad (2.65)$$

where $\mathbf{t}(i)$ stands for an $L_z \times 1$ vector that is inserted in between the blocks, and

$$\mathbf{P} = \mathbf{I}_{N+L_z}. \quad (2.66)$$

Note that the precoding expressed in the above can also account for the CP and ZP case, where in case of ZP, we have

$$\begin{aligned} \mathbf{t}(i) &= \mathbf{0}_{L_z \times 1}, \\ \mathbf{P} &= \mathbf{I}_{N+L_z}, \end{aligned} \quad (2.67)$$

while in case of CP, $\mathbf{t}(i)$ consists of the last L_z entries of $\mathbf{W}_N^H \mathbf{s}(i)$ and

$$\mathbf{P} = \begin{bmatrix} \mathbf{0}_{L \times N} & \mathbf{I}_L \\ \mathbf{I}_N & \mathbf{0}_{N \times L} \end{bmatrix}.$$

By assuming some non-zero values, the guard interval $\mathbf{t}(i)$ can also serve for channel estimation, which can be of interest for the sake of bandwidth efficiency [25, 79]. We will not discuss the system model of the NZP-OFDM in depth, but reiterate some important properties here.

1. In case of TI channels, the resulting channel matrix will be circulant provided that the KSP is sufficiently long ($L_z \geq L$) and the KSP between each transmitted blocks are identical, i.e.,

$$\mathbf{t}(i) = \mathbf{t}(i'), \quad (2.68)$$

for $i \neq i'$. This will lead to a diagonal FD channel matrix

2. This diagonality will, however, in case of TV channels, be destroyed just as in the CP- and ZP-OFDM cases.

2.3.3 Single-carrier Transmission

Compared to OFDM, the transmitter in a single-carrier system does not need to go through the IFFT. Except for this, we can establish the channel I/O relationship analogously to the OFDM system: the k th data symbol block is first precoded to

$$\mathbf{x}(i) = \mathbf{T}\mathbf{s}(i), \quad (2.69)$$

by e.g., inserting the CP or ZP (NZP). It is next sent over the channel. The same channel I/O relationship as in the OFDM case (2.46) can be applied here. At the receiver, an $N \times (N + L)$ decoding matrix \mathbf{R} is deployed, which produces $\mathbf{y}_t(i)$ as :

$$\begin{aligned} \mathbf{y}_t(i) &= \mathbf{R}\mathbf{H}_t(i)\mathbf{x}(i) + \mathbf{R}\mathbf{H}_{\text{IBI}}(i)\mathbf{x}_{\text{IBI}}(i) + \mathbf{R}\mathbf{v}_t(i) \\ &= \underbrace{\mathbf{R}\mathbf{H}_t(i)\mathbf{T}}_{\mathbf{H}_{t,sc}(i)}\mathbf{s}(i) + \mathbf{R}\mathbf{v}_t(i), \end{aligned} \quad (2.70)$$

where in the last equality, we omit the IBI term by assuming an adequately long guard interval. In the above, the precoder \mathbf{T} and decoder \mathbf{R} correspond to \mathbf{T}_{cp} and \mathbf{R}_{cp} , respectively if a CP is inserted, whereas \mathbf{T} and \mathbf{R} correspond to \mathbf{T}_{zp} and \mathbf{R}_{zp} , respectively if ZP is inserted. Both cases lead to the same channel matrix

$$\mathbf{H}_{t,sc}(i) = \mathbf{H}_{t,cp}(i) = \mathbf{H}_{t,zp}(i). \quad (2.71)$$

It is in practice more common to transform the data model of the single-carrier system into the frequency domain as

$$\begin{aligned} \mathbf{y}_f(i) &= \mathbf{W}_N\mathbf{y}(i) \\ &= \mathbf{H}_{f,sc}(i)\mathbf{W}_N\mathbf{s}(i) + \mathbf{v}_f(i), \end{aligned} \quad (2.72)$$

where $\mathbf{H}_{f,sc}(i)$ stands for the FD channel matrix

$$\mathbf{H}_{f,sc}(i) = \mathbf{W}_N\mathbf{H}_{t,sc}(i)\mathbf{W}_N^H. \quad (2.73)$$

Equation (2.72) is especially significant to low-complexity equalization. If the channel stays stable within the block, $\mathbf{H}_{f,sc}(i)$ will be diagonal, which enables a simple one-tap equalizer [30]. For TV channels, it can be still appealing to equalize the channel in the frequency domain, though the diagonality cannot be relied upon any more. In that case, we can make use of Assumption 2.1, which views the FD channel matrix as banded. Supposing the bandwidth is smaller than the channel delay spread, it is still cheaper to implement an FD equalizer. This idea will be explored in Chapter 6.

Chapter 3

Channel Equalization in an OFDM System

3.1 Introduction

In a high-mobility scenario, where the channel time-variation within one OFDM symbol period cannot be neglected, the orthogonality among the subcarriers will be destroyed by the induced Doppler spread, which results in a FD channel matrix that is full instead of diagonal. This implies that a reliable equalizer must take into account not only the frequency response of the channel at each subcarrier, but also the inter-carrier interference (ICI), leading to a high complexity.

To enable a cheap equalizer, pre-processing at the receiver is indispensable. For instance, in [7, 8], an FIR filter is adopted in an endeavor to restore the diagonality of the channel matrix by ‘flattening’ the channel’s fluctuation. Such a technique works well for channels that are moderately spread in delay and Doppler dimensions. Practically, a strictly diagonal FD channel matrix is too difficult to achieve. We have mentioned in the previous chapter that for a practical Doppler spread, the FD channel matrix has most of its power concentrated in the vicinity of the diagonal in a circular sense, with those entries that are far away from the diagonal reducing fast [96, 14]. This implies that in practice, we can assume the FD channel matrix to be circularly banded. Many other existing works exploit this banded (rather than diagonal) structure when designing the equalizer, e.g., the block zero-forcing (ZF) equalizer in [46], the block minimum mean square error (MMSE) equalizer in [81, 82], the serial iterative MMSE equalizer in [90], the maximum likelihood (ML) equalizer in [71, 49], etc. It can be imagined that to enhance the equalization performance, especially at a moderate to high signal-to-noise ratio (SNR), the band approximation error must be reduced as much as possible. One solution can be the FIR filter of [8], but it generally requires a multiple antenna assumption and can be considered ‘over-designed’ since it aims to achieving a diagonal FD channel matrix. In comparison, for a banded assumption as will be adopted in the following, a reduced-order FIR filter with just a single tap could be adequate. Such a filter is referred to as a receiver window in [90, 82, 44], which is essentially an element-wise multiplication

Part of the results of this chapter appeared in [100].

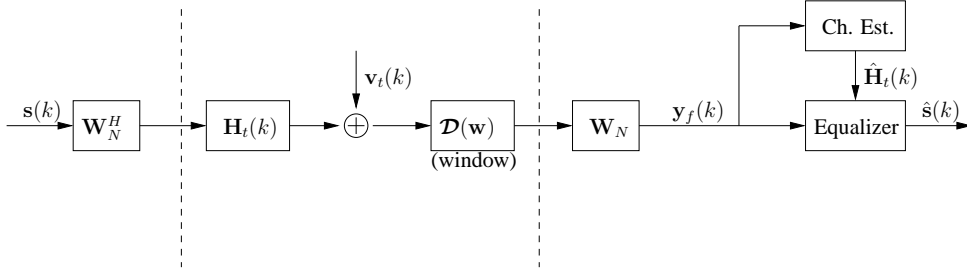


Figure 3.1: The transceiver block diagram

of the window entries with the received signal.

In the following, we will present an equalization scheme that utilizes a strictly banded channel matrix assumption to lower the complexity. However, unlike [90, 82, 44] where a strictly banded matrix is obtained by simply copying the significant diagonals from the original FD channel matrix, the band approximation problem in this chapter is tackled in a different way: we will devise a strictly banded matrix that is close to its full counterpart only in terms of the Frobenius norm. Although this leads eventually to the same solution as in [90, 82, 44], we can by this means translate the band approximation error in the frequency domain into a basis expansion modeling (BEM) error in the time domain. More specifically, we can establish the link between the band approximation with the (C)CE-BEM. This will serve an enlightening purpose when we design a low-complexity equalizer for a single-carrier system in Chapter 6.

3.2 System Model

Let us consider a CP-OFDM system with N subcarriers, as illustrated in Fig. 3.1. Following a similar approach as described in Sec. 2.3.1 but introducing a windowing operation in front of the FFT, we can readily establish the channel I/O relationship for the i th OFDM symbol $\mathbf{s}(i)$ as follows

$$\mathbf{y}_t(i) = \mathcal{D}\{\mathbf{w}\}\mathbf{H}_t(i)\mathbf{W}_N^H\mathbf{s}(i) + \mathcal{D}\{\mathbf{w}\}\mathbf{v}_t(i), \quad (3.1)$$

where $\mathcal{D}\{\mathbf{w}\}$ stands for the windowing operation, which is a diagonal matrix with

$$\mathbf{w} := [w_0, \dots, w_{N-1}]^T \quad (3.2)$$

as its diagonal. $\mathbf{v}_t(i)$ stands for the corresponding noise in the time domain. $\mathbf{H}_t(i)$ denotes the TD channel matrix without windowing. If we define $h_{n,l}$ as the l th

channel tap at the n th time-instance, and assume that $h_{n,l}$ has finite order L , i.e., $h_{n,l} = 0$ for $l < 0$ or $l > L$, then $h_{n,l}$ is related to $\mathbf{H}_t(k)$ as

$$[\mathbf{H}_t(i)]_{p,q} = h_{i(N+L_z)+L_z+p,\text{mod}(p-q,N)}, \quad (3.3)$$

where $\text{mod}(m, n)$ in the subscript stands for the residue of m divided by n , which corresponds to the ‘‘circulant’’ nature of the channel matrix $\mathbf{H}_t(i)$ as a result of inserting and removing the CP.

The received samples in $\mathbf{y}_t(i)$ are afterwards demodulated by the DFT matrix \mathbf{W}_N , resulting into

$$\begin{aligned} \mathbf{y}_f(i) &= \mathbf{W}_N \mathbf{y}_t(i) \\ &= \mathbf{H}_f(i) \mathbf{s}(i) + \mathbf{v}_f(i), \end{aligned} \quad (3.4)$$

with

$$\mathbf{H}_f(i) := \mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathbf{H}_t(i) \mathbf{W}_N^H \quad (3.5)$$

and

$$\mathbf{v}_f(i) := \mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathbf{v}_t(i)$$

denoting the channel matrix and noise in the frequency domain, respectively. Note that in Chapter 2, we have introduced the system model for CP-OFDM in (2.52). That is actually a special case of (3.4) here, which assumes a rectangular window

$$\mathbf{w} = \mathbf{1}_{N \times 1}. \quad (3.6)$$

Because the equalization discussed in this chapter will be implemented upon each OFDM symbol independently, we will drop the symbol index i in the sequel for the sake of brevity.

3.3 A Low-Complexity Equalization Scheme

We will in this section present a block linear minimum mean square error (LMMSE) equalizer, which is constructed based on a banded channel matrix approximation. As a matter of fact, most other equalization schemes than the LMMSE equalizer can also benefit from this idea to lower the complexity, e.g., those equalizers that are mentioned in the introduction of this chapter.

The following assumptions will be useful.

Assumption 3.1. *We assume that the TV channel is a wide-sense stationary uncorrelated scattering (WSSUS) process, for which*

$$\mathcal{E}_h \{h_{p,l} h_{p-m,l-n}\} = \sigma_l^2 \gamma_m \delta_n. \quad (3.7)$$

Here, δ_n denotes the Kronecker delta, σ_l^2 the variance of the l th channel tap, and γ_m the normalized temporal correlation.

Assumption 3.2. We assume the data symbols to be zero-mean white with unit variance, i.e.,

$$\mathcal{E}_s\{s_p s_{p-m}^*\} = \delta_m, \quad (3.8)$$

and the noise prior to windowing to be zero-mean white with variance σ^2 , i.e.,

$$\mathcal{E}_v\{v_p v_{p-m}^*\} = \sigma^2 \delta_m. \quad (3.9)$$

Starting from the I/O relationship in (3.4) and conform Assumption 3.2, the block LMMSE equalizer can be constructed following the similar approach as described in [56], which results in the data estimates $\hat{\mathbf{s}}$ as

$$\hat{\mathbf{s}} = \mathbf{H}_f^H (\mathbf{H}_f \mathbf{H}_f^H + \mathbf{R}_v)^{-1} \mathbf{y}_f, \quad (3.10)$$

where \mathbf{R}_v denotes the noise covariance matrix. Under the zero-mean white assumption of the noise and with the window taken into account, it can be shown that

$$\mathbf{R}_v := \mathcal{E}\{\mathbf{v}_f \mathbf{v}_f^H\} = \sigma^2 \mathbf{W}_N \mathcal{D}\{\mathbf{w} \odot \mathbf{w}^*\} \mathbf{W}_N^H. \quad (3.11)$$

It is easy to understand that most computational efforts in (3.10) are invested in inverting the covariance matrix, which are high considering that \mathbf{H}_f is in principle a full matrix. In order to lower the complexity, we will adopt Assumption 2.1 and approximate the full matrix \mathbf{H}_f in (3.10) by a strictly banded matrix $\hat{\mathbf{H}}_f$. The resulting LMMSE equalizer becomes

$$\hat{\mathbf{s}} \approx \hat{\mathbf{H}}_f^H (\hat{\mathbf{H}}_f \hat{\mathbf{H}}_f^H + \mathbf{R}_v)^{-1} \mathbf{y}_f. \quad (3.12)$$

In the above, $\hat{\mathbf{H}}_f$ is strictly banded with bandwidth $Q + 1$, which means that it has only non-zero entries on the main diagonal, the $Q/2$ super- and $Q/2$ sub-diagonals in a circular sense. Here, Q is a design parameter that can be chosen as a trade-off between complexity and performance. The resulting equalizer will be made cheaper with a smaller Q , but accordingly suffer from more out-of-band interference. In reality, the minimum value of Q is often related to the maximal Doppler spread.

It is noteworthy that a similar approach (band approximation and windowing) is reported in [90, 82, 44]. However, the difference is subtle: in [90, 82, 44], $\hat{\mathbf{H}}_f$ is obtained by simply taking the significant diagonals from \mathbf{H}_f , while in this thesis, this is not necessarily the case. As a matter of fact, we use the following cost function to design the non-zero diagonals of $\hat{\mathbf{H}}_f$,

$$\begin{aligned} & \arg \min_{\{\hat{\mathbf{H}}_f\}} \|\mathbf{H}_f - \hat{\mathbf{H}}_f\|^2, \\ & \text{s.t. } \hat{\mathbf{H}}_f = \hat{\mathbf{H}}_f \odot \mathbf{T}_Q. \end{aligned} \quad (3.13)$$

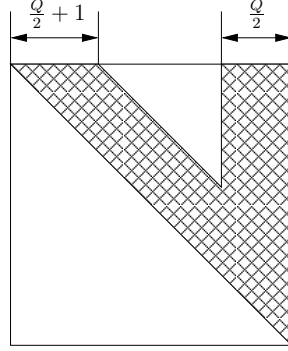


Figure 3.2: The “V-shape” structure of \mathbf{G} . The shaded part corresponds to non-zero entries.

where \mathbf{T}_Q is a matrix of appropriate dimension, which has ones on the main diagonal, the $Q/2$ super- and $Q/2$ sub-diagonals, and zeros on the remaining entries. We will discuss this matter in more detail in Remark 3.2.

Another important feature in (3.12) is that we also require the noise covariance matrix \mathbf{R}_v to be strictly banded with a bandwidth $2Q + 1$, just like the product $\hat{\mathbf{H}}_f \hat{\mathbf{H}}_f^H$. If this is achieved as we will see later on, we can apply the Cholesky factorization [37] on the covariance matrix

$$\hat{\mathbf{H}}_f \hat{\mathbf{H}}_f^H + \mathbf{R}_v = \mathbf{G} \mathbf{G}^H, \quad (3.14)$$

where the upper-triangular matrix \mathbf{G} will assume a sparse “V-shape” structure as illustrated in Fig. 3.2. By this means, the inverse of $\hat{\mathbf{H}}_f \hat{\mathbf{H}}_f^H + \mathbf{R}_v$ can be achieved by inverting \mathbf{G} and \mathbf{G}^H separately using Gaussian elimination.

For an overview of the total invoked complexity, we make a list of each step that is involved in computing (3.12) in the left panel of Table 3.1.

Taking the band structure of $\hat{\mathbf{H}}_f$ into account, we understand that step 1 requires $N(Q + 1)(1.5Q + 1)$ complex multiply and accumulate operations (MACs). Step 2 is realized by applying the Cholesky factorization, which entails $N(\frac{7}{8}Q^2 + \frac{13}{4}Q + 2) - \frac{1}{6}Q^3 - \frac{3}{2}Q^2 + \frac{5}{6}Q$ MACs [44]. Due to the sparse “V-shape” structure of \mathbf{G} , the Gaussian elimination required in steps 3 and 4 need $N(2Q + 1) - 2Q^2 - Q$ MACs each. Step 5 needs another $N(2Q + 1)$ MACs. The complexity of each step is listed in the right panel of Table 3.1. In summary, the considered block LMMSE equalizer has a complexity of $\mathcal{O}(NQ^2)$, which is linear in N and quadratic to Q .

We understand that the equalizer in (3.12) uses a banded $\hat{\mathbf{H}}_f$ to approximate \mathbf{H}_f . The complexity and performance of such an equalizer are determined by Q , the bandwidth of $\hat{\mathbf{H}}_f$. The question arises that whether it is possible to enhance

Table 3.1: Complexity of block LMMSE equalization for OFDM

Steps in (3.12)	MACs required per step
1. $\mathbf{M} = \hat{\mathbf{H}}_f \hat{\mathbf{H}}_f^H + \mathbf{R}_v;$	$N(Q+1)(1.5Q+1)$
2. $\mathbf{M} = \mathbf{G}\mathbf{G}^H;$	$N(\frac{7}{8}Q^2 + \frac{13}{4}Q + 2) - \frac{1}{6}Q^3 - \frac{3}{2}Q^2 + \frac{5}{6}Q$
3. $\boldsymbol{\rho} = \mathbf{G}^{-1}\mathbf{y}_f;$	$N(2Q+1) - 2Q^2 - Q$
4. $\mathbf{1} = \mathbf{G}^{-H}\boldsymbol{\rho};$	$N(2Q+1) - 2Q^2 - Q$
5. $\mathbf{x} := \hat{\mathbf{H}}_f^H \mathbf{1};$	$N(2Q+1)$

the equalization performance for a fixed complexity (Q). In other words, can we minimize the band approximation error $\|\mathbf{H}_f - \hat{\mathbf{H}}_f\|$. From the definition of \mathbf{H}_f in (3.5), we understand that this minimization problem is a function of the window. In addition, we need to design the window such that the noise covariance matrix \mathbf{R}_v in (3.11) will also be banded. These issues will be discussed next.

3.4 Window Design

Before we address the window design, let us first introduce the notation \mathbf{B} to denote the (C)CE-BEM matrix with scale $Q+1$, which is defined in (2.30). Besides, we present the following lemma, which will be useful not only for the ensuing part, but also for Chapter 6.

Lemma 3.1. *Any $N \times N$ strictly banded matrix \mathbf{M} with bandwidth $Q+1$ can be viewed as the Fourier transform of a sum of $Q+1$ matrices, where each term is a circulant matrix weighted by a special diagonal matrix. In mathematics, this can be expressed as*

$$\mathbf{M} = \mathbf{W}_N \sum_{q=-\frac{Q}{2}}^{\frac{Q}{2}} \mathcal{D}\{\mathbf{B}\mathbf{e}_q\} \mathbf{C}_q \mathbf{W}_N^H, \quad (3.15)$$

where \mathbf{e}_q stands for the q th column of an identity matrix, and \mathbf{C}_q for a circulant matrix with

$$\mathbf{c}_q = [c_{q,0}, \dots, c_{q,N-1}]^T \quad (3.16)$$

as its first column.

Proof. Let us use an $N \times 1$ vector \mathbf{m}_q to denote each diagonal of \mathbf{M} , for $q = -\frac{Q}{2}, \dots, \frac{Q}{2}$. As a result,

$$\mathbf{M} = \sum_q \mathbf{E}_q \mathcal{D}\{\mathbf{m}_q\}, \quad (3.17)$$

where \mathbf{E}_q represents a shift matrix defined in (2.31). Due to Property 2.1, we have

$$\mathbf{W}_N \mathcal{D}\{\mathbf{B}\mathbf{e}_q\} \mathbf{W}_N^H = \mathbf{E}_{q-\frac{Q}{2}}. \quad (3.18)$$

Further, it is easy to understand from (2.56) that if

$$\mathbf{c}_q = \mathbf{W}_N^H \mathbf{m}_q, \quad (3.19)$$

the diagonal matrix $\mathcal{D}\{\mathbf{m}_q\}$ can be related with a circulant matrix \mathbf{C}_q , which uses \mathbf{c}_q as its first column:

$$\mathcal{D}\{\mathbf{m}_q\} = \mathbf{W}_N \mathbf{C}_q \mathbf{W}_N^H. \quad (3.20)$$

By substituting (3.17) and (3.20) in (3.17), and swapping the DFT matrix and the summation, we reach (3.15). \square

To begin with the window design, we first study the noise shaping behavior of the window. The approach of [82] is adopted in this thesis, which is summarized in the following proposition.

Proposition 3.1. *The noise covariance matrix \mathbf{R}_v will be strictly banded with bandwidth $2Q + 1$ if we let the window \mathbf{w} be a weighted sum of $Q + 1$ complex exponentials:*

$$\mathbf{w} = \mathbf{B}\mathbf{d}. \quad (3.21)$$

Proposition 3.1 (see Appendix 3A for a proof) tells us that we can structure the window as a weighted sum of $Q + 1$ complex exponentials. In (2.56), we have shown that the subcarriers of OFDM over a time-invariant channel are related with the time-domain (TD) channel taps in the same way as (3.21). Hence, it is not difficult to understand that the windowing operation, which is an element-wise multiplication of the window with the received signal in the time domain, can also be interpreted as a filtering operation of the received signal in the frequency domain with an FIR filter. To be specific, this FIR filter is equipped with $Q + 1$ taps, which are collected in the vector \mathbf{d} . As a result, the window design boils down to the design of an FIR filter. We will come back to this later on.

Next, we will show how to minimize the band approximation error $\|\mathbf{H}_f - \hat{\mathbf{H}}_f\|$. To this end, we need to design the window and the banded matrix $\hat{\mathbf{H}}_f$ jointly. The following theorem proves to be important (see Appendix 3B for a proof).

Theorem 3.1. *The band approximation error $\|\mathbf{H}_f - \hat{\mathbf{H}}_f\|$, which is expressed in the frequency domain, can be transformed in the time domain as the error resulting from the (C)CE-BEM. In mathematics, this can be expressed as*

$$\|\mathbf{H}_f - \hat{\mathbf{H}}_f\| = \|\mathcal{D}\{\mathbf{w}\}\mathcal{H} - \mathbf{B}\mathbf{C}\|. \quad (3.22)$$

In the above, $\|\cdot\|$ stands for the Frobenius norm of a matrix; \mathcal{H} stands for an $N \times (L+1)$ matrix collecting all the channel taps $[\mathcal{H}]_{n,l} = h_{n,l}$, and \mathbf{C} for a $(Q+1) \times (L+1)$ matrix collecting all the coefficients $[\mathbf{C}]_{q,l} = c_{q,l}$.

The right-hand side (RHS) of (3.22) resembles those works that use a BEM, \mathbf{B} in this context, to fit to the (windowed) TV channel $\mathcal{D}\{\mathbf{w}\}\mathcal{H}$ [c.f. (2.23)]. In this sense, the band approximation error can be interpreted as a BEM modeling error (we will use these two terms in the sequel interchangeably), and thus $c_{q,l}$ is also referred to as a BEM coefficient. In particular, with the entries defined as

$$[\mathbf{B}]_{p,q} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi}{N} p(q - \frac{Q}{2})}, \quad (3.23)$$

this BEM corresponds to the classical definition of the (C)CE-BEM in (2.30), whose exponential period equals the BEM window size N .

In line with Theorem 3.1, we can now come up with the following design problem:

$$\begin{aligned} \min_{\{\mathbf{w}\}} \mathcal{E}_h \{ \min_{\{\mathbf{C}\}} \|\mathcal{D}\{\mathbf{w}\}\mathcal{H} - \mathbf{B}\mathbf{C}\|^2 \}, \\ \text{s.t. } \mathbf{w} = \mathbf{B}\mathbf{d} \text{ and } \|\mathbf{w}\|^2 = N. \end{aligned} \quad (3.24)$$

Note that the first constraint above is due to Proposition 3.1, and the second constraint is imposed to avoid a trivial all-zero window. We solve (3.24) first for \mathbf{C} :

$$\mathbf{C} = \mathbf{B}^\dagger \mathcal{D}\{\mathbf{w}\}\mathcal{H}. \quad (3.25)$$

Defining

$$\begin{aligned} \mathcal{P}_\mathbf{B} &:= \mathbf{I} - \mathbf{B}\mathbf{B}^\dagger, \\ \mathbf{R}_\mathcal{H} &:= \mathcal{E}_h \{ \mathcal{H}\mathcal{H}^H \}, \end{aligned} \quad (3.26)$$

we can show that the cost function then becomes

$$\begin{aligned} \mathcal{E}_h \{ \|\mathcal{D}\{\mathbf{w}\}\mathcal{H} - \mathbf{B}\mathbf{C}\|^2 \} &= \text{tr}(\mathcal{P}_\mathbf{B} \mathcal{D}\{\mathbf{w}\} \mathbf{R}_\mathcal{H} \mathcal{D}\{\mathbf{w}^H\} \mathcal{P}_\mathbf{B}^H) \\ &= \sum_{n=0}^{N-1} \mathbf{e}_n^T \mathcal{P}_\mathbf{B} \mathcal{D}\{\mathbf{w}\} \mathbf{R}_\mathcal{H} \mathcal{D}\{\mathbf{w}^H\} \mathcal{P}_\mathbf{B}^H \mathbf{e}_n. \end{aligned} \quad (3.27)$$

Under the WSSUS channel property of Assumption 3.1, it is easy to see that

$$[\mathbf{R}_{\mathcal{H}}]_{m,n} = \sum_{l=0}^L \sigma_l^2 \gamma_{m-n}. \quad (3.28)$$

For two vectors \mathbf{a} and \mathbf{b} of the same size,

$$\mathbf{a}^T \mathcal{D}\{\mathbf{b}\} = \mathbf{b}^T \mathcal{D}\{\mathbf{a}\}. \quad (3.29)$$

Applying this property to the second equality of (3.27) leads to

$$\mathcal{E}_h\{\|\mathcal{D}\{\mathbf{w}\}\mathcal{H} - \mathbf{B}\mathbf{C}\|^2\} = \mathbf{w}^T \left(\sum_{n=0}^{N-1} \mathcal{D}\{\mathbf{e}_n^T \mathcal{P}_{\mathbf{B}}\} \mathbf{R}_{\mathcal{H}} \mathcal{D}\{\mathcal{P}_{\mathbf{B}}^H \mathbf{e}_n\} \right) \mathbf{w}^*. \quad (3.30)$$

If we substitute (3.21) in (3.30), the design problem in (3.24) can thus be equivalently rewritten as

$$\begin{aligned} \min_{\{\mathbf{d}\}} \mathbf{d}^T \mathcal{X} \mathbf{d}^*, \\ \text{s.t. } \|\mathbf{d}\|^2 = N, \end{aligned} \quad (3.31)$$

with

$$\mathcal{X} := \mathbf{B}^T \left(\sum_{n=0}^{N-1} \mathcal{D}\{\mathbf{e}_n^T \mathcal{P}_{\mathbf{B}}\} \mathbf{R}_{\mathcal{H}} \mathcal{D}\{\mathcal{P}_{\mathbf{B}}^H \mathbf{e}_n\} \right) \mathbf{B}^*. \quad (3.32)$$

Hence, \mathbf{d} can be computed as the eigenvector corresponding to the least significant eigenvalue of \mathcal{X}^* .

A few remarks are in order now.

Remark 3.1.

In view of (3.22), we are able to interpret the band approximation error between \mathbf{H}_f and $\hat{\mathbf{H}}_f$ in the frequency domain as the (C)CE-BEM modeling error in the time-domain. It is thus not difficult to realize that the equalization performance is influenced by the (C)CE-BEM modeling performance. This effect will be explored in Chapter 6.

Remark 3.2.

With \mathbf{C} obtained in (3.25), we can show that the strictly banded matrix $\hat{\mathbf{H}}_f$ indeed takes the $Q + 1$ most significant diagonals of \mathbf{H}_f , i.e.,

$$\hat{\mathbf{H}}_f = \mathbf{H}_f \odot \mathbf{T}_Q. \quad (3.33)$$

where \mathbf{T}_Q denotes a selection matrix that is defined in (3.13). Accordingly, the discrepancy $\|\mathbf{H}_f - \hat{\mathbf{H}}_f\|$ is actually the out-of-band interference

The proof of (3.33) is given in Appendix 3C. It is worth underlining that although this observation coincides with the canonical band approximation approach that is adopted in [90, 82, 44], the underlying difference is not trivial. In this thesis, $\hat{\mathbf{H}}_f$ is sought using the criterion in (3.14). In Chapter 6 where single-carrier systems are treated, we will see that this criterion leads to a different band approximation than (3.33).

Remark 3.3.

A similar windowing strategy is presented in [90], which maximizes the signal to interference (band approximation error) and noise ratio directly in the frequency domain. As a matter of fact, the band approximation error considered in [90] can also be translated into a (C)CE-BEM modeling error just like in this thesis. Indeed, in the noiseless case we can show that the window of [90] will admit the same expression as the window here (see Appendix 3D for a proof). As a result, the performances of these two windows will be very close to each other if the same equalizer is employed.

3.5 Simulation Results

We consider an OFDM system with $N = 256$ subcarriers. To test the proposed algorithms over a TV channel following Jakes' Doppler profile [45], we use the TV channel generator that is given in [124]. The channel is assumed to have $L + 1 = 6$ channel taps with the l th tap having a variance $\sigma_l^2 = c e^{-\frac{l}{10}}$. Here, c denotes a normalization constant such that $\sum_l \sigma_l^2 = 1$. QPSK modulated data symbols are transmitted.

We first examine the proposed equalizer for TV channels with normalized Doppler spread $\nu_D = 0.0008$. To lower the equalization complexity, we approximate the FD channel matrix as a banded matrix. Windows that are optimally designed for bandwidths $Q = 4, 8, 20$ are exhibited in the upper plot of Fig. 3.3. For equalization, we test the windows for $Q = 4$ and $Q = 8$. Their BER performances are plotted in Fig. 3.4. In addition, we include the performance of the equalizer based on the same band approximation but without windowing. Eventually, we use the equalizer based on the full FD channel matrix as the performance benchmark (denoted as "full block" in Fig. 3.4). It is clear that with respect to the full block equalizer, the equalizers without windowing suffer from a large performance gap, and a high noise floor emerges at high SNR. In contrast, the equalizers with windowing renders a close performance to the full block equalizer.

The same simulations are carried out for TV channels with $\nu_D = 0.004$, where the windows for $Q = 4, 8, 20$ are depicted in the lower plot of Fig. 3.3. The equalization performance is shown in Fig. 3.5, from which similar remarks can be made as in the slower channel case.

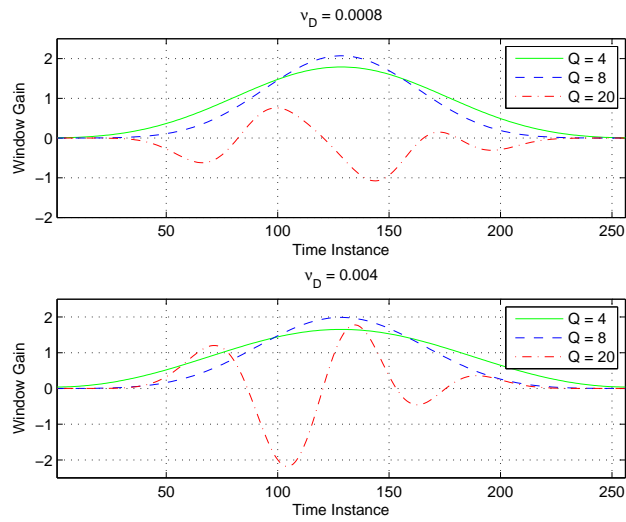


Figure 3.3: Windows as Function of the Normalized Doppler Spread and Chosen Bandwidth

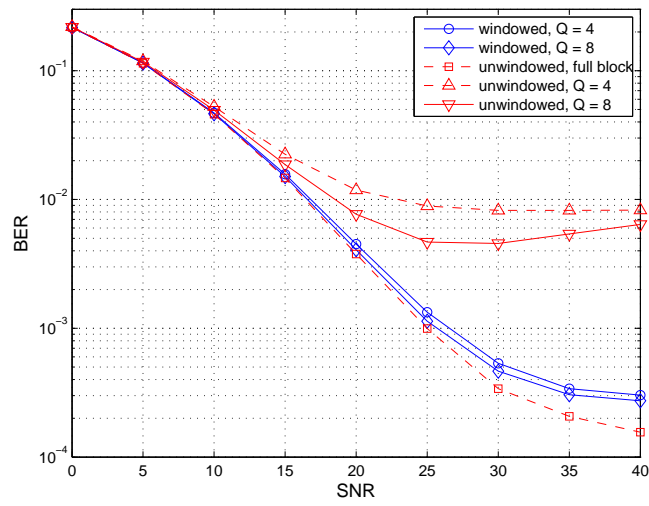


Figure 3.4: Equalization Performance for TV channels with $\nu_D = 0.0008$.

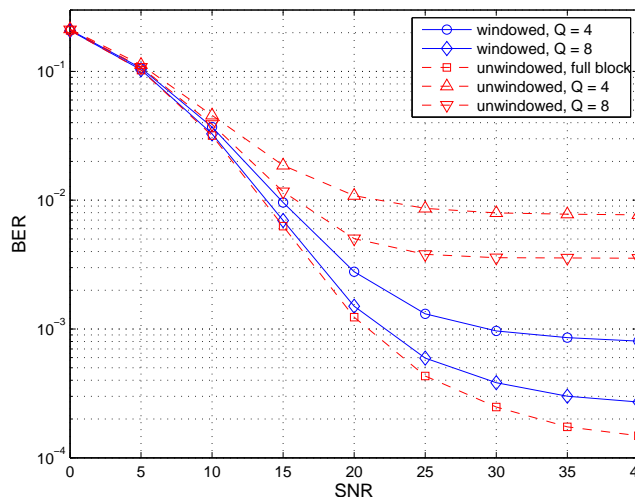


Figure 3.5: Equalization Performance for TV channels with $\nu_D = 0.004$.

3.6 Summary

In this chapter, we have discussed how to equalize a time-varying channel in an OFDM system. Due to the time-variation, the channel matrix in the frequency domain is not diagonal but full.

To enable a low-complexity equalizer, we have explored the fact that for realistic Doppler spreads, most of the channel power is concentrated in the vicinity of the diagonal, such that the FD channel matrix can be approximated as banded. We have shown that the out-of-band interference can be transformed into the modeling error of the (C)CE-BEM in the time domain, which can be further reduced by applying a window at the receiver.

We used the block LMMSE equalizer to equalize the channel. Thanks to the “banded” structure, the inversion of the covariance matrix, which is most complex part of the block LMMSE equalizer, can be simplified using structured Cholesky factorization. We have shown that the proposed block LMMSE equalizer can yield a satisfactory performance with a complexity that is linear in the number of subcarriers, just like that is for time-invariant channels. To compare, traditional techniques inflict a complexity that is cubic in the number of subcarriers in order to achieve a similar performance.

For the approaches proposed in this chapter, the channel knowledge is indispensable. Channel estimation will be discussed in the next two chapters.

Appendix 3A: Proof of Proposition 3.1

Let us rewrite the noise covariance matrix in (3.11) as

$$\mathbf{R}_v = \sigma^2 \underbrace{\mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathbf{W}_N^H}_{\mathcal{W}} \underbrace{\mathbf{W}_N \mathcal{D}\{\mathbf{w}^*\} \mathbf{W}_N^H}_{\mathcal{W}^H}. \quad (3.34)$$

Obviously, \mathbf{R}_v will be banded if

$$\mathcal{W} := \mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathbf{W}_N^H \quad (3.35)$$

is banded. Using Lemma 3.1, we express it in a general form as

$$\mathcal{W} = \mathbf{W}_N \sum_q \mathcal{D}\{\mathbf{B}\mathbf{e}_q\} \mathbf{C}_q \mathbf{W}_N^H. \quad (3.36)$$

Comparing (3.36) with (3.35), we require that

$$\sum_q \mathcal{D}\{\mathbf{B}\mathbf{e}_q\} \mathbf{C}_q = \mathcal{D}\{\mathbf{w}\}, \quad (3.37)$$

which can be accomplished by e.g., letting $\mathcal{D}\{\mathbf{B}\mathbf{e}_q\} \mathbf{C}_q$ be a diagonal matrix. For the circulant \mathbf{C}_q , this is possible if we let its first column \mathbf{c}_q be equal to

$$\mathbf{c}_q = \left[[\mathbf{d}]_q, \mathbf{0}_{1 \times (N-1)} \right]^T. \quad (3.38)$$

Substituting the above in (3.37), and looking only at the non-zero main diagonal leads to (3.21).

Appendix 3B: Proof of Theorem 3.1

Let us use

$$\hat{\mathbf{H}}_t := \mathbf{W}_N^H \hat{\mathbf{H}}_f \mathbf{W}_N, \quad (3.39)$$

to denote the time-domain counterpart of $\tilde{\mathbf{H}}_f$. Following Lemma 3.1, we can easily derive that

$$\hat{\mathbf{H}}_t = \sum_q \mathcal{D}\{\mathbf{B}\mathbf{e}_q\} \mathbf{C}_q, \quad (3.40)$$

and hence the band approximation error can be expressed in the time domain as

$$\|\mathbf{H}_f - \hat{\mathbf{H}}_f\| = \|\mathcal{D}\{\mathbf{w}\}\mathbf{H}_t - \sum_q \mathcal{D}\{\mathbf{B}\mathbf{e}_q\}\mathbf{C}_q\|. \quad (3.41)$$

The above holds because the Frobenius norm is considered here, which will not be affected by the existence of the unitary matrix \mathbf{W}_N . For the same reason, we can reshape $\mathcal{D}\{\mathbf{w}\}\mathbf{H}_t$ by stacking each of its diagonals in the columns of the following matrix

$$\mathcal{D}\{\mathbf{w}\}\mathbf{H}_t \sim \mathcal{D}\{\mathbf{w}\} [\mathcal{H} \quad \mathbf{0}_{N \times (N-L+1)}]. \quad (3.42)$$

Similarly, we coin a new matrix out of $\sum_q \mathcal{D}\{\mathbf{B}\mathbf{e}_q\}\mathbf{C}_q$

$$\sum_q \mathcal{D}\{\mathbf{B}\mathbf{e}_q\}\mathbf{C}_q \sim \mathbf{B} [\mathbf{C} \quad \mathbf{C}^\perp], \quad (3.43)$$

where \mathbf{C}^\perp stands for an $(Q+1) \times (N-L+1)$ matrix with entries $[\mathbf{C}^\perp]_{n,l} = c_{q,L+l+1}$. As a result, we obtain from (3.41) that

$$\|\mathbf{H}_f - \hat{\mathbf{H}}_f\| = \|\mathcal{D}\{\mathbf{w}\} [\mathcal{H} \quad \mathbf{0}_{N \times (N-L+1)}] - \mathbf{B} [\mathbf{C} \quad \mathbf{C}^\perp]\|. \quad (3.44)$$

Since we ultimately want to minimize the difference above, as a first step, we can already let

$$\mathbf{C}^\perp = \mathbf{0}_{N \times (N-L+1)}, \quad (3.45)$$

which leads to (3.22).

Appendix 3C: Proof of (3.33)

We have defined the BEM matrix \mathbf{B} to be the first $Q/2 + 1$ and the last $Q/2$ columns of the unitary DFT matrix \mathbf{W}_N . Let us use \mathbf{B}^\perp to denote an $N \times (N-Q-1)$ matrix, which is comprised of the remaining columns of \mathbf{W}_N , and thus spans the noise subspace of \mathbf{B} . Obviously, the (C)CE-BEM modeling error resulting from (3.25) can be expressed as

$$\mathcal{D}\{\mathbf{w}\}\mathcal{H} - \mathbf{B}\mathbf{C} = \mathbf{B}^\perp\mathbf{C}^\perp, \quad (3.46)$$

where the $(N-Q-1) \times (L+1)$ matrix \mathbf{C}^\perp has entries $[\mathbf{C}^\perp]_{m,l} = c_{Q+1+m,l}$. For the time-domain channel matrix $\mathcal{D}\{\mathbf{w}\}\mathbf{H}_t$, we thus have

$$\mathcal{D}\{\mathbf{w}\}\mathbf{H}_t = \underbrace{\sum_{q=0}^Q \mathcal{D}\{\mathbf{B}\mathbf{e}_q\}\mathbf{C}_q}_{\hat{\mathbf{H}}_t} + \underbrace{\sum_{q=Q+1}^N \mathcal{D}\{\mathbf{B}^\perp\mathbf{e}_{q-Q-1}\}\mathbf{C}_q}_{\hat{\mathbf{H}}_t^\perp}, \quad (3.47)$$

where \mathbf{C}_q is a circulant matrix with the first column defined as

$$\mathbf{C}_q \mathbf{e}_0 = [c_{q,0}, \dots, c_{q,L}, \mathbf{0}_{1 \times (N-L-1)}]^T. \quad (3.48)$$

The BEM modeling error transformed in the frequency domain

$$\mathbf{W}_N \hat{\mathbf{H}}_t^\perp \mathbf{W}_N^H = \sum_{q=Q+1}^N \mathbf{W}_N \mathcal{D}\{\mathbf{B}^\perp \mathbf{e}_{q-Q-1}\} \mathbf{W}_N^H \mathbf{W}_N \mathbf{C}_q \mathbf{W}_N^H \quad (3.49)$$

is a summation of $N - Q - 1$ matrices, where each factor $\mathbf{W}_N \mathbf{C}_q \mathbf{W}_N^H$ is a diagonal matrix, and

$$\mathbf{W}_N \mathcal{D}\{\mathbf{B}^\perp \mathbf{e}_{q-Q-1}\} \mathbf{W}_N^H = \mathbf{E}_q \quad (3.50)$$

due to Property 2.1. Hence, $\mathbf{W}_N \hat{\mathbf{H}}_t^\perp \mathbf{W}_N^H$ will occupy the $(Q/2 + 2)$ nd until $(N - Q/2)$ th super-diagonals of \mathbf{H}_f in a circular sense, which thus does not overlap with $\hat{\mathbf{H}}_f$.

Appendix 3D: Equivalence of the Proposed Window and the Window of [90]

We rewrite the cost function in (3.27) as

$$\begin{aligned} \mathbf{w} &= \arg \min \text{tr}(\mathcal{P}_B \mathcal{D}\{\mathbf{w}\} \mathbf{R}_{\mathcal{H}} \mathcal{D}\{\mathbf{w}^H\}), \\ \text{s.t. } &\|\mathbf{w}\|^2 = N. \end{aligned} \quad (3.51)$$

where we have used the fact that

$$\mathcal{P}_B \mathcal{P}_B^H = \mathcal{P}_B. \quad (3.52)$$

It is straightforward to understand that

$$\mathcal{P}_B = \mathbf{B}^\perp \mathbf{B}^{\perp H}, \quad (3.53)$$

where the $N \times (N - Q - 1)$ matrix \mathbf{B}^\perp is defined in Appendix 3.6, and thus has entries

$$[\mathbf{B}^\perp]_{m,n} = \frac{1}{\sqrt{N}} e^{j \frac{2\pi}{N} m(n + \frac{Q}{2} + 1)}. \quad (3.54)$$

Hence, the cost function in (3.51) becomes

$$\text{tr}(\mathbf{B}^{\perp H} \mathcal{D}\{\mathbf{w}\} \mathbf{R}_{\mathcal{H}} \mathcal{D}\{\mathbf{w}^H\} \mathbf{B}^\perp) = \mathbf{w}^T \underbrace{\left(\sum_{k=0}^{N-Q-2} \mathcal{D}\{\mathbf{B}^{\perp*} \mathbf{e}_k\} \mathbf{R}_{\mathcal{H}} \mathcal{D}\{\mathbf{B}^\perp \mathbf{e}_k\} \right)}_{\mathbf{\Gamma}} \mathbf{w}^*. \quad (3.55)$$

For each entry of $\mathbf{\Gamma}$, we have

$$[\mathbf{\Gamma}]_{m,n} = \frac{1}{N} [\mathbf{R}_{\mathcal{H}}]_{m,n} \sum_{k=0}^{N-Q-2} e^{j \frac{2\pi}{N} (k + \frac{Q}{2} + 1)(n-m)}. \quad (3.56)$$

Because

$$\sum_{k=0}^{N-Q-2} e^{j \frac{2\pi}{N} (k + \frac{Q}{2} + 1)(n-m)} = \begin{cases} N - Q - 1 & \text{if } m = n, \\ -\frac{\sin(\pi(Q+1)(n-m)/N)}{\sin(\pi(n-m)/N)} & \text{otherwise,} \end{cases} \quad (3.57)$$

we can show after some algebra that

$$\mathbf{\Gamma} = \frac{1}{N} \mathbf{R}_{\mathcal{H}} \odot \mathbf{I}_N - \mathbf{R}_{\mathcal{H}} \odot \mathbf{A}, \quad (3.58)$$

with \mathbf{A} being an $N \times N$ matrix with entries

$$[\mathbf{A}]_{m,n} = \frac{\sin(\pi(Q+1)(n-m)/N)}{N \sin(\pi(n-m)/N)}. \quad (3.59)$$

Substituting (3.58) in (3.55), we can rewrite the minimization problem as

$$\begin{aligned} \mathbf{w} &= \arg \min \frac{1}{N} \mathbf{w}^T (\mathbf{R}_{\mathcal{H}} \odot \mathbf{I}_N) \mathbf{w}^* - \mathbf{w}^T (\mathbf{R}_{\mathcal{H}} \odot \mathbf{A}) \mathbf{w}^*, \\ \text{s.t. } &\|\mathbf{w}\|^2 = N. \end{aligned} \quad (3.60)$$

Under the WSSUS channel property in Assumption 3.1, we know that $\mathbf{R}_{\mathcal{H}}$ has equal diagonal elements and thus $\mathbf{R}_{\mathcal{H}} \odot \mathbf{I}_N$ is a scaled identity matrix. As a result, the above minimization problem can be equivalently rewritten as

$$\begin{aligned} \mathbf{w} &= \arg \max \mathbf{w}^T (\mathbf{R}_{\mathcal{H}} \odot \mathbf{A}) \mathbf{w}^*, \\ \text{s.t. } &\|\mathbf{w}\|^2 = N. \end{aligned} \quad (3.61)$$

which is exactly the same cost function as in [90] in the absence of noise and IBI.

Chapter 4

Channel Estimation Based on a Single OFDM Symbol

4.1 Introduction

In order to design an equalizer for OFDM, the channel state information (CSI), or specifically, the value of the FD channel matrix \mathbf{H}_f , is indispensable, which can be acquired using the methods presented in this and the next chapter.

What is common to the channel estimators that will be described in the following is that they are aimed at first acquiring in the time domain the channel tap $h_{n,l}$ in (3.3). This means that there will be $N(L + 1)$ channel unknowns with N the OFDM size and L the channel order in the time domain, which are generally too many to allow for an efficient channel estimator. The idea of the BEM will be adopted here such that the number of channel unknowns can be reduced from $N(L + 1)$ to $(L + 1)(Q + 1)$, where Q stands for the BEM scale. Assuming that the BEM only induces a negligible modeling error by choosing a proper Q , the channel estimation boils down to estimating the BEM coefficients.

The works in [55, 19, 111] belong to the few that focus on blind BEM channel estimation, which rely on a large number of observation samples. In the context of TV channels, the channel coherence time is much shorter, which implies that a large number of relevant observation samples might not be always available. [10, 94, 59] propose pilot-assisted channel estimators for single-carrier transmission systems, where pilots are clustered in the time domain such that the channel estimation can be realized without interference from neighboring data symbols. For frequency-domain communication systems such as OFDM, it is not clear what is the best strategy to place the pilots. Because of ICI, the FD channel matrix is in principle full, which makes it impossible to find an observation sample that is solely dependent on pilots and thus not contaminated by data symbols. For this reason, many existing works view the frequency-domain channel matrix either as diagonal [69, 5, 122] thus ignoring ICI completely, or strictly banded like [47] that relies on a (C)CE-BEM

The results of this chapter appeared in [103, 97].

assumption. These approaches suffer from a large estimation error for faster channels (as we know that the (C)CE-BEM is not a tight model). An opposite approach is to respect the full property of the FD channel matrix and let the pilots occupy a whole OFDM symbol [18, 38, 21]. Since in this approach, the data has to be sent in other OFDM symbols, where the channel is obviously not the same, some extra- or interpolation techniques must be employed. The latter implies that multiple pilot OFDM symbols need to be deployed for channel estimation, an idea that will be discussed more in detail in the next chapter.

In this chapter, we discuss how to acquire the CSI within one OFDM symbol, where pilots will be clustered and interleaved with data symbols in the frequency domain. We will view the FD channel matrix as approximately banded, which means that we still consider a full channel matrix, but take the fact into account that the channel power is mostly concentrated around the main diagonal and getting smaller in the anti-diagonal direction. This view complies with the observations made in [96, 14] and can be represented by most BEMs discussed in Chapter 2 except for the (C)CE-BEM (the last will result in an exactly banded matrix). However, the bandwidth of such an approximately banded channel matrix is ambiguous to define, and if we artificially select a clear-cut bandwidth, the out-of-band entries will give rise to interference. We will show that by taking the out-of-band interference smartly into account in traditional estimator designs, such as the LMMSE estimator or the best linear unbiased estimator (BLUE), we can improve the estimation accuracy. This is in contrast to the least squares (LS) estimator, which requires the interference to be kept as small as possible. In other words, the amount of interference we take into account has a significant impact on each individual estimator. This effect will be analyzed and a criterion to select the optimal amount of interference for different types of channel estimators will be proposed.

4.2 System Model

4.2.1 OFDM System Model

The same system model that was presented in the previous chapter will be used. For the sake of self-completeness, we summarize the main points here. In the time domain, the received samples in the i th block are first stripped off from the CP and subsequently windowed. The resulting signals are related to the transmitted symbols as

$$\mathbf{y}_t(i) = \mathcal{D}\{\mathbf{w}\}\mathbf{H}_t(i)\mathbf{W}_N^H\mathbf{s}(i) + \mathcal{D}\{\mathbf{w}\}\mathbf{v}_t(i), \quad (4.1)$$

where $\mathcal{D}\{\mathbf{w}\}$ represents the windowing operation at the receiver; $\mathbf{v}_t(i)$ stands for the corresponding noise in the time domain. $\mathbf{H}_t(i)$ stands for the TD channel matrix

without windowing. If we use $h_{n,l}$ to represent the l th channel tap at the n th time-instance, and assume that it has a finite order L , i.e., $h_{n,l} = 0$ for $l < 0$ or $l > L$, then $h_{n,l}$ is related to the TD channel matrix $\mathbf{H}_t(i)$ as

$$[\mathbf{H}_t(i)]_{p,q} = h_{i(N+L_z)+L_z+p,\text{mod}(p-q,N)}. \quad (4.2)$$

In the above, L_z stands for the length of the CP; the notation $\text{mod}(a, b)$ in the subscript stands for the residue of a divided by b , which is used to account for the ‘circulant’ nature of $\mathbf{H}_t(i)$. More details can be found in Section 2.3.1.

The received samples in $\mathbf{y}_t(i)$ are afterwards demodulated by the DFT matrix \mathbf{W}_N , resulting in

$$\begin{aligned} \mathbf{y}_f(i) &= \mathbf{W}_N \mathbf{y}_t(i) \\ &= \mathbf{H}_f(i) \mathbf{s}(i) + \mathbf{v}_f(i), \end{aligned} \quad (4.3)$$

with

$$\mathbf{H}_f(i) := \mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathbf{H}_t(i) \mathbf{W}_N^H \quad (4.4)$$

denoting the FD channel matrix, and $\mathbf{v}_f(i)$ the FD noise which is similarly defined as $\mathbf{y}_f(i)$.

4.2.2 BEM Model in the Presence of a Window

Because the low-complexity equalizer discussed in the previous chapter relies on a windowed channel, it can be convenient to estimate the windowed channel directly, which is approximated first in this section by using a BEM. To this end, let us introduce an $N \times 1$ vector $\mathbf{h}_l(i)$, which collects the realization of the l th channel tap within the period of the i th OFDM symbol after windowing as

$$\mathbf{h}_l(i) := [w_0 h_{i(N+L_z)+L_z,l}, \dots, w_{N-1} h_{(i+1)(N+L_z)-1,l}]^T. \quad (4.5)$$

It admits an expression as

$$\mathbf{h}_l(i) = \mathbf{B} [c_{0,l}(i), \dots, c_{Q,l}(i)]^T + \boldsymbol{\epsilon}_l(i), \quad (4.6)$$

where \mathbf{B} is an $N \times (Q+1)$ BEM matrix, and $c_{q,l}(i)$ represents the q th BEM coefficient for the l th channel tap in the k th OFDM symbol. $\boldsymbol{\epsilon}_l(i)$ represents an $N \times 1$ vector corresponding to the BEM modeling error, which is assumed to be minimized in the Mean Squared Error (MSE) sense. Stacking all the channel taps within the i th observation block in one vector

$$\begin{aligned} \mathbf{h}(i) &:= [w_0 h_{i(N+L_z)+L_z,0}, \dots, w_0 h_{i(N+L_z)+L_z,L}, \dots, \\ &\quad w_{N-1} h_{(i+1)(N+L_z)-1,0}, \dots, w_{N-1} h_{(i+1)(N+L_z)-1,L}]^T, \end{aligned} \quad (4.7)$$

we obtain

$$\mathbf{h}(i) = (\mathbf{B} \otimes \mathbf{I}_{L+1})\mathbf{c}(i) + \boldsymbol{\epsilon}(i), \quad (4.8)$$

where

$$\mathbf{c}(i) := [c_{0,0}(i), \dots, c_{0,L}(i), \dots, c_{Q,0}(i), \dots, c_{Q,L}(i)]^T, \quad (4.9)$$

and $\boldsymbol{\epsilon}(i)$ is similarly defined as $\mathbf{h}(i)$.

With abuse of notation, let us here use the notation $\mathbf{B}_{\text{canonical}}$ to represent the various canonical BEM designs that have been described in Chapter 2. These canonical BEMs were actually designed for approximating TV channels without windowing. The question arises whether $\mathbf{B}_{\text{canonical}}$ is still a good BEM design now that a window is present. Basically, we have two options

$$\mathbf{B} = \begin{cases} \mathbf{B}_{\text{canonical}} & \text{(i),} \\ \mathcal{D}\{\mathbf{w}\}\mathbf{B}_{\text{canonical}}\mathbf{Q} & \text{(ii).} \end{cases} \quad (4.10)$$

In the above, we introduce a square matrix \mathbf{Q} to ensure that the columns of \mathbf{B} are orthonormal, just like $\mathbf{B}_{\text{canonical}}$. Its usage is in general not mandatory but can simplify the analysis as will be clear later on. We observe in (4.10) that option (i) ignores the window in the BEM design and sticks to the traditional ‘unwindowed’ BEMs to model the windowed channels whereas option (ii) includes the knowledge of the window in the BEM design. We will show in Section 4.5 that option (ii) generally yields a tighter fit. This is because the window itself brings some additional time-variation on top of the unwindowed channel, which probably requires more basis functions for the canonical “unwindowed” BEM to maintain a tight fit. However, the time-variation due to the window is totally predictable and hence can be counteracted by simply absorbing it in our BEM design. The (C)CE-BEM is an exception especially for channels with high Doppler spreads, in which case it is better to use option (i). This is because the window is usually designed to make the frequency-domain channel matrix as banded as possible, and the (C)CE-BEM channel matrix is already perfectly banded in the frequency domain. Therefore, it is not necessary to include the window in the (C)CE-BEM design.

4.2.3 OFDM System Model in Light of BEM

From now on, we will describe the OFDM system model derived above in light of the BEM, and drop the symbol index i in the remainder of the chapter since all the algorithms in this paper will be based on a single OFDM symbol. Substituting (4.8)

in (4.3), we obtain after some algebra that

$$\begin{aligned} \mathbf{y}_f &= \mathbf{H}_f \mathbf{s} + \mathbf{v}_f \\ &= \sum_{q=0}^Q \mathbf{D}_q \mathbf{\Delta}_q \mathbf{s} + \mathbf{v}_f + \boldsymbol{\delta}. \end{aligned} \quad (4.11)$$

In the second equality above, we have used a BEM expression $\sum_q \mathbf{D}_q \mathbf{\Delta}_q$ to replace the FD channel matrix \mathbf{H}_f , and put the effect of the corresponding BEM fitting error in the vector $\boldsymbol{\delta}$. It can be shown that in the BEM channel expression, \mathbf{D}_q is a circulant matrix with as first column the frequency response of the q th basis function

$$\mathbf{D}_q := \mathbf{W}_N \mathcal{D}\{\mathbf{b}_q\} \mathbf{W}_N^H, \quad (4.12)$$

and $\mathbf{\Delta}_q$ is a diagonal matrix with as diagonal the frequency response of the BEM coefficients corresponding to the q th basis function

$$\mathbf{\Delta}_q := \mathcal{D}\{\mathbf{G}_L [c_{q,0}, \dots, c_{q,L}]^T\}. \quad (4.13)$$

Here, \mathbf{G}_L stands for the first $L + 1$ columns of the matrix $\sqrt{N} \mathbf{W}_N$.

Note that (4.11) subsumes the expression for TI channels, in which case $Q = 0$ and \mathbf{D}_q becomes a scaled identity matrix. For TV channels, the non-diagonal entries of \mathbf{D}_q are in general not zero any more. This leads to a loss of orthogonality among the subcarriers known as ICI.

Before going on further, we will make the following assumption

Assumption 4.1. *The BEM approximation holds perfectly, i.e., $\boldsymbol{\epsilon} = \mathbf{0}$ or $\boldsymbol{\delta} = \mathbf{0}$.*

This assumption is motivated by the fact that we will mainly focus on BEMs that allow for a very good fit. For other BEMs that fail to capture the time-variation adequately, such as the (C)CE-BEM for instance, we should actually take the modeling error into account. This topic is partly treated in [80]. However, even if we include this error term to derive the best estimator possible, we still do not have a reliable channel estimate simply because the BEM itself is not capable of fitting the true channel. This suggests that it makes not much sense to take the modeling error into account and explains why we apply Assumption 4.1 for all possible BEMs.

4.3 Data Model for Channel Estimation

Instead of estimating the true bulky FD channel matrix \mathbf{H}_f , we will estimate the $(L + 1)(Q + 1)$ BEM coefficients collected in \mathbf{c} with the aid of pilots. We assume that there are M pilot clusters of length $D + 1$, with the m th pilot cluster denoted as \mathbf{t}_m



Figure 4.1: The pilot placement pattern

for $m = 0, 1, \dots, M - 1$. The pilot structure is indicated in Fig. 4.1. All these pilot clusters stacked together form the pilot vector

$$\mathbf{t} = \mathbf{s}^{\{\mathcal{T}\}} := [\mathbf{t}_0, \dots, \mathbf{t}_{M-1}]^T, \quad (4.14)$$

where \mathcal{T} is used to denote the set collecting the positions of the pilots. Here, we use the notation $\mathbf{a}^{\{\mathcal{S}_r\}}$ ($\mathbf{A}^{\{\mathcal{S}_r\}}$) to represent the elements of the vector \mathbf{a} (the rows of the matrix \mathbf{A}), whose indices are collected in the set \mathcal{S}_r . Likewise, we use \mathcal{D} to denote the set of all the data symbol positions, and hence all the data symbols can be collectively represented by $\mathbf{s}^{\{\mathcal{D}\}}$. For frequency-multiplexed pilot scheme (where a null subcarrier will be viewed as carrying a zero pilot), this means obviously that $\mathcal{T} \cap \mathcal{D} = \emptyset$.

It is not clear in terms of the channel estimation MSE what is the optimal pilot structure, e.g., the value and position of each pilot. The implication of the pilot structure on the channel identifiability will be discussed in more detail in the next chapter. Despite its significance, we will not discuss this issue here but allow our receiver design to be applicable to any frequency-multiplexed pilot placement scheme.

For a certain frequency-multiplexed pilot placement scheme, it is up to the receiver to decide which of the received samples must be used for channel estimation. This is crucial for a TV OFDM system since, due to ICI (or in other words, the non-zero off-diagonal entries of \mathbf{D}_q), the training power is spread out over the whole frequency band. A judicious choice of the observation samples will enhance the channel estimation performance.

Generally speaking, \mathbf{D}_q is approximately (circularly) banded suggesting that the ICI primarily comes from adjacent subcarriers [96, 14]. An extreme case occurs in the case of a (C)CE-BEM channel model, where the corresponding \mathbf{D}_q is an identity matrix but circularly shifted over $q - \frac{Q}{2}$ columns, which implies that only the Q neighboring subcarriers give rise to interference.

To clarify the notations that will be used, we plot the structure of \mathbf{D}_q in Fig. 4.2, where the columns of \mathbf{D}_q are related to the positions of the pilots and data, which operate on \mathbf{D}_q through the diagonal matrix $\mathbf{\Delta}_q$. The rows of \mathbf{D}_q are related to the

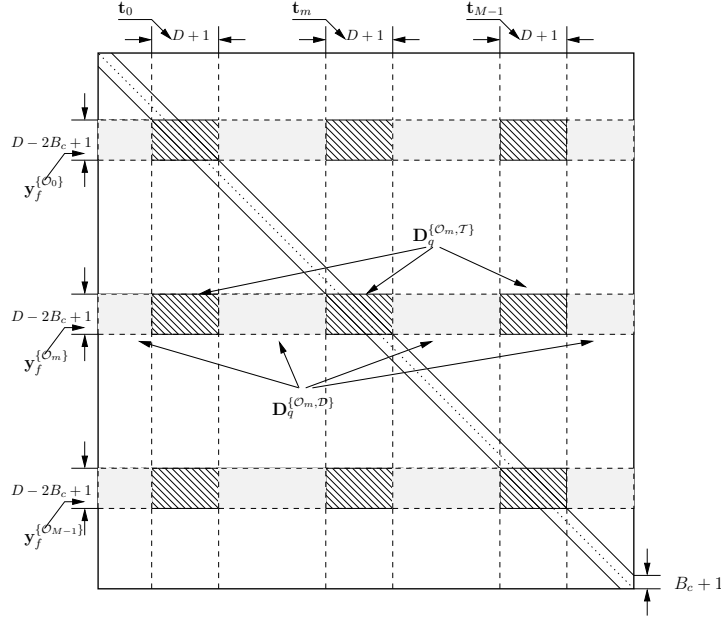


Figure 4.2: The partitioning of \mathbf{D}_q . Its rows corresponds to the positions of the received samples; its columns to the positions of the pilots and data.

observation samples. For the m th pilot cluster

$$\mathbf{t}_m = \mathbf{s}^{\{\mathcal{T}_m\}}, \quad (4.15)$$

with

$$\mathcal{T}_m := \{P_m, \dots, P_m + D\}, \quad (4.16)$$

where P_m stands for the position of the first pilot in \mathbf{t}_m , let us consider the following vector consisting of $D - 2B_c + 1$ observation samples:

$$\mathbf{y}_f^{\{\mathcal{O}_m\}} := [[\mathbf{y}_f]_{P_m+B_c}, \dots, [\mathbf{y}_f]_{P_m+D-B_c}]^T, \quad (4.17)$$

where B_c is a parameter that controls the amount of observation samples used for channel estimation. Here, \mathcal{O}_m denotes the set that collects the positions of the m th observation sample cluster:

$$\mathcal{O}_m := \{P_m + B_c, \dots, P_m + D - B_c\}. \quad (4.18)$$

To explain the function of the parameter B_c , we understand that if \mathbf{D}_q were strictly banded with only $2B_c + 1$ non-zero diagonals, $\mathbf{y}_f^{\{\mathcal{O}_m\}}$ would be the vector of maximal

length that exclusively depends on the pilot cluster \mathbf{t}_m . In this sense, B_c could be interpreted as the fictive bandwidth of \mathbf{D}_q , as suggested in Fig. 4.2. However, we must be cautious with this interpretation, because \mathbf{D}_q is not really banded for most BEMs. As a matter of fact, B_c just provides a handle on the amount of interference that we want to take into account. More importantly, B_c is not confined to positive values as we will see later on, in which case a negative B_c cannot account for the bandwidth physical interpretation any more. It has only meaning for (4.17).

To formulate the above discussion in mathematical expressions with notations indicated in Fig. 4.2, we obtain

$$\mathbf{y}_f^{\{\mathcal{O}_m\}} = \sum_{q=0}^Q \mathbf{D}_q^{\{\mathcal{O}_m, \mathcal{T}\}} \mathbf{\Delta}_q^{\{\mathcal{T}, \mathcal{T}\}} \mathbf{s}^{\{\mathcal{T}\}} + \underbrace{\sum_{q=0}^Q \mathbf{D}_q^{\{\mathcal{O}_m, \mathcal{D}\}} \mathbf{\Delta}_q^{\{\mathcal{D}, \mathcal{D}\}} \mathbf{s}^{\{\mathcal{D}\}}}_{\mathbf{d}_m} + \mathbf{v}_f^{\{\mathcal{O}_m\}}, \quad (4.19)$$

where we use the notation $\mathbf{A}^{\{\mathcal{S}_r, \mathcal{S}_c\}}$ to denote the intersection of the rows and columns of the matrix \mathbf{A} , where the row indices are collected in the set \mathcal{S}_r and the column indices are collected in the set \mathcal{S}_c . Conform such a definition, we understand that $\mathbf{D}_q^{\{\mathcal{O}_m, \mathcal{T}\}}$ is a $(D - 2B_c + 1) \times M(D + 1)$ matrix, representing the hatched parts of \mathbf{D}_q in Fig. 4.2. Likewise, $\mathbf{\Delta}_q^{\{\mathcal{T}, \mathcal{T}\}}$ is an $M(D + 1) \times M(D + 1)$ diagonal matrix, which is carved out of $\mathbf{\Delta}_q$ corresponding to the pilot-carrying subcarriers; $\mathbf{D}_q^{\{\mathcal{O}_m, \mathcal{D}\}}$ is a $(D - 2B_c + 1) \times (N - M(D + 1))$ matrix, representing the shaded parts of \mathbf{D}_q in Fig. 4.2; $\mathbf{\Delta}_q^{\{\mathcal{D}, \mathcal{D}\}}$ is an $(N - M(D + 1)) \times (N - M(D + 1))$ diagonal matrix, which is carved out of $\mathbf{\Delta}_q$ corresponding to the data-carrying subcarriers; finally, $\mathbf{v}_f^{\{\mathcal{O}_m\}}$ stands for the noise related to $\mathbf{y}_f^{\{\mathcal{O}_m\}}$.

In (4.19), we have thus uncoupled the effect of the data from the pilots, and put it in a separate term \mathbf{d}_m . This term, which poses a nuisance to channel estimation, is in general not zero since \mathbf{D}_q is not strictly banded. Let us rewrite (4.19) as a function of the BEM coefficients \mathbf{c} :

$$\mathbf{y}_f^{\{\mathcal{O}_m\}} = \mathbf{D}^{\{\mathcal{O}_m, \mathcal{T}\}} \mathbf{U} \mathbf{c} + \mathbf{d}_m + \mathbf{v}_f^{\{\mathcal{O}_m\}}, \quad (4.20)$$

with

$$\begin{aligned} \mathbf{D}^{\{\mathcal{O}_m, \mathcal{T}\}} &:= [\mathbf{D}_0^{\{\mathcal{O}_m, \mathcal{T}\}}, \dots, \mathbf{D}_Q^{\{\mathcal{O}_m, \mathcal{T}\}}], \\ \mathbf{U} &:= \mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{T}\}}\} \mathbf{G}_L^{\{\mathcal{T}\}}. \end{aligned} \quad (4.21)$$

Here, $\mathbf{G}_L^{\{\mathcal{T}\}}$ collects the rows of \mathbf{G}_L corresponding to the positions of the pilots. Further, we want to underline that the interference term \mathbf{d}_m carries also channel information \mathbf{c} as can be seen from

$$\mathbf{d}_m = \mathbf{D}^{\{\mathcal{O}_m, \mathcal{D}\}} (\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\}) \mathbf{G}_L^{\{\mathcal{D}\}} \mathbf{c}, \quad (4.22)$$

where

$$\mathbf{D}^{\{\mathcal{O}_m, \mathcal{D}\}} := [\mathbf{D}_0^{\{\mathcal{O}_m, \mathcal{D}\}}, \dots, \mathbf{D}_Q^{\{\mathcal{O}_m, \mathcal{D}\}}], \quad (4.23)$$

and $\mathbf{G}_L^{\{\mathcal{D}\}}$ collects the rows of \mathbf{G}_L corresponding to the positions of the data symbols.

We repeat the above operations for all the observation vectors and stack the results together in a vector $\mathbf{y}_f^{\{\mathcal{O}\}}$, where the symbol \mathcal{O} in the superscript denotes the set that collects the positions of all the observation samples

$$\mathcal{O} := \mathcal{O}_0 \cup \dots \cup \mathcal{O}_{M-1}. \quad (4.24)$$

Accordingly, we have

$$\mathbf{y}_f^{\{\mathcal{O}\}} := [\mathbf{y}_f^{\{\mathcal{O}_0\}T}, \dots, \mathbf{y}_f^{\{\mathcal{O}_{M-1}\}T}T]. \quad (4.25)$$

By introducing further the notations

$$\begin{aligned} \mathbf{d} &:= [\mathbf{d}_0^T, \dots, \mathbf{d}_{M-1}^T]^T, \\ \mathbf{v}_f^{\{\mathcal{O}\}} &:= [\mathbf{v}_f^{\{\mathcal{O}_0\}T}, \dots, \mathbf{v}_f^{\{\mathcal{O}_{M-1}\}T}T], \end{aligned}$$

we can express $\mathbf{y}_f^{\{\mathcal{O}\}}$ as :

$$\mathbf{y}_f^{\{\mathcal{O}\}} = \mathcal{P}\mathbf{c} + \mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}}, \quad (4.26)$$

where

$$\mathcal{P} := \mathbf{Z}\mathbf{U}, \quad (4.27)$$

with

$$\begin{aligned} \mathbf{Z} &= [\mathbf{D}^{\{\mathcal{O}_0, \mathcal{T}\}T}, \dots, \mathbf{D}^{\{\mathcal{O}_{M-1}, \mathcal{T}\}T}T \\ &= \begin{bmatrix} \mathbf{D}_0^{\{\mathcal{O}_0, \mathcal{T}\}} & \dots & \mathbf{D}_Q^{\{\mathcal{O}_0, \mathcal{T}\}} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_0^{\{\mathcal{O}_{M-1}, \mathcal{T}\}} & \dots & \mathbf{D}_Q^{\{\mathcal{O}_{M-1}, \mathcal{T}\}} \end{bmatrix} \end{aligned}$$

and \mathbf{U} defined in (4.21). Likewise, we can express the interference term \mathbf{d} as

$$\mathbf{d} = \mathbf{\Gamma}(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\}\mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c}, \quad (4.28)$$

with

$$\begin{aligned} \mathbf{\Gamma} &:= [\mathbf{D}^{\{\mathcal{O}_0, \mathcal{D}\}T}, \dots, \mathbf{D}^{\{\mathcal{O}_{M-1}, \mathcal{D}\}T}T \\ &= \begin{bmatrix} \mathbf{D}_0^{\{\mathcal{O}_0, \mathcal{D}\}} & \dots & \mathbf{D}_Q^{\{\mathcal{O}_0, \mathcal{D}\}} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_0^{\{\mathcal{O}_{M-1}, \mathcal{D}\}} & \dots & \mathbf{D}_Q^{\{\mathcal{O}_{M-1}, \mathcal{D}\}} \end{bmatrix}. \end{aligned}$$

The above interference analysis is not restricted to any specific BEM. However, note that for the (C)CE-BEM \mathbf{D}_q is strictly banded, and thus

$$\mathbf{D}_q^{\{\mathcal{O}_m, \mathcal{D}\}} = \mathbf{0},$$

if $B_c \geq Q/2$. This case is considered in [47].

4.4 Channel Estimation and B_c Optimization

In this section, we will discuss channel estimation based on the data model that has been established in the previous section. We make the following assumptions:

Assumption 4.2. *The time-domain noise prior to windowing is assumed to be zero-mean white with variance σ^2 .*

Note that the same assumption is also adopted in the previous chapter. Further, we require that

Assumption 4.3. *The data symbols $\mathbf{s}^{\{\mathcal{D}\}}$ are assumed to be zero-mean white with variance σ_s^2 , and uncorrelated with the noise \mathbf{v}_f , i.e.,*

$$\mathcal{E}\{\mathbf{s}^{\{\mathcal{D}\}} \mathbf{v}_f^H\} = \mathbf{0}. \quad (4.29)$$

We will propose three channel estimators in this section: the LMMSE estimator relies on the statistics of \mathbf{c} , while the LS estimator and the BLUE treat \mathbf{c} as a deterministic variable. The performance of each channel estimator is sensitive to B_c . This can be understood from (4.26) and (4.28), where the pilot-related \mathbf{Z} is an $M(D - 2B_c + 1) \times (Q + 1)M(D + 1)$ matrix, and the interference-related $\mathbf{\Gamma}$ is an $M(D - 2B_c + 1) \times (Q + 1)(N - M(D + 1))$ matrix. Intuitively, one would reduce the interference term by setting B_c as large as possible. The same idea is adopted in [96] though the authors address the problem from a different point of view. To explain this using the physical interpretation of B_c : a larger B_c corresponds to a more accurate band approximation of \mathbf{D}_q , and thus to a smaller interference. On the other hand, a larger B_c give rise to a “fatter” \mathbf{Z} , which is often detrimental for a linear channel estimator. We will examine the effect of B_c individually for each estimator.

4.4.1 The LMMSE Estimator

The LMMSE estimator treats \mathbf{c} as a stochastic variable. To be more specific, we introduce the following assumption:

Assumption 4.4. The channel vector \mathbf{c} is assumed to be uncorrelated with the noise \mathbf{v}_f and the information symbols $\mathbf{s}^{\{\mathcal{D}\}}$, i.e.,

$$\mathcal{E}\{\mathbf{c}\mathbf{v}_f^H\} = \mathbf{0}, \quad (4.30)$$

and

$$\mathcal{E}\{\mathbf{c}\mathbf{s}^{\{\mathcal{D}\}H}\} = \mathbf{0}. \quad (4.31)$$

We seek a linear filter \mathbf{F} such that the MSE between the estimated BEM coefficients

$$\hat{\mathbf{c}} = \mathbf{F}\mathbf{y}_f^{\{\mathcal{O}\}} \quad (4.32)$$

and the true BEM coefficients \mathbf{c} is minimal. In other words, we solve

$$\mathbf{F}_{\text{LMMSE}} = \arg \min_{\{\mathbf{F}\}} \text{tr} \{ \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f} \{ (\mathbf{F}\mathbf{y}_f^{\{\mathcal{O}\}} - \mathbf{c})(\mathbf{F}\mathbf{y}_f^{\{\mathcal{O}\}} - \mathbf{c})^H \} \}. \quad (4.33)$$

Substituting (4.26) in the above, we obtain

$$\begin{aligned} & \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f} \{ (\mathbf{F}\mathbf{y}_f^{\{\mathcal{O}\}} - \mathbf{c})(\mathbf{F}\mathbf{y}_f^{\{\mathcal{O}\}} - \mathbf{c})^H \} \\ &= \mathbf{F}(\mathcal{P}\mathbf{R}_c\mathcal{P}^H + \mathbf{R}_d + \mathbf{R}_v^{\{\mathcal{O}\}} + 2\Re\{\Gamma\mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}} \{ (\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}\}\mathbf{R}_c\}\mathcal{P}^H)\mathbf{F}^H \\ & - 2\Re\{\mathbf{R}_c\mathcal{P}^H\mathbf{F}^H + \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}} \{ \mathbf{R}_c(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}*}\})\mathbf{G}_L^{\{\mathcal{D}\}H}\}\Gamma^H\mathbf{F}^H\} + \mathbf{R}_c. \end{aligned} \quad (4.34)$$

In the above, we have introduced the covariance matrix notations

$$\begin{aligned} \mathbf{R}_c &:= \mathcal{E}_{\mathbf{c}}\{\mathbf{c}\mathbf{c}^H\}, \\ \mathbf{R}_d &:= \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}}\{\mathbf{d}\mathbf{d}^H\}, \\ \mathbf{R}_v^{\{\mathcal{O}\}} &:= \mathcal{E}_{\mathbf{v}_f^{\{\mathcal{O}\}}}\{\mathbf{v}_f^{\{\mathcal{O}\}}\mathbf{v}_f^{\{\mathcal{O}\}H}\}, \end{aligned} \quad (4.35)$$

whose computations are given in Appendix 4A. Making use of Assumption 4.3 and 4.4, we can simplify (4.34) to

$$\begin{aligned} & \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f} \{ (\mathbf{F}\mathbf{y}_f^{\{\mathcal{O}\}} - \mathbf{c})(\mathbf{F}\mathbf{y}_f^{\{\mathcal{O}\}} - \mathbf{c})^H \} \\ &= \mathbf{F}(\mathcal{P}\mathbf{R}_c\mathcal{P}^H + \mathbf{R}_d + \mathbf{R}_v^{\{\mathcal{O}\}})\mathbf{F}^H - 2\Re(\mathbf{R}_c\mathcal{P}^H\mathbf{F}^H) + \mathbf{R}_c, \end{aligned} \quad (4.36)$$

Taking the derivative on the RHS of (4.36) with respect to \mathbf{F} , and setting the result to zero, we get

$$\mathbf{F}(\mathcal{P}\mathbf{R}_c\mathcal{P}^H + \mathbf{R}_d + \mathbf{R}_v^{\{\mathcal{O}\}}) = \mathbf{R}_c\mathcal{P}^H. \quad (4.37)$$

Assuming that the covariance matrix in the above is non-singular, we can find the LMMSE estimator as

$$\begin{aligned} \mathbf{F}_{\text{LMMSE}} &= \mathbf{R}_c\mathcal{P}^H(\mathcal{P}\mathbf{R}_c\mathcal{P}^H + \mathbf{R}_d + \mathbf{R}_v^{\{\mathcal{O}\}})^{-1} \\ &= \mathbf{R}_c\mathcal{P}^H(\mathcal{P}\mathbf{R}_c\mathcal{P}^H + \mathbf{R}_{\mathcal{I}})^{-1}. \end{aligned} \quad (4.38)$$

with

$$\mathbf{R}_{\mathcal{I}} := \mathbf{R}_d + \mathbf{R}_v^{\{\mathcal{O}\}}. \quad (4.39)$$

Note that although (4.38) bears a similar form to the classical LMMSE estimator [48], it has the extra task to process the interference term \mathbf{d} , which contains the information of \mathbf{c} itself. For this purpose, the proposed LMMSE treats \mathbf{d} as a random vector resorts to the assumed statistics of \mathbf{c} and $\mathbf{s}^{\{\mathcal{D}\}}$. However, the statistics of \mathbf{c} are difficult to retrieve in practice and not always reliable. For instance, the Doppler spread could only be roughly known or the assumed Doppler spectrum deviates from the true value. In such cases, the proposed LMMSE estimator is sub-optimal. We will show some examples in Section 4.5.

We want to use $\hat{\mathbf{c}}$ to reconstruct the BEM channel, and examine how close it is to the best BEM fit of the real channel. To this end, we adopt the MSE criterion as

$$\begin{aligned} \text{MSE} &:= \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \{ \|\mathbf{B} \otimes \mathbf{I}_{L+1} \hat{\mathbf{c}} - (\mathbf{B} \otimes \mathbf{I}_{L+1}) \mathbf{c}\|^2 \} \\ &= \text{tr} \{ (\mathbf{B} \otimes \mathbf{I}_{L+1}) \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \{ (\hat{\mathbf{c}} - \mathbf{c})(\hat{\mathbf{c}} - \mathbf{c})^H \} (\mathbf{B} \otimes \mathbf{I}_{L+1})^H \} \\ &= \text{tr} \{ \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \{ (\hat{\mathbf{c}} - \mathbf{c})(\hat{\mathbf{c}} - \mathbf{c})^H \} \}, \end{aligned} \quad (4.40)$$

where $\text{tr}\{\mathbf{A}\}$ stands for the trace of the matrix \mathbf{A} . The last equality above holds since we have designed \mathbf{B} to have orthonormal columns [c.f.(4.10)]. (4.40) suggests that the channel fitting MSE equals the MSE of the estimated BEM coefficients. Therefore, the MSE resulting from the LMMSE channel estimator can be expressed as:

$$\text{MSE}_{\text{LMMSE}} = \text{tr} \{ (\mathcal{P}^H \mathbf{R}_{\mathcal{I}}^{-1} \mathcal{P} + \mathbf{R}_c^{-1})^{-1} \}. \quad (4.41)$$

$\text{MSE}_{\text{LMMSE}}$ is a function of the chosen bandwidth B_c because it determines the size and content of \mathcal{P} and $\mathbf{R}_{\mathcal{I}}$. Later on, we will show how to find the optimal B_c to minimize $\text{MSE}_{\text{LMMSE}}$.

4.4.2 The Least Squares Estimator

The Least Squares (LS) estimator \mathbf{F}_{LS} treats \mathbf{c} as a deterministic variable. It is straightforward to obtain

$$\mathbf{F}_{\text{LS}} := \mathcal{P}^\dagger \quad (4.42)$$

such that

$$\hat{\mathbf{c}}_{\text{LS}} = \mathbf{c} + \mathcal{P}^\dagger (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}}). \quad (4.43)$$

The LS estimator is the most robust estimator, requiring no knowledge about the channel and noise statistics. However, it performs inferior when the interference is prominent: we will show in the simulation part that the LS estimator suffers from a

large performance gap in comparison with the Cramer-Rao Bound (CRB) (derived in Appendix 4B). In addition, the performance of the LS estimator relies heavily on the condition number of \mathcal{P} as we can see from the resulting MSE

$$\begin{aligned} \text{MSE}_{\text{LS}} &:= \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \{ \text{tr} \{ \mathcal{P}^\dagger (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}}) (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}})^H \mathcal{P}^\dagger H \} \} \\ &= \text{tr} \{ \mathcal{P}^\dagger \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \{ (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}}) (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}})^H \} \mathcal{P}^\dagger H \} \\ &= \text{tr} \{ \mathcal{P}^\dagger \mathbf{R}_{\mathcal{I}} \mathcal{P}^\dagger H \}, \end{aligned} \quad (4.44)$$

which is again a function of B_c .

4.4.3 An Iterative BLUE

From (4.26), we can find an expression for the BLUE following similar steps as in [48, Appendix 6B] by treating the interference \mathbf{d} and noise $\mathbf{v}_f^{\{\mathcal{O}\}}$ as a single disturbance term such that

$$\begin{aligned} \hat{\mathbf{c}}_{\text{BLUE}} &= \mathbf{F}_{\text{BLUE}} \mathbf{y}_f^{\{\mathcal{O}\}}, \\ \mathbf{F}_{\text{BLUE}} &= (\mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{c}) \mathcal{P})^{-1} \mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{c}), \end{aligned} \quad (4.45)$$

where $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$ denotes the covariance matrix of the disturbance. Here, \mathbf{c} is again viewed as a deterministic variable and therefore

$$\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c}) := \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \{ (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}}) (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}})^H \}. \quad (4.46)$$

Due to Assumption 4.3, we have

$$\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c}) = \tilde{\mathbf{R}}_d(\mathbf{c}) + \mathbf{R}_v^{\{\mathcal{O}\}}, \quad (4.47)$$

with

$$\tilde{\mathbf{R}}_d(\mathbf{c}) := \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}} \{ \mathbf{d} \mathbf{d}^H \}, \quad (4.48)$$

whose derivation can be found in Appendix 4A.

However, (4.45) is not implementable in closed form since its computation entails the knowledge of \mathbf{c} itself. A recursive approach can therefore be applied. Suppose at the k th iteration, an estimate for \mathbf{c} is available denoted as $\hat{\mathbf{c}}_{\text{BLUE}}^{(k)}$. Next, we use this estimate to update the covariance matrix $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$, which in turn is used to produce the BLUE for the next iteration and so on:

$$\begin{aligned} \mathbf{F}_{\text{BLUE}}^{(k+1)} &= (\mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\hat{\mathbf{c}}_{\text{BLUE}}^{(k)}) \mathcal{P})^{-1} \mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\hat{\mathbf{c}}_{\text{BLUE}}^{(k)}), \\ \hat{\mathbf{c}}_{\text{BLUE}}^{(k+1)} &= \mathbf{F}_{\text{BLUE}}^{(k+1)} \mathbf{y}_f^{\{\mathcal{O}\}}. \end{aligned}$$

Note that a similar idea is adopted in [33] though applied in a different context. To initialize, we can set

$$\hat{\mathbf{c}}_{\text{BLUE}}^{(0)} = \mathbf{0}, \quad (4.49)$$

which results in the following expression for the first iteration:

$$\mathbf{F}_{\text{BLUE}}^{(1)} = (\mathcal{P}^H (\mathbf{R}_v^{\{\mathcal{O}\}})^{-1} \mathcal{P})^{-1} \mathcal{P}^H (\mathbf{R}_v^{\{\mathcal{O}\}})^{-1}. \quad (4.50)$$

Conform Assumption 4.2, the above expression is actually the maximum likelihood estimator (MLE) [48] that is obtained by ignoring the interference \mathbf{d} . The resulting

$$\hat{\mathbf{c}}_{\text{BLUE}}^{(1)} = \mathbf{F}_{\text{BLUE}}^{(1)} \mathbf{y}_f^{\{\mathcal{O}\}} \quad (4.51)$$

is the least squares fit as obtained in the previous section but weighted by the noise covariance. If \mathbf{d} is small, which is often the case by carefully selecting B_c , $\hat{\mathbf{c}}_{\text{BLUE}}^{(1)}$ is already very close to \mathbf{c} .

Assuming that $\hat{\mathbf{c}}_{\text{BLUE}}^{(k)} \rightarrow \hat{\mathbf{c}}_{\text{BLUE}}$, we use (4.45) to find the MSE of the channel estimator

$$\begin{aligned} \text{MSE}_{\text{BLUE}} &= \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \{ \text{tr} \{ \mathbf{F}_{\text{BLUE}} (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}}) (\mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}})^H \mathbf{F}_{\text{BLUE}}^H \} \} \\ &= \mathcal{E}_{\mathbf{c}} \{ \text{tr} (\mathbf{F}_{\text{BLUE}} \tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c}) \mathbf{F}_{\text{BLUE}}^H) \} \\ &= \mathcal{E}_{\mathbf{c}} \{ \text{tr} ((\mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{c}) \mathcal{P})^{-1}) \}. \end{aligned} \quad (4.52)$$

The above expression provides a lower-bound on the performance of the iterative BLUE. This MSE is, however, difficult to evaluate in closed form due to the inversion of $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$, which forces us to resort to the Monte Carlo method. As will be evident later on, the MSE resulting from the BLUE depends also on the choice of B_c .

The LS estimator and BLUE produce both unbiased estimates of \mathbf{c} . Their performances are hence lower bounded by the Cramer-Rao bound (CRB), which is derived in Appendix 4B. We will compare them in the simulation part.

4.4.4 Optimization of B_c

In this section, we will optimize the number of observation samples used for channel estimation, or in other words seek the optimal B_c that will minimize the estimator variance given in (4.41), (4.44) and (4.52) for the LMMSE, LS estimator and BLUE, respectively:

$$B_c = \arg \min_{\{B_c\}} \text{MSE}. \quad (4.53)$$

It is difficult to find a closed-form solution for the above equation, especially for the BLUE. An alternative is to evaluate (4.53) exhaustively, which is feasible since

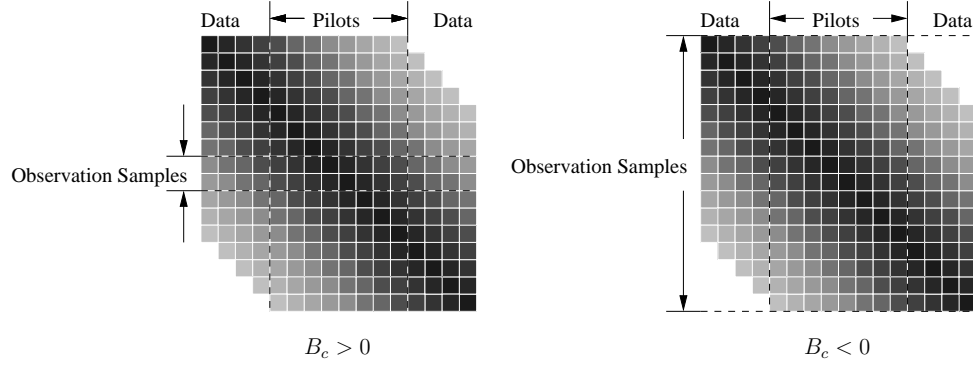


Figure 4.3: The FD channel matrix \mathbf{H}_f , whose power degradation is represented by the tints. The rows of \mathbf{H}_f corresponds to the positions of the observation samples and its columns to the positions of the data and pilots. Left plot $B_c > 0$. Right plot $B_c < 0$.

there is only a limited range of possible values for B_c . First of all, the number of observation samples $M(D - 2B_c + 1)$ cannot exceed the number of subcarriers N . This leads to

$$\frac{D+1}{2} - \frac{N}{2M} \leq B_c. \quad (4.54)$$

Second, we desire that the $M(D - 2B_c + 1) \times (L+1)(Q+1)$ matrix \mathcal{P} be of full rank, which is essential for the channel estimators to have a good performance in the absence of interference and noise. We will discuss this topic in detail in the next chapter. Here, we suffice with a necessary condition for \mathcal{P} to have full column-rank: it should be tall or at least square, which yields

$$\frac{D+1}{2} - \frac{(L+1)(Q+1)}{2M} \geq B_c. \quad (4.55)$$

Combining (4.54) and (4.55), we obtain the possible range of B_c :

$$\frac{D+1}{2} - \frac{N}{2M} \leq B_c \leq \frac{D+1}{2} - \frac{(L+1)(Q+1)}{2M}, \quad (4.56)$$

Fortunately, even the exhaustive search might be avoided as will become evident from the simulations, where the MSE-versus- B_c curves for each channel estimator exhibit a monotonous track: the LS channel estimator yields the best performance when B_c is maximized, i.e.,

$$B_c = \frac{D+1}{2} - \frac{(L+1)(Q+1)}{2M},$$

whereas the LMMSE estimator and BLUE perform best when B_c is minimized, i.e.,

$$B_c = \frac{D+1}{2} - \frac{N}{2M}.$$

To explain these, we recall that a larger B_c leads to a more accurate band assumption of the channel matrix and hence to a smaller out-of-band interference. This effect is depicted in the left plot of Fig. 4.3, where the FD channel matrix \mathbf{H}_f is drawn. It can be seen that for a larger B_c , the number of observation samples is typically smaller than the number of pilots. A smaller out-of-band interference is beneficial to the LS estimator, which is not good at suppressing it due to the lack of statistical knowledge.

Opposed to the LS estimator, the LMMSE estimator and BLUE require the B_c to be as small as possible. For practical setups, this often implies a negative B_c , in which case the observation samples outnumber the pilots as illustrated in the right plot of Fig. 4.3. We can see that some of the observation samples, e.g., in the two boundary areas, will suffer a very low signal to interference and noise ratio (SINR) because for these observation samples the unknown data are magnified by the high-power diagonals of the channel matrix and are hence much more prominent than the pilots. However, this casts no serious problems to the LMMSE estimator and BLUE since both of them can take the interference into account in a positive way. We will come back to this issue in the next section.

4.5 Simulations

We build up an OFDM system with $N = 256$ subcarriers, where roughly 80% of the subcarriers are used for transmitting data symbols. The remaining subcarriers are reserved for pilots, which are grouped in $M = 6$ equidistant clusters, each containing $D + 1 = 9$ pilot tones. Inside each cluster, we adopt a scheme referred to as ‘‘Frequency-Domain Kronecker Delta’’ (FDKD) in [47], where a non-zero pilot is located in the middle of the cluster with zero guard bands on both sides.

The TV channel following the Jakes’ Doppler spectrum [45] can be generated following the methods in [124]. To be able to approximate the TV channel by a BEM, we set the BEM scale to $Q = 4$. Further, we assume the channel to be an FIR filter with $L + 1 = 6$ taps, which are independent random variables with an exponential power intensity profile. More specifically, we set the variance of the l th channel tap to be $\sigma_l^2 = c \cdot e^{-l/10}$, where c denotes a normalization constant. In short, we can characterize the TV channel with $(L + 1)(Q + 1) = 30$ BEM coefficients.

Test Case 1. The (Windowed) BEM Justification.

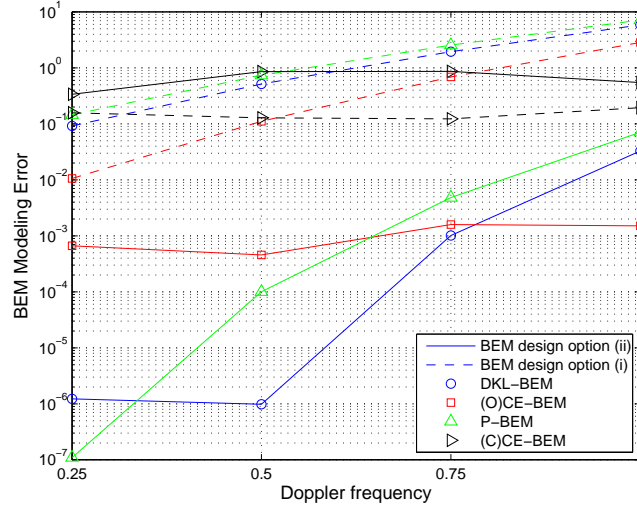


Figure 4.4: Justification of (windowed) BEM

We first study the modeling performance (e.g., the channel fitting performance in the absence of noise) of the DKL-BEM, (C)CE-BEM, (O)CE-BEM, and P-BEM for a range of normalized Doppler spreads ν_D . The DKL-BEM is constructed based on the Jakes' Doppler spectrum assumption but fixed at $\nu_D = 0.002$, which is thus sub-optimal for other ν_D 's. The BEM-modeled channel is compared to the true channel after windowing in terms of the modeling error $\mathcal{E}_{h_t} \{\|\epsilon\|^2\}$. For the window design, we adopt the window presented in the previous chapter, which is a sum of 3 complex exponentials (Proposition 3.1). In Section 4.2.2, we have argued about two BEM designs in the presence of the windowing: the first BEM design follows option (i) in (4.10), which is the traditional BEM design ignoring the windowing; the second BEM design follows option (ii) in (4.10), and is adapted to the windowing. From the results that are sketched in Fig. 4.4, we can observe that by taking the windowing into account, the BEMs following option (ii) yield in general a tighter fit with the windowed channel, with the only exception of the (C)CE-BEM, which, by following option (i), performs better within the tested Doppler frequency range. Further, it can be seen that the DKL-BEM and P-BEM have the smallest modeling error at low Doppler frequencies but lose track if the channel varies faster. Apparently for the DKL-BEM, the mismatch due to an underestimated Doppler frequency is much more harmful than the mismatch due to an overestimated Doppler frequency. The

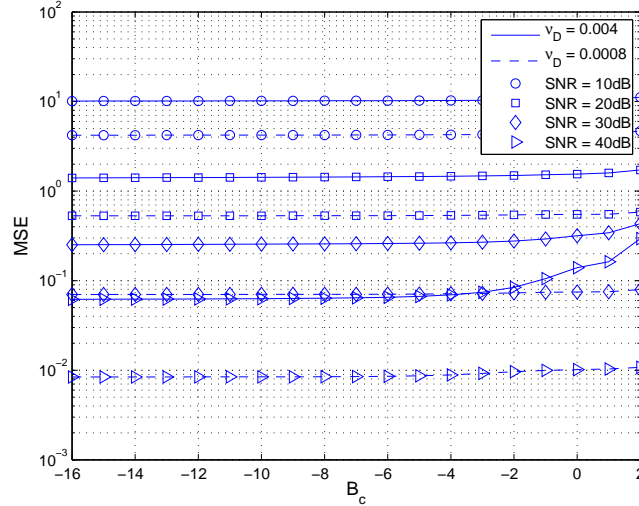


Figure 4.5: MSE vs. B_c for the LMMSE estimator. Solid curves $\nu_D = 0.004$; Dashed curves $\nu_D = 0.0008$.

(O)CE-BEM that is virtually independent of the Doppler frequency is more robust in this sense.

For the following simulations, we will concentrate on TV channels at two Doppler spreads: (1) $\nu_D = 0.0008$, (2) $\nu_D = 0.004$. We will use the DKL-BEM when constructing the LMMSE estimator because both of them rely on the knowledge of channel statistics. For both the DKL-BEM and the LMMSE estimator, we will allow for a statistical mismatch by assuming a fixed $\nu_D = 0.002$ for all the Doppler spreads under test. For the LS estimator and the BLUE, we will just use the (O)CE-BEM since both the channel estimators and the BEM are independent of the channel statistics. In addition, we will also compare our results with the channel estimation method for the (C)CE-BEM presented in [47]. Note that this method resembles our proposed LMMSE estimator (without mismatch) but uses a data model that is only applicable to the (C)CE-BEM, i.e., the channel matrix is viewed as strictly banded.

Test Case 2. Seeking the optimal B_c .

First, we need to find an optimal B_c for the different channel estimators. From (4.56), B_c is bounded by

$$-16 \leq B_c \leq 2. \quad (4.57)$$

For these values, we evaluate the MSE of the LMMSE estimator (4.41) for SNR = 10, 20, 30, 40 dB, and depict the results in Fig. 4.5. We observe that at low SNR, the

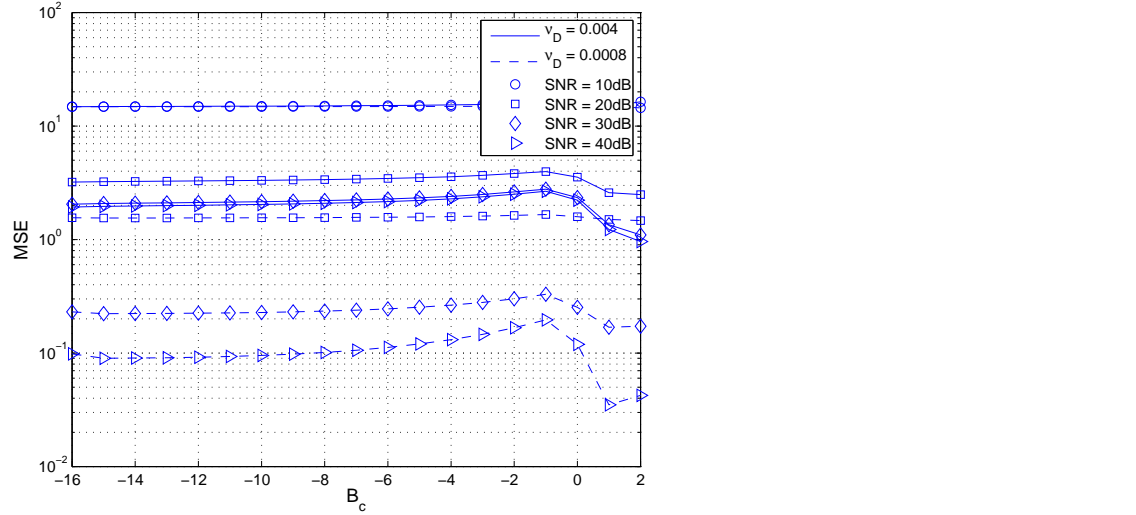


Figure 4.6: MSE vs. B_c for the LS estimator. Solid curves $\nu_D = 0.004$; Dashed curves $\nu_D = 0.0008$.

effect of B_c is not pronounced, whereas at high SNR, a smaller B_c corresponds to a smaller MSE. This is especially true for $\nu_D = 0.004$, where the ICI is still severe in spite of windowing. Tuning B_c is of greater importance in that case. Therefore, we choose $B_c = -16$ as the optimal value, which implies that the whole OFDM symbol will be invoked for channel estimation.

The results for the LS estimator are plotted in Fig. 4.6, where we observe that B_c must be chosen as large as possible, i.e., $B_c = 2$.

For the BLUE in Fig. 4.7, a smaller B_c always yields a lower MSE just like the LMMSE estimator and we should also take $B_c = -16$. However, complexity plays a crucial role in this case, because the BLUE has to be computed recursively and the procedure must be repeated for every OFDM symbol (note that the LMMSE estimator is in essence time-invariant and can thus be precomputed and stored off-line). In practice, a smaller B_c often requires more iterations to reach convergence, and during each iteration, it inflicts a larger computational effort because more observation samples have to be processed. Observing that the MSE curve descends only slowly for $B_c < -3$, we select $B_c = -3$ as a good compromise between complexity and performance for the BLUE.

Test Case 3. The estimator performance.

Having determined $B_c = -16$ for the LMMSE estimator, $B_c = 2$ for the LS esti-

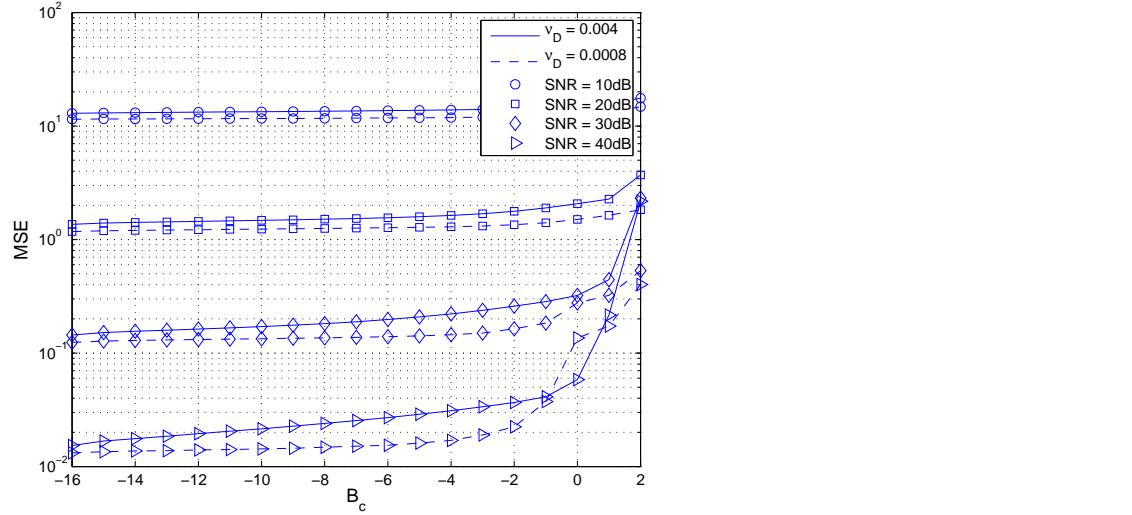


Figure 4.7: MSE vs. B_c for the BLUE. Solid curves $\nu_D = 0.004$; Dashed curves $\nu_D = 0.0008$.

mator, and $B_c = -3$ for the BLUE, we inspect their channel estimation performance for a wide range of SNRs. Next to the MSE defined in (4.40), which we will refer to as the “BEM MSE”, we will also look at the so-called “channel MSE” which we define by

$$\text{MSE-CH} := \mathcal{E}_{\mathbf{h}_t} \{ \|\mathbf{h}_t - (\mathbf{B} \otimes \mathbf{I}_{L+1})\hat{\mathbf{c}}\|^2 \}. \quad (4.58)$$

Note that the channel MSE differs from the BEM MSE in that it explicitly takes the BEM modeling error into account, whereas the BEM MSE merely indicates how close the estimated channel is to the best possible BEM fit. Fig. 4.8 depicts the performance in terms of the BEM MSE, whereas Fig. 4.9 depicts the performance in terms of the channel MSE. We observe that these two performances are in general very close to each other, which suggests that Assumption 4.1 brings no harm to channel estimation. Further, we remark that the LMMSE estimator, which is sub-optimal due to the Doppler frequency mismatch, performs much better under TV channels with $\nu_D = 0.0008$ than with $\nu_D = 0.004$. This suggests that underestimating the Doppler frequency is more harmful than overestimating it.

The BEM MSE is also compared with the CRB (see the derivation in Appendix 4B) in Fig. 4.8. The CRB is based on $B_c = -16$ and obtained using the Monte Carlo method, thereby exploring the channel statistics. We observe that the performance of the BLUE is very close to the CRB.

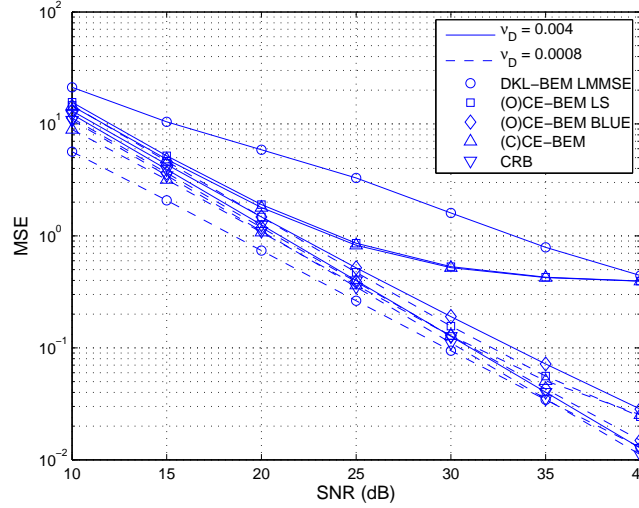


Figure 4.8: MSE of the BEM vs. SNR. Solid curves $\nu_D = 0.004$; Dashed curves $\nu_D = 0.0008$.

Test Case 4. Equalization performance based on the estimated channel

For this test case, we will transmit QPSK modulated data symbols, and use the low-complexity block LMMSE equalizer that is presented in the previous chapter. For the sake of complexity, the FD channel matrix will be approximated as a banded matrix with nine non-zero diagonals*.

To examine the influence of the channel estimation error on the equalization, we construct the equalizer utilizing the estimated channel obtained from our LMMSE estimator, LS estimator and BLUE, respectively. As a comparison, we also list the equalization performances, which are based on the estimated (C)CE-BEM channel and the perfect CSI. It can be seen that due to the Doppler diversity the equalizer renders a better performance for faster TV channels, but it is plagued by a higher BER floor. Comparing the results in Fig. 4.9 with those in Fig. 4.10, one can remark that the equalization performance is consistent with the corresponding channel estimation performance for each Doppler frequency. Take the DKL-BEM and the (C)CE-BEM at $\nu_D = 0.004$ for instance, the channel estimation for the (C)CE-BEM is better at low SNR but worse at high SNR than that for the DKL-BEM. A corresponding

*In the previous chapter, we used Q to represent the number of diagonals we take into account in the equalization design. However, that must not be confused with the BEM scale Q here, which is only meant for channel estimation

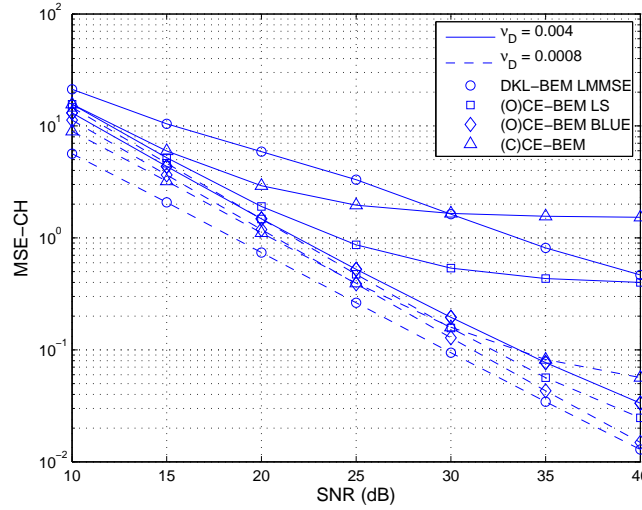


Figure 4.9: Channel MSE vs. SNR. Solid curves $\nu_D = 0.004$; Dashed curves $\nu_D = 0.0008$.

observation can be made from Fig. 4.10.

4.6 Summary

In this chapter, we discussed how to estimate the TV channel in the OFDM system with aid of pilots. The pilots are interleaved with data in the frequency domain. The TV channel is approximated by an arbitrary BEM, which must be very tight with respect to the true channel. Channel estimation is achieved by estimating the BEM coefficients.

Due to the Doppler spread, the channel matrix in the frequency domain is not diagonal but full. This means that even if it is approximately banded, we cannot find any observation sample that is only contributed by the pilots and free from interference. Note that by employing more observation samples for channel estimation, we can benefit from more pilot information to suppress the noise, but at the same time we have to deal with more interference.

Three channel estimators were discussed, the LMMSE estimator, the LS estimator and the BLUE. The LMMSE estimator and the BLUE can both explicitly take the interference into account. The former requires more statistical knowledge than the latter, but can be attained in one shot. Simulation results show that for the LMMSE

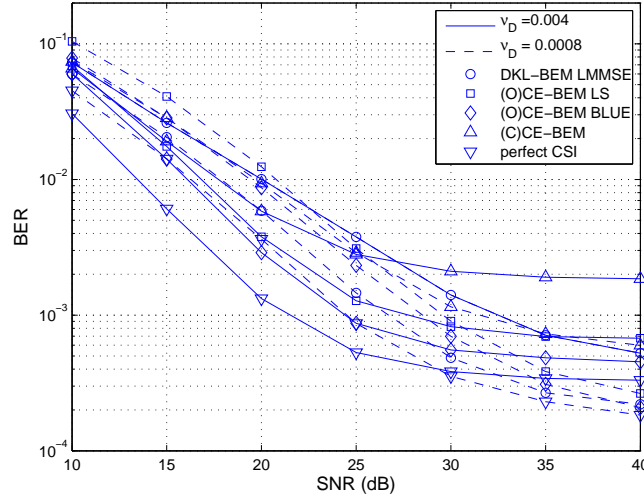


Figure 4.10: BER vs. SNR. Solid curves $\nu_D = 0.004$; Dashed curves $\nu_D = 0.0008$.

estimator and the BLUE, the whole OFDM symbol must be employed to achieve the optimum performance, while for the LS estimator, the optimum performance is achieved when the number of observation samples is kept to the minimum. These effects can be accounted for by the amounts of information and interference that are cast to the channel estimator, and the different capabilities of the channel estimators to suppress the interference.

Channel estimation is carried out in this chapter within a single OFDM symbol. Two questions arise: first, how many pilots are needed to make the channel identifiable; second, with the same amount of pilots (bandwidth efficiency), can we further improve the channel estimation performance? These two questions will be answered in the next chapter, where the channel estimation will be based on multiple OFDM symbols.

Appendix 4A: Derivation of Covariance Matrices $\mathbf{R}_n^{\{O\}}$, \mathbf{R}_d and $\tilde{\mathbf{R}}_d(\mathbf{c})$

In this section, we will give the computations of the covariance matrices that are used in Section 4.4. Let us start with the covariance matrix of the noise term $\mathbf{v}_f^{\{O\}}$,

i.e.,

$$\mathbf{R}_v^{\{\mathcal{O}\}} := \mathcal{E}_{\mathbf{v}_f} \{ \mathbf{v}_f^{\{\mathcal{O}\}} \mathbf{v}_f^{\{\mathcal{O}\}H} \}. \quad (4.59)$$

It is clear that $\mathbf{R}_v^{\{\mathcal{O}\}}$ is extracted from a larger matrix

$$\mathbf{R}_v = \mathcal{E}_{\mathbf{v}_f} \{ \mathbf{v}_f \mathbf{v}_f^H \}, \quad (4.60)$$

which can be easily obtained from Assumption 4.2 by taking the windowing and demodulation into account:

$$\mathbf{R}_v = \sigma^2 \mathbf{W}_N \mathcal{D} \{ \mathbf{w} \} \mathcal{D} \{ \mathbf{w}^H \} \mathbf{W}_N^H. \quad (4.61)$$

Hence, $\mathbf{R}_v^{\{\mathcal{O}\}}$ is comprised of the rows and columns of \mathbf{R}_v , corresponding to the positions of $\mathbf{y}_f^{\{\mathcal{O}\}}$.

Since the interference term \mathbf{d} depends on the BEM coefficients \mathbf{c} , we can make a distinction between two types of covariance matrices. The first one is based on a stochastic channel assumption, whereas the other one is based on a deterministic channel assumption, which leads to \mathbf{R}_d and $\tilde{\mathbf{R}}_d(\mathbf{c})$, respectively. Note that for both cases the data symbols $\mathbf{s}^{\{\mathcal{D}\}}$ are always viewed as a stochastic variable.

We first assume that \mathbf{c} is stochastic with

$$\mathbf{R}_c = \mathcal{E}_c \{ \mathbf{c} \mathbf{c}^H \}. \quad (4.62)$$

To derive its expression, let us introduce the $(Q+1) \times (Q+1)$ matrix \mathbf{R}_{c_l} , which is defined as the autocorrelation of the BEM coefficients for tap l :

$$\mathbf{R}_{c_l} := \mathcal{E} \{ \mathbf{c}_l \mathbf{c}_l^H \}. \quad (4.63)$$

with

$$\mathbf{c}_l := [c_{0,l}, \dots, c_{Q,l}]^T. \quad (4.64)$$

By taking the windowing into account, we can easily derive that

$$\mathbf{R}_{c_l} = \mathbf{B}^\dagger \mathcal{D} \{ \mathbf{w} \} \mathbf{R}_{h_l} \mathcal{D} \{ \mathbf{w}^H \} \mathbf{B}^{\dagger H}, \quad (4.65)$$

where

$$\mathbf{R}_{h_l} := \mathcal{E}_{\mathbf{h}_l} \{ \mathbf{h}_l \mathbf{h}_l^H \} \quad (4.66)$$

is defined as the covariance matrix of the l th unwindowed channel tap, which is thus related to the specific Doppler spectrum. In addition, if we use an $(L+1) \times (L+1)$ matrix $\mathbf{R}_{\text{multipath}}$ to describe the correlation due to the channel's frequency selectivity, which depends on the correlation between the channel taps and the power intensity profile, i.e.,

$$[\mathbf{R}_{\text{multipath}}]_{p,q} := \mathcal{E} \{ h_{n,p} h_{n,q}^* \}, \quad (4.67)$$

we obtain

$$\mathbf{R}_c = \mathbf{R}_{c_l} \otimes \mathbf{R}_{\text{multipath}} / \sigma_l^2. \quad (4.68)$$

with σ_l^2 denoting the power of the l th channel tap defined in (2.17).

With aid of \mathbf{R}_c , we can derive the covariance matrix of the interference term \mathbf{d} as

$$\begin{aligned} \mathbf{R}_d &:= \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}}\{\mathbf{d}\mathbf{d}^H\} \\ &= \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}}\{\boldsymbol{\Gamma}(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}\}\mathbf{c}\mathbf{c}^H(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}{}^H\boldsymbol{\Gamma}^H\} \\ &= \boldsymbol{\Gamma}\mathbf{R}_x\boldsymbol{\Gamma}^H, \end{aligned} \quad (4.69)$$

where using Assumption 4.4 we have

$$\begin{aligned} \mathbf{R}_x &:= \mathcal{E}_{\mathbf{c}, \mathbf{s}^{\{\mathcal{D}\}}}\{(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}\}\mathbf{c}\mathbf{c}^H(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}{}^H\} \\ &= \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}}\{(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}\}\mathbf{R}_c(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}{}^H\} \\ &= \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}}\{(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\boldsymbol{\mathcal{X}}(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})^H\}, \end{aligned}$$

with

$$\boldsymbol{\mathcal{X}} := (\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{R}_c(\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})^H. \quad (4.70)$$

Utilizing Assumption 4.3, we can easily verify that

$$[\mathbf{R}_x]_{m,n} = \begin{cases} \sigma_s^2[\boldsymbol{\mathcal{X}}]_{m,n} & \text{if } \text{mod}(m-n, N_d) = 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $N_d = N - M(D+1)$ is the total number of data symbols in $\mathbf{s}^{\{\mathcal{D}\}}$.

In contrast with \mathbf{R}_d , $\tilde{\mathbf{R}}_d(\mathbf{c})$ is obtained by assuming that \mathbf{c} is deterministic:

$$\begin{aligned} \tilde{\mathbf{R}}_d(\mathbf{c}) &:= \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}}\{\mathbf{d}\mathbf{d}^H\} \\ &= \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}}\{\boldsymbol{\Gamma}(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}\}\mathbf{c}\mathbf{c}^H(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}{}^H\boldsymbol{\Gamma}^H\}, \\ &= \boldsymbol{\Gamma}\tilde{\mathbf{R}}_x(\mathbf{c})\boldsymbol{\Gamma}^H, \end{aligned} \quad (4.71)$$

where using Assumption 4.3 we have

$$\begin{aligned} \tilde{\mathbf{R}}_x(\mathbf{c}) &:= \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}}\{(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}\}\mathbf{c}\mathbf{c}^H(\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{s}^{\{\mathcal{D}\}}\})\mathbf{G}_L^{\{\mathcal{D}\}}{}^H\} \\ &= \mathcal{D}\{(\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c}\}\mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}}\{(\mathbf{1}_{Q+1,1} \otimes \mathbf{s}^{\{\mathcal{D}\}})(\mathbf{1}_{Q+1,1} \otimes \mathbf{s}^{\{\mathcal{D}\}})^H\}\mathcal{D}\{(\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c}\}^H \\ &= \sigma_s^2\mathcal{D}\{(\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c}\}(\mathbf{1}_{Q+1,Q+1} \otimes \mathbf{I}_{N_d})\mathcal{D}\{(\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c}\}^H. \end{aligned} \quad (4.72)$$

Appendix 4B: Cramer-Rao Bound

In this section, we will derive the CRB of the BEM coefficient estimate \mathbf{c} . We start from (4.26)

$$\mathbf{y}_f^{\{\mathcal{O}\}} = \mathcal{P}\mathbf{c} + \mathbf{d} + \mathbf{v}_f^{\{\mathcal{O}\}}. \quad (4.73)$$

For the sake of simplicity, we assume that $\mathbf{y}_f^{\{\mathcal{O}\}}$ is Gaussian distributed with mean $\mathcal{P}\mathbf{c}$ and covariance matrix $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$. The latter is defined in (4.46) as the covariance of the interference and noise taking \mathbf{c} as a deterministic variable and the data and noise as stochastic. The Gaussian distribution assumption of $\mathbf{y}_f^{\{\mathcal{O}\}}$ is supported by Assumption 4.3, and by the fact that the OFDM symbol size N is large enough to make \mathbf{d} approximately normal-distributed due to the central limit theorem. The negative Gaussian log-likelihood function \mathcal{L} can hence be written as

$$-\mathcal{L} = \mathcal{C} \log(\det(\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c}))) + (\mathbf{y}_f^{\{\mathcal{O}\}} - \mathcal{P}\mathbf{c})^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{c})(\mathbf{y}_f^{\{\mathcal{O}\}} - \mathcal{P}\mathbf{c}), \quad (4.74)$$

which leads to the following Fisher Information Matrix (FIM)

$$\mathcal{J}(\mathbf{c}) := \mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f^{\{\mathcal{O}\}}} \left\{ \left(\frac{\partial \mathcal{L}}{\partial \mathbf{c}} \right)^T \left(\frac{\partial \mathcal{L}}{\partial \mathbf{c}} \right) \right\}. \quad (4.75)$$

Adapting the results given in [23], we can formulate the real FIM as

$$\mathcal{J}(\mathbf{c}) = 2 \begin{bmatrix} \Re(\mathbf{J}_{\theta\theta}) & -\Im(\mathbf{J}_{\theta\theta}) \\ \Im(\mathbf{J}_{\theta\theta}) & \Re(\mathbf{J}_{\theta\theta}) \end{bmatrix} + 2 \begin{bmatrix} \Re(\mathbf{J}_{\theta\theta^*}) & -\Im(\mathbf{J}_{\theta\theta^*}) \\ \Im(\mathbf{J}_{\theta\theta^*}) & \Re(\mathbf{J}_{\theta\theta^*}) \end{bmatrix}, \quad (4.76)$$

where

$$\begin{aligned} [\mathbf{J}_{\theta\theta}]_{i,j} &:= [\mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \mathcal{P}]_{i,j} + \text{tr} \left(\tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{c}]_i^*} \tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{c}]_j^*} \right), \\ [\mathbf{J}_{\theta\theta^*}]_{i,j} &:= \text{tr} \left(\tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{c}]_i^*} \tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{c}]_j^*} \right). \end{aligned}$$

Let us now focus on computing $\frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{c}]_j^*}$. Since $[\frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{c}]_j^*}]_{m,n} = \frac{\partial [\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n}}{\partial [\mathbf{c}]_j^*}$, we want to formulate $[\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n}$ as a function of $[\mathbf{c}]_j^*$. This is achieved by realizing that

$$[\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n} = \mathbf{e}_m^T (\Gamma \tilde{\mathbf{R}}_x(\mathbf{c}) \Gamma^H + \mathbf{R}_v^{\{\mathcal{O}\}}) \mathbf{e}_n, \quad (4.77)$$

with $\tilde{\mathbf{R}}_x(\mathbf{c})$ defined in (4.72). Following the derivative rules given in [65], we have

$$\begin{aligned} \frac{\partial [\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n}}{\partial [\mathbf{c}]_j^*} &= \mathbf{e}_m^T \Gamma \frac{\partial \tilde{\mathbf{R}}_x(\mathbf{c})}{\partial [\mathbf{c}]_j^*} \Gamma^H \mathbf{e}_n, \\ \frac{\partial \tilde{\mathbf{R}}_x(\mathbf{c})}{\partial [\mathbf{c}]_j^*} &= \sigma_s^2 ((\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}}) \mathbf{c}) (\mathbf{1}_{Q+1, Q+1} \otimes \mathbf{I}_{N_d}) \frac{\partial \mathcal{D} \{ (\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}}) \mathbf{c} \}^H}{\partial [\mathbf{c}]_j^*}. \end{aligned}$$

To work out the last equation, we realize that in the matrix $\mathcal{D}((\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c})^H$, only the $(1 + \lfloor \frac{j-1}{L+1} \rfloor N_d)$ th until the $(N_d + \lfloor \frac{j-1}{L+1} \rfloor N_d)$ th diagonal entries are associated with $[\mathbf{c}]_j^*$, with the coefficients $\mathbf{G}_L^{\{\mathcal{D}\}*} \mathbf{e}_{\text{mod}(j-1, L+1)+1}$. Therefore,

$$\begin{aligned} & \frac{\partial \mathcal{D}((\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c})^H}{\partial [\mathbf{c}]_j^*} \\ &= \mathcal{D}\left\{[\mathbf{0}_{1, \lfloor \frac{j-1}{L+1} \rfloor N_d}, \mathbf{e}_{\text{mod}(j-1, L+1)+1}^T \mathbf{G}_L^{\{\mathcal{D}\}}, \mathbf{0}_{1, (Q - \lfloor \frac{j-1}{L+1} \rfloor) N_d}]^T\right\}. \end{aligned} \quad (4.78)$$

Coining some new matrix definitions:

$$\begin{aligned} \mathcal{F}_j &:= \mathbf{1}_{Q+1,1} \otimes \mathcal{D}\{\mathbf{G}_L^{\{\mathcal{D}\}*} \mathbf{e}_{\text{mod}(j-1, L+1)+1}\}, \\ \Gamma_j &:= \Gamma[\mathbf{0}_{N_d, \lfloor \frac{j-1}{L+1} \rfloor N_d} \mathbf{I}_{N_d} \mathbf{0}_{N_d, (Q - \lfloor \frac{j-1}{L+1} \rfloor) N_d}]^T, \end{aligned} \quad (4.79)$$

we can easily show

$$\frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{c}]_j^*} = \sigma_s^2 \Gamma \mathcal{D}((\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}\}})\mathbf{c}) \mathcal{F}_j \Gamma_j^H. \quad (4.80)$$

With the obtained FIM, we find a lower bound on the channel estimator's variance [48]

$$\text{CRB} = \text{tr}\{(\mathcal{J}(\mathbf{c}))^{-1}\} \leq \text{tr}\{\mathcal{E}_{\mathbf{s}^{\{\mathcal{D}\}}, \mathbf{v}_f}\{(\hat{\mathbf{c}} - \mathbf{c})(\hat{\mathbf{c}} - \mathbf{c})^H\}\}, \quad (4.81)$$

which is also a lower bound on the MSE of the BEM channel, as can be seen from (4.40).

Chapter 5

Channel Estimation based on Multiple OFDM Symbols

5.1 Introduction

In the previous chapter, the channel is estimated separately for each received OFDM symbol. From information theory, we understand that if more observation samples are available, the channel estimation performance can be improved. This idea will be explored in this chapter. It is nonetheless noteworthy that in the context of TV channels, the channel coherence time is relatively short, which means that only a limited number of OFDM symbols can be utilized for channel estimation.

With multiple OFDM symbols, we are endowed with the freedom to decide how to distribute the pilots among these OFDM symbols. In order to differentiate between diverse pilot patterns, we borrow in this chapter the terms that are used in [20]. The first scheme, referred to as the comb-type, is adopted in [58, 119, 89, 5, 67, 88, 104, 15]. In this scheme, pilots occupy only a fraction of the subcarriers, but such pilots are carried by each OFDM symbol. This is actually the pilot scheme that we discussed in the previous chapter but is now extended to multiple OFDM symbols. In contrast, the channel estimators in [18, 21, 91] adopt a block-type scheme, where the pilots occupy the entire OFDM symbol, and such pilot OFDM symbols are interleaved with data OFDM symbols. A third pilot scheme, considered in [17], is referred to as the mixed-type, which is a compromise between the comb- and block-types. To be more specific, the pilots only occupy a fraction of the subcarriers, and such pilot OFDM symbols are interleaved with data OFDM symbols*. An example of these three pilot schemes is sketched in Fig. 5.1.

The question arises as which scheme is able to yield the most reliable channel estimates under the same bandwidth/power restriction. Conflicting results are reported, e.g., the works in [69, 84, 17] advocate the comb- or mixed-type, while the block-type scheme is preferred in [21, 91]. Common to all pilot schemes, channel

The results of this chapter appeared in [104, 98] and [102, Chapter 3].

*In this chapter, we will refer to an OFDM symbol that contains pilots as a pilot OFDM symbol, and an OFDM symbol that contains exclusively data as a data OFDM symbol.

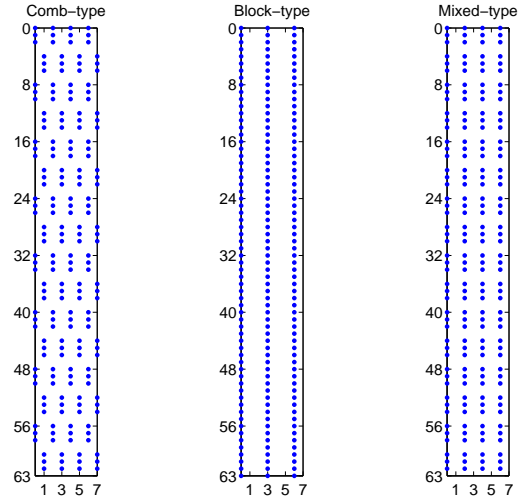


Figure 5.1: The pilot schemes. The horizontal axis corresponds to the time; the vertical axis corresponds to the subcarrier positions; the position where a pilot symbol is located is represented by a dot.

estimation can be decomposed into a two-step approach: first, preliminary channel estimates are acquired for individual pilot OFDM symbols; next, these preliminary results are interpolated to attain the final channel estimates with the aid of, e.g., channel second-order statistics (SOS), or a first-order polynomial. Usually, the block-type scheme can yield a better channel estimation for each pilot OFDM symbol because more pilots can be utilized. It is also noteworthy that the comb- or mixed-type pilots are usually adopted in the works which assume the channel is constant within the OFDM symbol except in [104, 15]. In the TV channel case where ICI cannot be ignore any more, they will suffer from a noise floor, whereas the pilots in the block-type scheme, by occupying the entire subcarriers, are not plagued by this problem. On the other hand, in the second step of channel estimation where the interpolation is invoked, the block-type scheme will in general suffer from a larger interpolation error because in the comb- and mixed-type schemes, the pilot OFDM symbols are placed much closer to each other.

This chapter aims to provide a better overview on the impact of diverse pilot placement schemes on the channel estimation performance. Compared to the existing works, improvements are made in two aspects:

1. in the comb- and mixed-type schemes, the channel will not be assumed con-

stant within the OFDM symbol, and the ICI will be more effectively addressed. To this end, the pilots will be lumped in clusters like we have done in the previous chapter. Besides, we will employ a best linear unbiased estimator (BLUE) [48] to combat ICI, which is similarly applied as in the previous chapter [97];

2. for interpolation (as well as channel modeling for each OFDM symbol), we will adopt a general BEM assumption. Note that the first-order polynomial interpolation used in [67] can be put in the framework of the P-BEM, and the interpolation using channel SOS (especially the reduced rank approach in [91]) corresponds to the DKL-BEM.

5.2 System Model and BEM for Multiple OFDM Symbols

A similar system model as considered in the previous chapter will be used here, where we consider the OFDM system to have N subcarriers. The TV channel that takes the transmit filter, the propagation environment, and the receive filter (and a possible window) into account is assumed to be an FIR with a maximal order L . If we use $h_{p,l}$ to denote the l th channel tap at the p th time-instance, then $h_{p,l} = 0$ for $l < 0$ or $l > L$. For the i th OFDM symbol, the data symbols $\mathbf{s}(i)$ are first modulated on N subcarriers as

$$\mathbf{x}(i) = \mathbf{W}_N^H \mathbf{s}(i). \quad (5.1)$$

Making abstraction of the digital-to-analog and analog-to-digital conversions, $\mathbf{x}(i)$ is next concatenated by a cyclic prefix (CP) of length $L_z \geq L$, sent over the channel, stripped from the CP, windowed and finally demodulated. The received data stream resulting from the i th OFDM symbol can be expressed in the frequency domain as

$$\mathbf{y}_f(i) = \mathbf{H}_f(i) \mathbf{s}(i) + \mathbf{v}_f(i), \quad (5.2)$$

where

$$\mathbf{v}_f(i) := \mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathbf{v}_t(i) \quad (5.3)$$

represents the noise in the frequency domain with $\mathbf{v}_t(i)$ being its counterpart in the time domain. $\mathcal{D}\{\mathbf{w}\}$ corresponds to the windowing operation with

$$\mathbf{w} := [w_0, \dots, w_{N-1}]^T \quad (5.4)$$

as the window elements. $\mathbf{H}_f(i)$ stands for the FD channel matrix, which is related with the TD channel matrix $\mathbf{H}_t(i)$ as

$$\mathbf{H}_f(i) := \mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathbf{H}_t(i) \mathbf{W}_N^H. \quad (5.5)$$

Conform the FIR assumption of the channel and letting $L_z = L$ without loss of generality, we understand that the entries of $\mathbf{H}_t(i)$ admits the expression as

$$[\mathbf{H}_t(i)]_{p,n} = h_{i(N+L)+p+L, \text{mod}(p-n, N)}. \quad (5.6)$$

In the following of the chapter, we will consider R consecutive OFDM symbols and assume that the symbol index i is ranged from 0 until $R - 1$. By employing multiple OFDM symbols, we are able to benefit from their mutual time correlation to enhance the estimation precision. Besides, this is indispensable for the block-type scheme, where the data symbols are transmitted in a different OFDM symbol than the pilots.

To introduce the BEM, let us first define an $NR \times 1$ vector \mathbf{h}_l that collects all the l th channel taps after windowing

$$\mathbf{h}_l := [\mathbf{h}_l(0)^T, \dots, \mathbf{h}_l(R-1)^T]^T, \quad (5.7)$$

for $l = 0, \dots, L$, with

$$\mathbf{h}_l(i) := [w_0 h_{i(N+L), l}, \dots, w_{N-1} h_{i(N+L)+N-1, l}]^T \quad (5.8)$$

representing the l th channel tap within the i th OFDM symbol after windowing. Note that there is a gap between the last index of $\mathbf{h}_l(i)$ and the first index of $\mathbf{h}_l(i+1)$ because the channel in between is discarded with the CP. As a result, if we want to use a BEM to approximate \mathbf{h}_l as

$$\mathbf{h}_l \approx \underbrace{\begin{bmatrix} \mathbf{b}_0(0) & \cdots & \mathbf{b}_Q(0) \\ \vdots & \ddots & \vdots \\ \mathbf{b}_0(R-1) & \cdots & \mathbf{b}_Q(R-1) \end{bmatrix}}_{\mathbf{B}} [c_{0,l}, \dots, c_{Q,l}]^T, \quad (5.9)$$

we need to first set up a larger BEM matrix with a dimension of $R(N+L)$ by $Q+1$, and designate $\mathbf{b}_q(i)$ as its q th column starting from row $i(N+L)$ until row $(i+1)(N+L)+N-1$. Hence for the i th OFDM symbol in particular,

$$\mathbf{h}_l(i) \approx \underbrace{[\mathbf{b}_0(i), \dots, \mathbf{b}_Q(i)]}_{\mathbf{B}(i)} [c_{0,l}, \dots, c_{Q,l}]^T, \quad (5.10)$$

Assuming that the BEM induces only a negligible modeling error, we can express the TD channel matrix in (5.6) in terms of the BEM as

$$\mathcal{D}\{\mathbf{w}\} \mathbf{H}_t(i) = \sum_{q=0}^Q \mathcal{D}\{\mathbf{b}_q(i)\} \mathbf{C}_q, \quad (5.11)$$

where \mathbf{C}_q denotes a circulant matrix with $[\mathbf{c}_q^T, \mathbf{0}_{1 \times (N-L-1)}]^T$ as its first column. Here,

$$\mathbf{c}_q := [c_{q,0}, \dots, c_{q,L}]^T. \quad (5.12)$$

As a result, the I/O relationship in (5.2) can be accordingly rewritten as

$$\begin{aligned} \mathbf{y}_f(i) &= \sum_{q=0}^Q \mathbf{W}_N \mathcal{D}\{\mathbf{b}_q(i)\} \mathbf{C}_q \mathbf{W}_N^H \mathbf{s}(i) + \mathbf{v}_f(i), \\ &= \sum_{q=0}^Q \mathbf{W}_N \mathcal{D}\{\mathbf{b}_q(i)\} \mathbf{W}_N^H \mathcal{D}\{\mathbf{G}_L \mathbf{c}_q\} \mathbf{s}(i) + \mathbf{v}_f(i), \end{aligned} \quad (5.13)$$

where the second equality is derived due to the fact that the circulant matrix \mathbf{C}_q corresponds in the frequency domain to a diagonal matrix

$$\mathcal{D}\{\mathbf{G}_L \mathbf{c}_q\} = \mathbf{W}_N \mathbf{C}_q \mathbf{W}_N^H, \quad (5.14)$$

with \mathbf{G}_L standing for an $N \times (L+1)$ matrix that consists of the first $L+1$ columns of \mathbf{W}_N , and is scaled by \sqrt{N} .

5.3 Channel Estimators and Pilot Schemes

5.3.1 Data Model for Channel Estimation

The channel estimation is accomplished in a similar way as in the previous chapter. For the i th OFDM symbol, let us use $\mathcal{T}(i)$ to denote the set that contains the indices of the pilot-carrying subcarriers, and $\mathcal{D}(i)$ to denote the set that contains the indices of the data-carrying subcarriers. Accordingly, the pilots carried in the i th OFDM symbol can be expressed as

$$\mathbf{t}(i) = \mathbf{s}^{\{\mathcal{T}(i)\}}(i), \quad (5.15)$$

and the data symbols carried in the i th OFDM symbol can be expressed

$$\mathbf{d}(i) = \mathbf{s}^{\{\mathcal{D}(i)\}}(i). \quad (5.16)$$

In addition, we use $\mathcal{O}(i)$ to denote the positions of the observation samples. for which we can derive their relationship with the pilots and data symbols from (5.13) as:

$$\mathbf{y}_f^{\{\mathcal{O}(i)\}}(i) = \mathcal{P}(i) \mathbf{c} + \mathcal{I}(i) \mathbf{c} + \mathbf{v}_f^{\{\mathcal{O}(i)\}}(i), \quad (5.17)$$

with $\mathbf{c} := [\mathbf{c}_0^T, \dots, \mathbf{c}_Q^T]^T$ and

$$\begin{aligned} \mathcal{P}(i) &:= \left[\mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{b}_0(i)\} \mathbf{W}_N^{\{\mathcal{T}(i)\}H}, \dots, \mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{b}_Q(i)\} \mathbf{W}_N^{\{\mathcal{T}(i)\}H} \right] \\ &\quad (\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{t}(i)\} \mathbf{G}_L^{\{\mathcal{T}(i)\}}), \\ \mathcal{I}(i) &:= \left[\mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{b}_0(i)\} \mathbf{W}_N^{\{\mathcal{D}(i)\}H}, \dots, \mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{b}_Q(i)\} \mathbf{W}_N^{\{\mathcal{D}(i)\}H} \right] \\ &\quad (\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{d}(i)\} \mathbf{G}_L^{\{\mathcal{D}(i)\}}). \end{aligned} \quad (5.18)$$

In (5.17), the contribution to the observation samples is decomposed into two parts, where $\mathcal{P}(i)$ is entirely dependent on pilots, while $\mathcal{I}(i)$ is entirely dependent on unknown data symbols. Note that (5.17) is the same as (4.26), but in a slightly different appearance.

Further, let us use \mathcal{V} to denote the index set for those OFDM symbols that contain pilots with i_v standing for the v th component. We suppose there are in total V pilot OFDM symbols with

$$\mathcal{V} = \{i_0, \dots, i_{V-1}\}. \quad (5.19)$$

By the definition of the diverse pilot scheme, we understand that

- **Comb-type scheme** has $|\mathcal{V}| = R$ since each OFDM symbol carries pilot, and the pilots occupy only a fraction of the subcarriers, i.e., $|\mathcal{T}(i)| < N$.
- **Mixed-type scheme** has $|\mathcal{V}| < R$ since the pilot OFDM symbols are interleaved with the data OFDM symbols, and the pilots occupy only a fraction of the subcarriers, i.e., $|\mathcal{T}(i)| < N$ if $i \in \mathcal{V}$.
- **Block-type scheme** has $|\mathcal{V}| < R$ since the pilot OFDM symbols are interleaved with the data OFDM symbols, but the pilots occupy all the subcarriers, i.e., $|\mathcal{T}(i)| = N$ if $i \in \mathcal{V}$.

Note that in a general comb/mixed-type, it is not mandatory that the positions of the pilot subcarriers in different OFDM symbols are aligned with each other, i.e., it is allowed that $\mathcal{T}(i) \neq \mathcal{T}(i')$. An example of these three pilot schemes is sketched in Fig. 5.1.

With the aid of the above notations, we can apply a similar operation as in (5.17) to each pilot OFDM symbol, and stack the results in one vector. This leads to

$$\mathbf{y}_f = \mathcal{P}\mathbf{c} + \mathcal{I}\mathbf{c} + \mathbf{v}_f, \quad (5.20)$$

where

$$\mathbf{y}_f := [\mathbf{y}_f^{\{\mathcal{O}(i_0)\}T}(i_0), \dots, \mathbf{y}_f^{\{\mathcal{O}(i_{V-1})\}T}(i_{V-1})]^T, \quad (5.21)$$

and \mathcal{P} , \mathcal{I} , and \mathbf{v}_f are similarly defined as \mathbf{y}_f .

It is important to remark here that in the block-type scheme, the effect of the pilots are detached from the data, and thus the observation samples are free from the interference. In that case, (5.20) reduces to

$$\mathbf{y}_f = \mathcal{P}\mathbf{c} + \mathbf{v}_f. \quad (5.22)$$

5.3.2 Channel Estimators

We consider the BLUE, which can be analogously implemented as in the previous chapter. Here, we briefly summarize its working principle: from (5.20), suppose $\hat{\mathbf{c}}^{(k)}$ denotes the channel estimate obtained at the k th iteration. In the next iteration, we compute:

$$\hat{\mathbf{c}}_{\text{BLUE}}^{(k+1)} = (\mathcal{P}^H \mathbf{R}^{(k)-1} \mathcal{P})^{-1} \mathcal{P}^H \mathbf{R}^{(k)-1} \mathbf{y}, \quad (5.23)$$

with $\mathbf{R}^{(k)}$ standing for the covariance matrix of the interference and noise resulting from the current estimate $\hat{\mathbf{c}}^{(k)}$. Assuming that the noise and data are uncorrelated and independent in different OFDM symbols, we have

$$\mathbf{R}^{(k)} := \mathcal{E}\{\mathcal{I}\hat{\mathbf{c}}^{(k)}\hat{\mathbf{c}}^{(k)H}\mathcal{I}^H + \mathbf{v}_f\mathbf{v}_f^H\}, \quad (5.24)$$

$$= \begin{bmatrix} \mathbf{R}^{(k)}(i_0) & & \\ & \ddots & \\ & & \mathbf{R}^{(k)}(i_{V-1}) \end{bmatrix}, \quad (5.25)$$

with

$$\mathbf{R}^{(k)}(i) := \mathcal{E}_{\mathbf{d}(i), \mathbf{v}_f(i)}\{\mathcal{I}(i)\hat{\mathbf{c}}^{(k)}\hat{\mathbf{c}}^{(k)H}\mathcal{I}^H(i) + \mathbf{v}_f^{\{\mathcal{O}(i)\}}(i)\mathbf{v}_f^{\{\mathcal{O}(i)\}H}(i)\}. \quad (5.26)$$

In the sequel of this paper, the following assumption will be made.

Assumption 5.1. *All the received samples of the pilot OFDM symbols will be used as the observation samples for channel estimation, i.e.,*

$$|\mathcal{O}(i)| = N. \quad (5.27)$$

For the block-type scheme, this assumption speaks for itself. For the comb- and mixed-type schemes, this is required for the BLUE to achieve the optimal performance. Requiring $|\mathcal{O}(i)| = N$ is equivalent to minimizing the assumed bandwidth B_c in Fig. 4.2. See the previous chapter for more details.

It is noteworthy that since the block-type scheme is free from interference, (5.26) becomes

$$\mathbf{R}^{(k)}(i) := \mathcal{E}_{\mathbf{v}_f(i)}\{\mathbf{v}_f^{\{\mathcal{O}(i)\}}(i)\mathbf{v}_f^{\{\mathcal{O}(i)\}H}(i)\}. \quad (5.28)$$

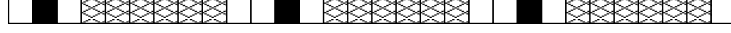


Figure 5.2: The “FDKD” pilot scheme, where the black boxes stand for non-zero pilots, the white boxes for zero pilots. The remaining boxes corresponds to the data symbols.

If we further assume that the noise prior to windowing is zero-mean white with power σ^2 , the BLUE in (5.23) will reduce to

$$\hat{\mathbf{c}} = (\mathcal{P}^H \mathbf{R}_v^{-1} \mathcal{P})^{-1} \mathcal{P}^H \mathbf{R}_v^{-1} \mathbf{y}_f, \quad (5.29)$$

with

$$\mathbf{R}_v = \sigma^2 \mathbf{W}_N \mathcal{D}\{\mathbf{w}\} \mathcal{D}\{\mathbf{w}^H\} \mathbf{W}_N^H. \quad (5.30)$$

The estimator in (5.29) is a weighted LS estimator, and can be attained in one shot.

5.4 Channel Identifiability

Channel identifiability is defined in this chapter as the existence of a unique channel estimator. For the LS estimator used in the block-type scheme, this is equivalent to requiring that the matrix \mathcal{P} in (5.29) be of full column rank. The same requirement is posed on the BLUE in the comb- and mixed-type schemes, which is adequate since the covariance matrix $\mathbf{R}^{(k)}(i)$ defined in (5.26) is always positive definite for a zero-mean white noise and a practical SNR, and so is its inverse. In that case, $\mathcal{P}^H \mathbf{R}^{(k)-1} \mathcal{P}$ in (5.23) will be non-singular if \mathcal{P} has full rank.

Conform Assumption 5.1, \mathcal{P} is a $VN \times (L+1)(Q+1)$ matrix, which is usually tall. Hence, the channel identifiability is paraphrased in this chapter as

$$\text{Rank}\{\mathcal{P}\} = (L+1)(Q+1). \quad (5.31)$$

In practice, the rank condition on \mathcal{P} depends on many factors, such as the choice of BEM, or the pilot structure. Especially for the latter, it turns out to be very hard to give a rigorous formulation on (5.31) if a general pilot structure is used. In this chapter, we will adopt a specific pilot structure assumption for each pilot OFDM symbol, which is referred to as the “FDKD” pilot scheme in [47].

Assumption 5.2. Assuming that total number of subcarriers equals

$$N = PM \quad (5.32)$$

with P and M integers, we group the pilots into P (cyclically) equi-distant clusters, where each cluster takes on a pulse shape form, i.e., a non-zero pilot is guarded by $\frac{D}{2}$ zero pilots on both sides.

Note that for the derivation that follows, the position of the non-zero pilot inside one cluster is actually not important. In that sense, the zero guard pilots might be a misnomer. An example of such “FDKD” pilots is schematically illustrated in Fig. 5.2.

Although the optimal pilot scheme remains to be explored, the “FDKD” pilot scheme (and its variant “TDKD” pilot scheme for single-carrier systems [59]) are shown to be optimal under a special (C)CE-BEM assumption in [47]. For a general BEM assumption as taken in [97], the “FDKD” seems also to yield a good performance experimentally.

In what follows, we find it instrumental to first explore the rank condition on $\mathcal{P}(i)$ for the single OFDM symbol case, and then extend the results to the multiple pilot OFDM symbol case.

5.4.1 Rank Condition of $\mathcal{P}(i)$

To begin with, let us describe the “FDKD” pilots contained in the i th pilot OFDM symbol in a vector form as

$$\mathbf{t}(i) = \bar{\mathbf{t}}(i) \otimes [\mathbf{0}_{1 \times \frac{D}{2}}, 1, \mathbf{0}_{1 \times \frac{D}{2}}]^T, \quad (5.33)$$

where $\bar{\mathbf{t}}(i)$ contains all the P non-zero pilots. Let us further use $\bar{T}(i)$ to denote the positions of the subcarriers carrying these non-zero pilots. Under the equi-distance assumption, these non-zero pilots are also equally distributed with distance M . Hence,

$$\bar{T}(i) := \{\mu(i), \mu(i) + M, \dots, \mu(i) + M(P - 1)\}. \quad (5.34)$$

with $\mu(i)$ denoting the position of the first non-zero pilot in the i th OFDM symbol. Discarding the columns (rows) corresponding to the positions of the zero pilots, we are allowed to rewrite the matrix $\mathcal{P}(i)$ defined in (5.18) as

$$\mathcal{P}(i) = \mathbf{Z}(i)\mathbf{U}(i), \quad (5.35)$$

with

$$\begin{aligned} \mathbf{Z}(i) &= \mathbf{W}_N^{\{\mathcal{O}(i)\}} [\mathcal{D}\{\mathbf{b}_0(i)\} \quad \dots \quad \mathcal{D}\{\mathbf{b}_Q(i)\}] (\mathbf{I}_{Q+1} \otimes \mathbf{W}_N^{\{\bar{T}(i)\}H}), \\ \mathbf{U}(i) &= \mathbf{I}_{Q+1} \otimes \mathcal{D}\{\bar{\mathbf{t}}(i)\} \mathbf{G}_L^{\{\bar{T}(i)\}}. \end{aligned} \quad (5.36)$$

Because $\mathbf{G}_L^{\{\bar{T}(i)\}}$ consists of P equi-distant rows of \mathbf{G}_L , which is introduced in (5.14), it is a $P \times (L + 1)$ Vandermonde matrix, which implies that

$$\text{Rank}\{\mathbf{U}(i)\} = \min(P, L + 1)(Q + 1). \quad (5.37)$$

Under Assumption 5.1 and 5.2, we can show that (see Appendix 5A for a proof)

$$\text{Rank}\{\mathbf{Z}(i)\} = \min\left(\frac{N}{P}, Q + 1\right)P. \quad (5.38)$$

For an $m \times k$ matrix \mathbf{A} and a $k \times n$ matrix \mathbf{B} , the rank inequality reads [43]

$$(\text{Rank}\{\mathbf{A}\} + \text{Rank}\{\mathbf{B}\}) - k \leq \text{Rank}\{\mathbf{AB}\} \leq \min(\text{Rank}\{\mathbf{A}\}, \text{Rank}\{\mathbf{B}\}). \quad (5.39)$$

Now that $\mathbf{Z}(i)$ is a $PM \times P(Q + 1)$ matrix, and $\mathbf{U}(i)$ is a $P(Q + 1) \times (L + 1)(Q + 1)$ matrix, we understand that in order for the matrix product $\mathbf{Z}(i)\mathbf{U}(i)$ to have a rank $(L + 1)(Q + 1)$, it is sufficient that

$$P \geq (L + 1)$$

from (5.37) and

$$M = \frac{N}{P} \geq Q + 1$$

from (5.38). Let us summarize the above results in the following theorem,

Theorem 5.1. *Within a single pilot OFDM symbol, the channel will be identifiable for a practical BEM if the equi-distant “FDKD” pilots are used and if the number of pilot clusters P satisfies*

$$\frac{N}{Q + 1} \geq P \geq L + 1. \quad (5.40)$$

The following remarks are in order at this stage.

Remark 5.1.

It is noteworthy that the length of the guard interval D of each pilot cluster is not mentioned in Theorem 5.1. For the comb- and mixed- type pilot schemes, this implies that it is even possible to take $D = 0$ in some special cases. For a fast fading channel, the power of the interference $\mathcal{I}(i)$ could therefore be dominant, which is, however, not the biggest concern for the BLUE. Later on, we will give a lower bound on the total number of pilots in terms of an asymptotic expression of the BLUE.

Remark 5.2.

Theorem 5.1 holds in general true irrespective of the specific pilot value. However, for the block-type pilot scheme, we must be cautious if an all non-zero pilot structure is to be taken. Such a pilot structure can be viewed as a special form of the equi-distant “FDKD” pilots with $P = N$ and $D = 0$. In this case, Theorem 5.1, especially the first inequality in (5.40), can only be satisfied if $Q = 0$, which happens only for time-invariant channels. In this case, $\mathcal{P}(i)$ is not necessarily of full rank, which will depend on the specific value of the pilots $\mathbf{t}(i)$. As a result, the channel is not guaranteed to be identifiable.

5.4.2 Rank Condition of \mathcal{P}

In many practical situations, Theorem 5.1 is hard to satisfy. We have mentioned one such example in Remark 5.2. Another scenario that can be commonly encountered is when the channel delay spread L is large. In that case, it will be difficult, especially for the comb-type scheme, to afford enough pilot clusters P to match the channel length as the second inequality of (5.40) requires. If channel estimation on a single OFDM symbol is not possible, a solution is to employ multiple pilot OFDM symbols.

First of all, the following two lemmas proves useful.

Lemma 5.1. *If each pilot symbol contains the same equi-distant “FDKD” pilots: the pilot values and positions are identical, i.e.,*

$$\begin{aligned}\bar{\mathbf{t}}(i_0) &= \cdots = \bar{\mathbf{t}}(i_{V-1}), \\ \mu(i_0) &= \cdots = \mu(i_{V-1}),\end{aligned}\tag{5.41}$$

then \mathcal{P} will have full rank provided that

$$\frac{VN}{Q+1} \geq P \geq (L+1).\tag{5.42}$$

The proof will be given in Appendix 5B. In the above, we have used the notation $\mu(i)$ that is defined in (5.34) to denote the position of the first non-zero pilot in the i th pilot OFDM symbol. Because the pilots in each OFDM symbol are grouped in equi-distant clusters, each cluster having only one non-zero pilot (Assumption 5.2), the positions of all the pilots in the i th pilot OFDM symbol are uniquely determined by $\mu(i)$.

Lemma 5.2. *If the positions of the non-zero pilots in each pilot OFDM symbol are shifted with respect to each other, i.e.,*

$$\mu(i_0) \neq \cdots \neq \mu(i_{V-1}),\tag{5.43}$$

then the matrix \mathcal{P} will have full rank provided that

$$\frac{N}{Q+1} \geq P \geq \frac{L+1}{V}.\tag{5.44}$$

The proof will be given in Appendix 5C. In addition, we will need the following assumption:

Assumption 5.3. *We assume that the entire number of pilot OFDM symbols satisfies*

$$V = V_a V_b\tag{5.45}$$

with V_a and V_b being integers. Let us group the pilot OFDM symbols into V_a clusters, each containing V_b pilot OFDM symbols. Inside each cluster, we will make the pilots exactly identical to each other [c.f. (5.41)], whereas among different clusters, we will displace the positions of the non-zero pilots in each pilot OFDM symbol [c.f. (5.43)].

Combining Lemma 5.1 and 5.2, and conform Assumption 5.3, we come up with the following theorem that guarantees the full-rank condition of \mathcal{P} .

Theorem 5.2. *For multiple OFDM symbols, the channel will be identifiable for a practical BEM if the “FDKD” pilot structure follows Assumption 5.3 and if the number of pilot clusters P inside each OFDM symbol satisfies*

$$\frac{V_b N}{Q+1} \geq P \geq \frac{L+1}{V_a}. \quad (5.46)$$

The proof is in a great deal analogous to the proof of Lemma 5.1 and 5.2, and will be omitted here.

5.4.3 Channel Identifiability at a High SNR

Theorem 5.2 (as well as Theorem 5.1) does not impose an explicit restriction on the length of the guard band, D , in the “FDKD” pilots. For the LS estimator in (5.29) that is used for the block-type scheme, this restriction is implicitly imposed since the pilots occupy all the subcarriers, thus $P(D+1) = PM = N$. For the other two pilot schemes, where the BLUE is applied, D will only become significant to the channel identifiability at high SNR.

To understand this, recall that when we start to discuss the channel identifiability related to the BLUE, we have assumed that the covariance matrix $\mathbf{R}^{(k)}$ given in (5.28) is always positive-definite. Due to the white Gaussian noise assumption, this is reasonable for a practical SNR, but could lose its validity when the SNR becomes large. This makes sense since in that case the noise is small, the performance of the channel estimator will be more susceptible to the interference due to the data symbols. A longer guard band D within the pilot clusters will be helpful to reduce the impact of the data interference. To facilitate a deeper insight, let us express the covariance matrix $\mathbf{R}^{(k)}(i)$ of the data interference defined in (5.26) as

$$\mathbf{R}^{(k)}(i) = \mathbf{\Gamma}(i) \mathbf{\Phi}^{(k)}(i) \mathbf{\Gamma}^H(i) + \sigma^2 \mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{w}\} \mathcal{D}\{\mathbf{w}^H\} \mathbf{W}_N^{\{\mathcal{O}(i)\}H}. \quad (5.47)$$

It can be shown that [c.f. (4.28)]

$$\begin{aligned} \mathbf{\Gamma}(i) &:= \left[\mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{b}_0(i)\} \mathbf{W}_N^{\{\mathcal{D}(i)\}H}, \dots, \mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{b}_q(i)\} \mathbf{W}_N^{\{\mathcal{D}(i)\}H} \right], \\ \mathbf{\Phi}^{(k)}(i) &:= \mathcal{D}\{(\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}(i)\}}) \hat{\mathbf{c}}^{(k)}\} (\mathbf{1}_{Q+1, Q+1} \otimes \mathcal{E}\{\mathbf{d}(i) \mathbf{d}^H(i)\}) \\ &\quad \mathcal{D}\{(\mathbf{I}_{Q+1} \otimes \mathbf{G}_L^{\{\mathcal{D}(i)\}}) \hat{\mathbf{c}}^{(k)}\}^H. \end{aligned} \quad (5.48)$$

Assuming the data to be zero-mean white with variance σ_s^2 , we can rewrite $\Phi^{(k)}(i)$ after some algebra as

$$\Phi^{(k)}(i) = \sigma_s^2 \mathcal{G}^{(k)}(i) \mathcal{G}^{(k)H}(i), \quad (5.49)$$

with

$$\mathcal{G}^{(k)}(i) := \left[\mathcal{D}\{\mathbf{G}_L^{\{\mathcal{D}(i)\}} \hat{\mathbf{c}}_0^{(k)}\}, \dots, \mathcal{D}\{\mathbf{G}_L^{\{\mathcal{D}(i)\}} \hat{\mathbf{c}}_Q^{(k)}\} \right]^T. \quad (5.50)$$

Further, because the noise covariance matrix is positive definite, it can be decomposed as

$$\sigma^2 \mathbf{W}_N^{\{\mathcal{O}(i)\}} \mathcal{D}\{\mathbf{w}\} \mathcal{D}\{\mathbf{w}^H\} \mathbf{W}_N^{\{\mathcal{O}(i)\}H} = \sigma^2 \mathcal{M}^{-1}(i) \mathcal{M}^{-H}(i), \quad (5.51)$$

with $\mathcal{M}(i)$ standing for a non-singular square matrix. As a result, the inversion of the covariance matrix $\mathbf{R}^{(k)}(i)$ can be expressed as

$$\mathbf{R}^{(k)-1}(i) = \mathcal{M}^{-1}(i) \Psi(i) \mathcal{M}^{-H}(i), \quad (5.52)$$

with

$$\Psi(i) := (\mathcal{M}(i) \Gamma(i) \mathcal{G}^{(k)}(i) \mathcal{G}^{(k)H}(i) \Gamma^H(i) \mathcal{M}^H(i) + \sigma^2 \mathbf{I})^{-1}. \quad (5.53)$$

Utilizing the matrix inversion lemma, we can express the above as

$$\begin{aligned} \Psi(i) &= \sigma^{-2} \mathbf{I} - \sigma^{-2} \mathcal{M}(i) \Gamma(i) \mathcal{G}^{(k)}(i) \left(\frac{\sigma^2}{\sigma_s^2} \mathbf{I} + \mathcal{G}^{(k)H}(i) \Gamma^H(i) \mathcal{M}^H(i) \right. \\ &\quad \left. \mathcal{M}(i) \Gamma(i) \mathcal{G}^{(k)}(i) \right)^{-1} \mathcal{G}^{(k)H}(i) \Gamma^H(i) \mathcal{M}^H(i). \end{aligned} \quad (5.54)$$

In the high SNR region, where $\frac{\sigma_s^2}{\sigma^2} \rightarrow \infty$, the above becomes

$$\begin{aligned} \Psi(i) &\approx \sigma^{-2} \mathbf{I} - \sigma^{-2} \mathcal{M}(i) \Gamma(i) \mathcal{G}^{(k)}(i) (\mathcal{G}^{(k)H}(i) \Gamma^H(i) \mathcal{M}^H(i) \mathcal{M}(i) \Gamma(i) \mathcal{G}^{(k)}(i))^{-1} \\ &\quad \mathcal{G}^{(k)H}(i) \Gamma^H(i) \mathcal{M}^H(i). \end{aligned} \quad (5.55)$$

In the above, because $\mathcal{M}(i)$ is a non-singular matrix, the rank condition of $\Psi(i)$ depends on $\Gamma(i) \mathcal{G}^{(k)}(i)$, which is unfortunately not always of full column-rank: for the diagonal block components $\mathcal{D}\{\mathbf{G}_L^{\{\mathcal{D}(i)\}} \mathbf{c}_q^{(k)}\}$ of $\mathcal{G}^{(k)}(i)$ in (5.50), it is possible that they have common zeros at the same diagonal position. For such a rank-deficient $\Gamma(i) \mathcal{G}^{(k)}(i)$, $\mathbf{R}^{(k)}(i)$ can already become extremely ill-conditioned even at a moderate-to-high SNR. In that case, we have to replace the inverse operation in (5.23) with the pseudo-inverse operation, which leads to a sub-optimal BLUE

$$\hat{\mathbf{c}}_{\text{BLUE}}^{(k+1)} = (\mathcal{P}^H \mathbf{R}^{(k)\dagger} \mathcal{P})^{-1} \mathcal{P}^H \mathbf{R}^{(k)\dagger} \mathbf{y}. \quad (5.56)$$

For such a sub-optimal BLUE, the full rank condition of \mathcal{P} alone is not sufficient to guarantee the channel identifiability, but we must focus on the rank condition of $\mathcal{P}^H \mathbf{R}^{(k)\dagger} \mathcal{P}$. In general, its full rank condition is difficult to prove, and we can only find a necessary condition that is stated in the following lemma.

Lemma 5.3. *In order for the channel to be identifiable using the BLUE at high SNR, it is necessary that the total number of pilots be greater or equal to the number of unknowns, i.e.,*

$$VP(D+1) \geq (L+1)(Q+1). \quad (5.57)$$

Proof. The $PM \times PM$ matrix $\Psi(i)$ in (5.55) lies in the noise subspace of $\mathcal{M}(i)\Gamma(i)\mathcal{G}^{(k)}(i)$, i.e.,

$$\Psi(i)\mathcal{M}(i)\Gamma(i)\mathcal{G}^{(k)}(i) = \mathbf{0}. \quad (5.58)$$

For the non-singular $\mathcal{M}(i)$, and assuming that the $PM \times P(M-D-1)$ matrix $\Gamma(i)\mathcal{G}^{(k)}(i)$ has a full column-rank, we understand that

$$\text{Rank}\{\Psi(i)\} \leq P(D+1). \quad (5.59)$$

and hence

$$\text{Rank}\{\mathbf{R}^{(k)\dagger}\} \leq VP(D+1). \quad (5.60)$$

In light of (5.56), we desire that the $(L+1)(Q+1) \times (L+1)(Q+1)$ matrix $\mathcal{P}^H \mathbf{R}^{(k)\dagger} \mathcal{P}$ be non-singular. This is only possible if (5.57) is satisfied. \square

5.5 Some Simulation Results and Discussion

We test the different channel estimators/pilot schemes for TV channels that follow Jakes' Doppler profile [45] utilizing the TV channel generator that is given in [124]. We assume the channel to be an FIR filter with each filter tap having an independent Gaussian distribution with zero-mean and a uniform multi-path profile, i.e., $\mathcal{E}\{|h_{n,l}|^2\} = \frac{1}{L+1}$. Further, a flat window is used for channel estimation[†].

We approximate the channel time-variation by means of the (O)CE-BEM, which spans in total eight OFDM symbols including the cyclic prefix, with each OFDM symbol containing $N = 64$ subcarriers. Note that other BEMs are also viable, but will not be tested here. The scale of the BEM is associated with the normalized Doppler spread ν_D . In the simulation, we set $Q = 4$ if $\nu_D \leq 0.002$ and otherwise $Q = 8$.

The pilot schemes to be compared are plotted in Fig. 5.1. To be more specific, we let each OFDM symbol carry pilots in the comb-type scheme, thus $\mathcal{V} = \{0, \dots, 7\}$. The pilots are grouped in $P = 8$ equi-distant "FDKD" clusters, each of length $D+1 = 3$. In the block-type scheme, the indices of the pilot OFDM symbols are collected in $\mathcal{V} = \{0, 3, 6\}$. Inside each pilot OFDM symbols, we set further $[P, D] = [16, 3]$. Note

[†]We did not use a window that is introduced in the previous chapters because it is only significant for low-complexity channel equalization.

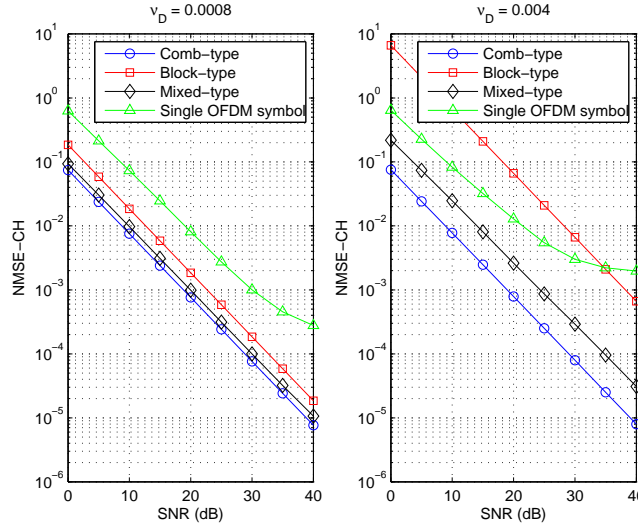


Figure 5.3: NMSE vs SNR for $L = 3$.

that different than the most traditional works, we let the block-type scheme also carry “FDKD” pilots to ensure the channel identifiability (see Remark 5.2). In the mixed-type scheme, we set $\mathcal{V} = \{0, 2, 4, 6\}$, and $[P, D] = [16, 2]$. In this way, all the pilot schemes result in an equal loss in bandwidth (37.5%).

We use the normalized mean square error (NMSE) with respect to the actual TV channel (hence the BEM approximation error is also taken into account) as performance measure [c.f. (5.10)]:

$$\text{NMSE-CH}(i) = \frac{1}{N} \sum_l \|\mathbf{h}_l(i) - \mathbf{B}(i)[\hat{c}_{0,l}, \dots, \hat{c}_{Q,l}]^T\|^2, \quad (5.61)$$

where $\mathbf{B}(i)$ is defined in (5.10) as the part of the BEM matrix that corresponds to the i th OFDM symbol. To combat the BEM modeling error due to a large BEM window size, we adopt a sliding window approach. In other words, we will consider the NMSE only for the fourth and fifth OFDM symbol in the comb- and mixed-type scheme ($i = 3$ and $i = 4$), and the fifth and sixth OFDM symbols in the block-type scheme ($i = 4$ and $i = 5$).

Test Case 1. Comparison of channel estimation performance

In the left and right plot of Fig. 5.3, we depict the performances of the three pilot schemes for short channels, $L = 3$, with a normalized Doppler spread $\nu_D = 0.0008$

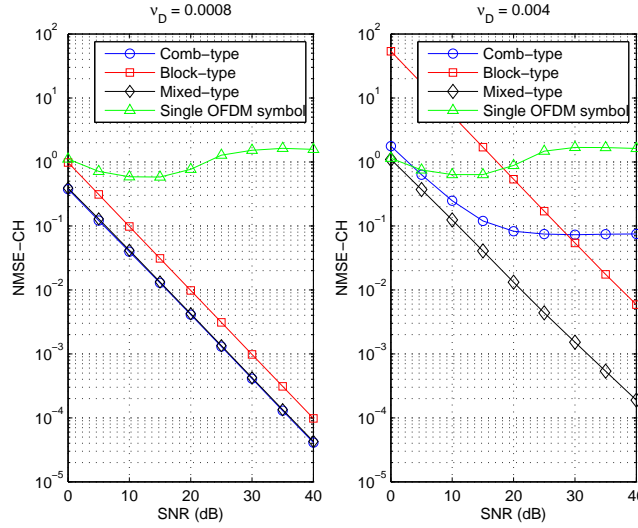


Figure 5.4: NMSE vs SNR for $L = 15$.

and $\nu_D = 0.004$, respectively. In addition, we list the performance based on a single OFDM symbol, which is a special case of the comb-type scheme but with $|\mathcal{V}| = 1$ (for this case, we set $Q = 2$). We can observe in the left plot of Fig. 5.3 that when the channel fading is slow, the three pilot schemes yield similar performances, which are all better than the performance of the single OFDM symbol case. When the channel fades faster as in the right plot, the block-type scheme endures more difficulty in tracking the channel compared to the other schemes.

To explain this, let us interpret the pilot-aided channel estimation as a kind of interpolation, where the positions of the pilot OFDM symbols correspond to the interpolation points. In the block-type scheme, these interpolation points are agglomerated in blocks, which are separated relatively farther apart from each other than in the other two schemes. It is hence not hard to imagine that such a pilot pattern is not capable of tracking promptly the ups-and-downs that arise more often in fast channels. This drawback is alleviated in the other two schemes, where the pilot OFDM symbols are much closer to each other. As can be seen, the performance of the block-type scheme is even inferior to that of the single OFDM symbol case in most SNR regimes.

The results for a much longer channel, $L = 15$, are depicted in Fig. 5.4. We observe that when the channel varies slow, the block-type scheme renders a similar

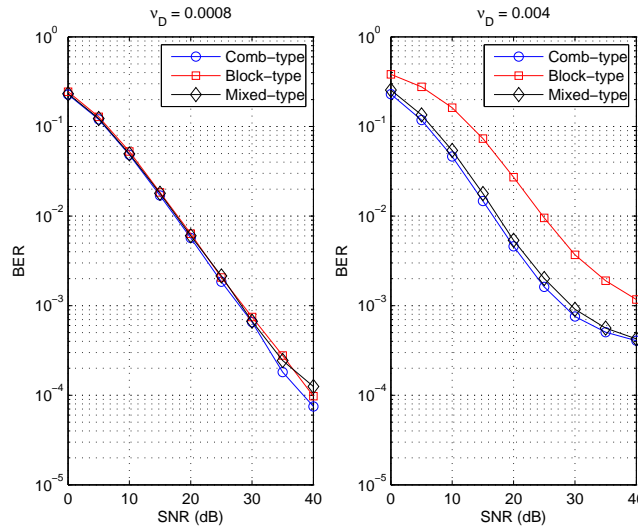


Figure 5.5: BER vs SNR for $L = 3$.

performance as the other two, but gets worse when the channel varies much faster as shown in the right plot. Interestingly, the performance of the comb-type scheme degrades in the right plot more severely, and suffers from a high noise floor at a moderate-to-high SNR. Compared to the other channel situations, this suggests that the comb-type is inferior for a channel that is overspread in both the Doppler and delay domain. We will study this effect in more detail later on. Further, note that the channel with $L = 15$ is not identifiable any more within a single OFDM symbol as evident in Fig. 5.4, since it only accommodates $P = 8$ pilot clusters, and consequently, the condition in Theorem 5.1 is violated.

Test Case 2. Comparison of channel equalization performance

It is practically of more significance to see how the estimation error will impact the BER performance. To this end, we construct a block LMMSE channel equalizer that is based on the full block matrix knowledge of the frequency-domain channel, estimated from the three pilot schemes. The BER results are depicted in Fig. 5.5 and Fig. 5.6, where one can easily remark that the channel equalization performance corresponds roughly to the channel estimation performance that is depicted in Fig. 5.3 and Fig. 5.4.

Test Case 3. Comparison of channel estimation performance at high SNR

In Fig. 5.7, we plot the channel estimation performance in relationship with the

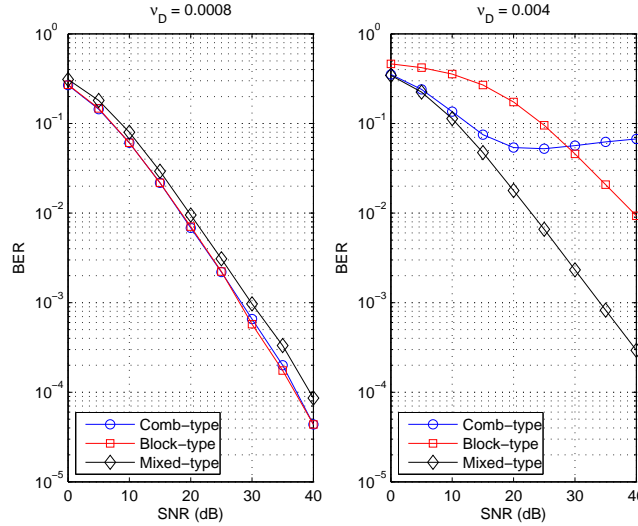


Figure 5.6: BER vs SNR for $L = 15$.

Doppler spread, which is obtained at an SNR equal to 40dB for $L = 3$ and $L = 15$, respectively. We observe that all the three pilot schemes are more or less robust to the Doppler spread. There is indeed a jump at the point $\nu_D = 0.0025$, where the BEM scale is enlarged to $Q = 8$, and thereby almost doubles the number of channel unknowns, which equals $(L + 1)(Q + 1)$. Nevertheless, the performance after this point becomes steady again.

However, the comb-type scheme is worth paying more attention to because it exhibits an extraordinary behavior with regard to this performance drop. For short channels with $L = 3$, it renders a superior channel estimation performance regardless of the Doppler spread change. Oppositely, when the channel is much longer, $L = 15$, its performance drops drastically at the point the BEM scale needs to be enlarged. We understand from Lemma 5.3 that the total number of channel unknowns must be smaller than the total number of pilots in order for the channel to be identifiable at high SNR. For a channel that is overspread in both the Doppler and delay domain, this condition seems to become more critical, and the performance of the comb-type scheme is apparently very sensitive to it.

In Fig. 5.8, we compare the channel estimation performance of the three pilot schemes with respect to the channel length L , which is obtained at an SNR equal to 40dB for $\nu_D = 0.0008$ and $\nu_D = 0.004$, respectively. Similar to the situation in

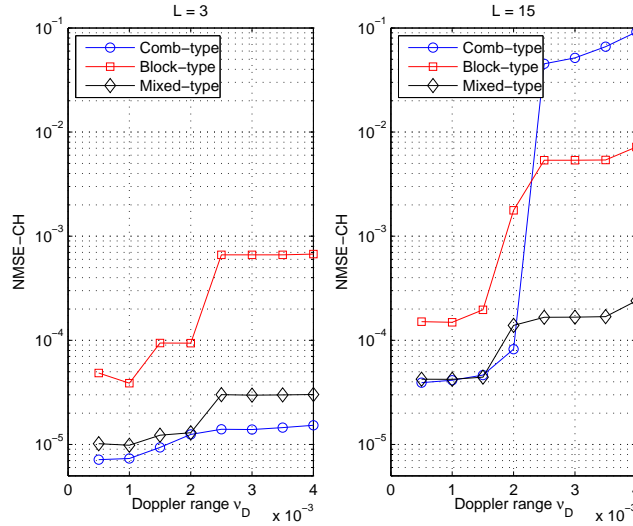


Figure 5.7: NMSE vs the normalized Doppler spread.

Fig. 5.7, one can remark that for slow channels, the performance of the comb-type scheme is superior and robust to the channel length. In contrast, for fast channels, its performance undergoes a drastic jump. Interestingly enough, this jump happens specifically at the point $L = 8$, where the comb-type scheme with its number of pilot clusters, $P = 8$, begins to break the condition in Theorem 5.1, i.e., the channel loses identifiability within a single pilot OFDM symbol. We see that when the total number of channel unknowns becomes larger, or in other words, the condition in Lemma 5.3 becomes more critical, Theorem 5.1 plays a more significant role even when Theorem 5.2 is satisfied. This is probably due to the fact that in that case, the matrix \mathcal{P} , though still of full rank, becomes quite ill conditioned, and the BLUE will be more susceptible to noise and interference and be more difficult to converge.

It is noteworthy that the mixed-type scheme satisfies both Theorem 5.1 and Theorem 5.2 for all the channel lengths. In addition, it withstands the fast channel variation much better than the block-type scheme thanks to the fact that the pilot OFDM symbols are closer to each other. These two factors endow the mixed-type scheme with a very robust channel estimation performance for all fading situations.

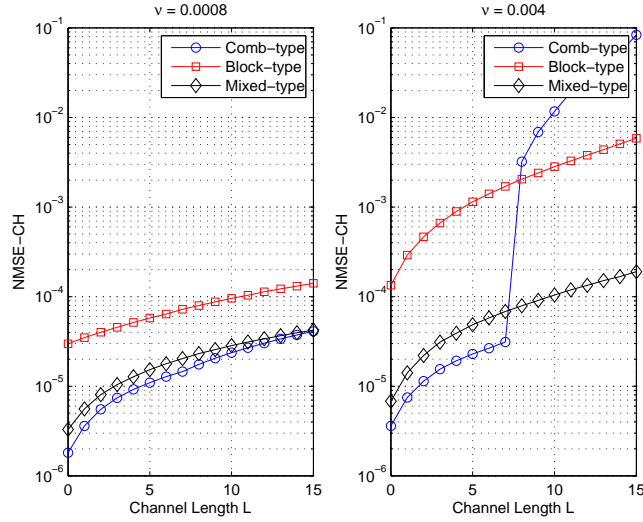


Figure 5.8: NMSE vs the channel length.

5.6 Summary

In this chapter, TV channel estimation based on multiple OFDM symbols was discussed. We compared three pilot arrangement schemes, namely, the comb-, block- and mixed-type schemes: in the comb-type scheme, each OFDM symbol carries pilots, which are interleaved with data in the frequency domain; in the block-type scheme, only a few of the OFDM symbols carries pilots, but these pilots occupy all the subcarriers; in the mixed-type scheme, only a fraction of the subcarriers are dedicated to pilots, and such pilot OFDM symbols are interleaved with all-data OFDM symbols in the time domain.

A BLUE channel estimator is adopted, which reduces to an LS channel estimator in the case of the block-type scheme. The existence of a channel estimate (channel identifiability) was discussed, which turns out to be dependent on the pattern in which pilots are distributed in the time and frequency domain.

Finally, via numerical examples, we have shown that the pilot schemes also have an impact on the channel estimation performance. In short, the block-type scheme, despite the fact that it allows for a simple channel estimator, is not suitable for channels that vary fast. The performances of the comb-type and mixed-type schemes are in general similar if the number of channel unknowns is limited, but the comb-

type scheme has a slight advantage in the precision. However, when the number of channel unknowns becomes large, the comb-type is incapable of estimating a long channel. In contrast, the mixed-type scheme is able to render a robust performance for a wide range of channel situations.

Appendix 5A: Rank Condition of $\mathbf{Z}(i)$

In this section, we will look at the rank condition of $\mathbf{Z}(i)$, which is considered in Section 5.4.1. Before starting, let us introduce the following facts. First, due to the fact that the set $\bar{\mathcal{T}}(i)$ contains equi-distant elements [c.f. (5.34)], the entries of $\mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}}$ can be expressed as

$$[\mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}}]_{m,n} = e^{-j\frac{2\pi}{MP}(mM+\mu(i))n}, \quad (5.62)$$

where $\mu(i)$ denotes the position of the first subcarrier that carries a non-zero pilot. Hence, it is not hard to see in matrix/vector form that

$$\mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}} = \boldsymbol{\theta}^{\mu(i)T} \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i)}\}, \quad (5.63)$$

with

$$\boldsymbol{\theta} := e^{-j\frac{2\pi}{M}[0, \dots, M-1]^T}, \quad (5.64)$$

$$\boldsymbol{\delta} := e^{-j\frac{2\pi}{MP}[0, \dots, P-1]^T}. \quad (5.65)$$

Second, in compliance with the equi-spaced clustered pilots, we are allowed as well to group the observation samples into P equi-distant clusters with each cluster containing $K + 1$ observation samples in a row. Notice from Fig. 4.2 that the centers of these observation sample clusters align with the centers of the pilot clusters. Therefore, the set of the observation sample positions $\mathcal{O}(i)$ can be expressed as

$$\mathcal{O}(i) = \bar{\mathcal{T}}(i) \oplus \left\{ -\frac{K}{2}, \dots, \frac{K}{2} \right\}, \quad (5.66)$$

where the operator \oplus stands for the set sum: if a belongs to the set \mathcal{A} and b belongs to the set \mathcal{B} , then $a + b$ belongs to the set $\mathcal{A} \oplus \mathcal{B}$. From the definition, we understand that if we arrange the elements of $\mathcal{O}(i)$ in an ascendant order and group them into $K + 1$ subsets by picking the $(K + 1)$ -spaced elements, then the k th subset will contain

$$\bar{\mathcal{T}}(i) + k - \frac{K}{2} = \left\{ \mu(i) + k - \frac{K}{2}, \dots, \mu(i) + M(P - 1) + k - \frac{K}{2} \right\},$$

for $k = 0, \dots, K - 1$. Such a partitioning of $\mathcal{O}(i)$ enables us to attain the following relationship

$$\mathbf{\Pi}_{K+1} \mathbf{W}_N^{\{\mathcal{O}(i)\}} = \begin{bmatrix} \mathbf{W}_N^{\{\bar{\mathcal{T}}(i) - \frac{K}{2}\}} \\ \vdots \\ \mathbf{W}_N^{\{\bar{\mathcal{T}}(i) + \frac{K}{2}\}} \end{bmatrix}, \quad (5.67)$$

where $\mathbf{\Pi}_d$ denotes a depth- d interleaver matrix. It is such defined that when multiplied with a vector $[a_0, a_1, \dots]^T$, it produces

$$\mathbf{\Pi}_d [a_0, a_1, \dots]^T = [a_0, a_d, \dots, a_1, a_{d+1}, \dots]^T. \quad (5.68)$$

The matrix $\mathbf{W}_N^{\{\bar{\mathcal{T}}(i)+k\}}$ is related to the matrix $\mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}}$ as

$$\mathbf{W}_N^{\{\bar{\mathcal{T}}(i)+k\}} = \mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}} \mathbf{\Lambda}^k, \quad (5.69)$$

with $\mathbf{\Lambda}$ denoting the diagonal matrix

$$\begin{aligned} \mathbf{\Lambda} &= \mathcal{D}\{e^{-j \frac{2\pi}{MP} [0, \dots, MP-1]^T}\} \\ &= \mathcal{D}\{\boldsymbol{\theta}\} \otimes \mathcal{D}\{\boldsymbol{\delta}\}. \end{aligned}$$

Using (5.63) along with the fact that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, we can derive that

$$\begin{aligned} \mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}} \mathbf{\Lambda}^k &= (\boldsymbol{\theta}^{\mu(i)T} \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i)}\}) (\mathcal{D}\{\boldsymbol{\theta}^k\} \otimes \mathcal{D}\{\boldsymbol{\delta}^k\}) \\ &= \boldsymbol{\theta}^{(\mu(i)+k)T} \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i)+k}\}, \end{aligned}$$

and thus the matrix $\mathbf{\Pi}_{K+1} \mathbf{W}_N^{\{\mathcal{O}(i)\}}$ in (5.67) can be further rewritten as

$$\mathbf{\Pi}_{K+1} \mathbf{W}_N^{\{\mathcal{O}(i)\}} = \begin{bmatrix} \mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}} \mathbf{\Lambda}^{-\frac{K}{2}} \\ \vdots \\ \mathbf{W}_N^{\{\bar{\mathcal{T}}(i)\}} \mathbf{\Lambda}^{\frac{K}{2}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(\mu(i) - \frac{K}{2})T} \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i) - \frac{K}{2}}\} \\ \vdots \\ \boldsymbol{\theta}^{(\mu(i) + \frac{K}{2})T} \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i) + \frac{K}{2}}\} \end{bmatrix}.$$

Taking both the above and (5.63) into account, we can show that the matrix $\mathbf{Z}(i)$ defined in (5.36), if permuted by $\mathbf{\Pi}_{K+1}$, will admit the following expression:

$$\begin{aligned} \mathbf{\Pi}_{K+1} \mathbf{Z}(i) &= \begin{bmatrix} \boldsymbol{\theta}^{(\mu(i) - \frac{K}{2})T} \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i) - \frac{K}{2}}\} \\ \vdots \\ \boldsymbol{\theta}^{(\mu(i) + \frac{K}{2})T} \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i) + \frac{K}{2}}\} \end{bmatrix} \times \\ &[\mathcal{D}\{\mathbf{b}_0(i)\} \quad \dots \quad \mathcal{D}\{\mathbf{b}_Q(i)\}] (\mathbf{I}_{Q+1} \otimes \boldsymbol{\theta}^{-\mu(i)} \otimes \mathcal{D}\{\boldsymbol{\delta}^{-\mu(i)}\} \mathbf{W}_P^H). \end{aligned} \quad (5.70)$$

Actually, $\mathbf{\Pi}_{K+1}\mathbf{Z}(i)$ is a block matrix comprised of $(K + 1)$ by $(Q + 1)$ square sub-matrices, each of size P . The (k, q) th sub-matrix $[\mathbf{\Pi}_{K+1}\mathbf{Z}(i)]_{kP:(k+1)P-1, qP:(q+1)P-1}$ can be expressed as

$$\begin{aligned} & [\mathbf{\Pi}_{K+1}\mathbf{Z}(i)]_{kP:(k+1)P-1, qP:(q+1)P-1} \\ &= (\boldsymbol{\theta}^{\mu(i)+k-\frac{K}{2}})^T \otimes \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i)+k-\frac{K}{2}}\} \mathcal{D}\{\mathbf{b}_q(i)\} (\boldsymbol{\theta}^{-\mu(i)} \otimes \mathcal{D}\{\boldsymbol{\delta}^{-\mu(i)}\} \mathbf{W}_P^H), \end{aligned} \quad (5.71)$$

for $k = 0, \dots, K$ and $q = 0, \dots, Q$. Because $\boldsymbol{\theta}$ is an $M \times 1$ vector [c.f. (5.64)], to work the above equality further out, we can likewise partition the q th expansion sequence that corresponds to the i th OFDM symbol $\mathbf{b}_q(i)$ into M sub-vectors

$$\mathbf{b}_q(i) = [\mathbf{b}_{q,0}^T(i), \dots, \mathbf{b}_{q,M-1}^T(i)]^T, \quad (5.72)$$

with each sub-vector containing P entries, i.e., for the m th sub-vector

$$\mathbf{b}_{q,m}(i) := [[\mathbf{b}_q(i)]_{mP}, \dots, [\mathbf{b}_q(i)]_{(m+1)P-1}]^T. \quad (5.73)$$

Accordingly, (5.71) can be expressed as

$$\begin{aligned} & [\mathbf{\Pi}_{K+1}\mathbf{Z}(i)]_{kN:(k+1)N-1, qN:(q+1)N-1} \\ &= \sum_{m=0}^{M-1} [\boldsymbol{\theta}^{\mu(i)+k-\frac{K}{2}}]_m \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{\mu(i)+k-\frac{K}{2}}\} \mathcal{D}\{\mathbf{b}_{q,m}(i)\} [\boldsymbol{\theta}^{-\mu(i)}]_m \mathcal{D}\{\boldsymbol{\delta}^{-\mu(i)}\} \mathbf{W}_P^H \\ &= \mathbf{W}_P \sum_{m=0}^{M-1} [\boldsymbol{\theta}^{k-\frac{K}{2}}]_m \mathcal{D}\{\mathbf{b}_{q,m}(i)\} \mathcal{D}\{\boldsymbol{\delta}^{k-\frac{K}{2}}\} \mathbf{W}_P^H \\ &= \mathbf{W}_P \mathcal{D}\{\boldsymbol{\delta}^{k-\frac{K}{2}}\} \mathcal{D}\{\boldsymbol{\Xi}_q(i)\boldsymbol{\theta}^{k-\frac{K}{2}}\} \mathbf{W}_P^H, \end{aligned} \quad (5.74)$$

with $\boldsymbol{\Xi}_q(i)$ defined as an $P \times M$ matrix

$$\boldsymbol{\Xi}_q(i) := [\mathbf{b}_{q,0}(i), \dots, \mathbf{b}_{q,M-1}(i)]. \quad (5.75)$$

We observe that the index $\mu(i)$ drops out from (5.74). As can be understood that the positions of equi-distant pilots are uniquely characterized by their starting position $\mu(i)$, this implies that $[\mathbf{\Pi}_{K+1}^T \mathbf{Z}(i)]_{k,q}$ (and thus $\mathbf{Z}(i)$ as well) is independent of the pilot positions.

In line with (5.74), $\mathbf{Z}(i)$ can be expressed as [c.f. (5.70)]

$$\mathbf{Z}(i) = \mathbf{\Pi}_{K+1}^T (\mathbf{I}_{K+1} \otimes \mathbf{W}_P) \mathcal{D}\{[\boldsymbol{\delta}^{-\frac{K}{2}T}, \dots, \boldsymbol{\delta}^{\frac{K}{2}T}]^T\} \mathbf{X}(i) (\mathbf{I}_{Q+1} \otimes \mathbf{W}_P^H), \quad (5.76)$$

with

$$\mathbf{X}(i) := \begin{bmatrix} \mathcal{D}\{\boldsymbol{\Xi}_0(i)\boldsymbol{\theta}^{-\frac{K}{2}}\} & \dots & \mathcal{D}\{\boldsymbol{\Xi}_Q(i)\boldsymbol{\theta}^{-\frac{K}{2}}\} \\ \vdots & \ddots & \vdots \\ \mathcal{D}\{\boldsymbol{\Xi}_0(i)\boldsymbol{\theta}^{\frac{K}{2}}\} & \dots & \mathcal{D}\{\boldsymbol{\Xi}_Q(i)\boldsymbol{\theta}^{\frac{K}{2}}\} \end{bmatrix}. \quad (5.77)$$

Apparently, $\mathbf{X}(i)$ determines the rank of $\mathbf{Z}(i)$. Observing that $\mathbf{X}(i)$ is a block matrix consisting of $K + 1$ by $Q + 1$ diagonal sub-matrices, each of dimension $P \times P$, we transform it into a block diagonal matrix to attain a better perception of its rank condition:

$$\mathbf{\Pi}_P \mathbf{X}(i) \bar{\mathbf{\Pi}}_P^T = \begin{bmatrix} \tilde{\mathbf{X}}_0(i) & & \\ & \ddots & \\ & & \tilde{\mathbf{X}}_{P-1}(i) \end{bmatrix}, \quad (5.78)$$

where $\mathbf{\Pi}_P$ and $\bar{\mathbf{\Pi}}_P$ are both depth- P interleave matrices with appropriate dimensions. $\tilde{\mathbf{X}}_p(i)$ is a $(K + 1) \times (Q + 1)$ matrix, which can be expressed after some algebra as

$$\tilde{\mathbf{X}}_p(i) = \mathbf{\Theta}^T \bar{\mathbf{B}}_p(i), \quad (5.79)$$

with

$$\mathbf{\Theta} := \begin{bmatrix} \theta^{-\frac{K}{2}} & \dots & \theta^{\frac{K}{2}} \end{bmatrix},$$

$$\bar{\mathbf{B}}_p(i) := [\Xi_0^T(i) \mathbf{e}_p \quad \dots \quad \Xi_{Q+1}^T(i) \mathbf{e}_p].$$

As the transpose of a stack of all the p th rows from the matrices $\Xi_q(i)$'s for $q = 0, \dots, Q$, the $M \times (Q + 1)$ matrix $\bar{\mathbf{B}}_p(i)$ is actually comprised of M equi-distant rows of the i th BEM matrix $\mathbf{B}(i)$, i.e.,

$$\bar{\mathbf{B}}_p(i) = [\mathbf{B}(i)^T \mathbf{e}_p, \dots, \mathbf{B}(i)^T \mathbf{e}_{p+(M-1)P}]^T. \quad (5.80)$$

This means for the CE-BEM or the P-BEM, $\bar{\mathbf{B}}_p(i)$ is a Vandermonde matrix, and thus

$$\text{Rank}\{\bar{\mathbf{B}}_p(i)\} = \min(M, Q + 1). \quad (5.81)$$

For the other BEM definitions, the above holds in general true as well, though we cannot rigorously prove it.

Due to Assumption 5.1 and 5.2, we understand that

$$P(K + 1) = PM = N. \quad (5.82)$$

Consequently, the $M \times (K + 1)$ matrix $\mathbf{\Theta}$ becomes a unitary M -point DFT matrix, and the rank of $\bar{\mathbf{B}}_p(i)$ is therefore also the rank of $\tilde{\mathbf{X}}_p(i)$ [‡].

Since $\tilde{\mathbf{X}}_p(i)$ determines the rank of $\mathbf{X}(i)$ in (5.77), and thus that of $\mathbf{Z}(i)$ in (5.76) as well, we reach the conclusion in (5.38).

[‡]A special exception is reported in [47], which does not adopt Assumption 5.1 and chooses $K + 1 = Q + 1 < M$. However, because it relies on a critical CE-BEM assumption, it can be shown that $\tilde{\mathbf{X}}_p(i) = \mathbf{I}_{Q+1}$.

Appendix 5B: Proof of Lemma 5.1

Due to the identical pilot assumption, we have

$$\mathbf{U}(i_0) = \cdots = \mathbf{U}(i_{V-1}), \quad (5.83)$$

and therefore

$$\begin{aligned} \mathcal{P} &= [\mathcal{P}^T(i_0), \dots, \mathcal{P}^T(i_{V-1})]^T \\ &= \mathbf{Z}\mathbf{U}(i), \end{aligned} \quad (5.84)$$

where

$$\mathbf{Z} := [\mathbf{Z}^T(i_0), \dots, \mathbf{Z}^T(i_{V-1})]^T, \quad (5.85)$$

which is thus an $PV(K+1) \times P(Q+1)$ matrix. Since we have already shown that the $P(Q+1) \times (L+1)(Q+1)$ matrix $\mathbf{U}(i)$ has full rank $(L+1)(Q+1)$ if $P \geq L+1$, to prove the full-rank condition of \mathcal{P} , it is sufficient to prove that \mathbf{Z} has a rank $P(Q+1)$. Following the similar steps as given in Appendix 5A, we can show that the rank of \mathbf{Z} is a V -multiple of the rank of $\tilde{\mathbf{X}}_p(i)$ that is defined in (5.79). In other words,

$$\begin{aligned} \tilde{\mathbf{X}}_p &:= [\tilde{\mathbf{X}}_p^T(i_0), \dots, \tilde{\mathbf{X}}_p^T(i_{V-1})]^T \\ &= (\mathbf{I}_V \otimes \Theta^T) [\bar{\mathbf{B}}_p^T(i_0), \dots, \bar{\mathbf{B}}_p^T(i_{V-1})]^T. \end{aligned} \quad (5.86)$$

As a result, we understand that [c.f. (5.81)]

$$\text{Rank}\{\tilde{\mathbf{X}}_p\} = \min(VM, Q+1) = \min(V(K+1), Q+1). \quad (5.87)$$

The rank of \mathbf{Z} will therefore be $P(Q+1)$ if $V(K+1) \geq Q+1$, which concludes the proof.

Appendix 5C: Proof of Lemma 5.2

In the case that the positions of the non-zero pilots in each pilot OFDM symbol are not aligned with each other, we have

$$\mathcal{P} = \mathbf{Z}\mathbf{U}, \quad (5.88)$$

with

$$\mathbf{Z} := \begin{bmatrix} \mathbf{Z}(i_1) & & \\ & \ddots & \\ & & \mathbf{Z}(i_V) \end{bmatrix}, \mathbf{U} := \begin{bmatrix} \mathbf{U}(i_1) \\ \vdots \\ \mathbf{U}(i_V) \end{bmatrix}.$$

Obviously, with the equality $K + 1 = M$, \mathbf{Z} is of full column-rank if $K + 1 \geq Q + 1$. Therefore, we only need to examine the rank of \mathbf{U} . By adopting a depth- $P(Q + 1)$ interleaver matrix $\mathbf{\Pi}_{P(Q+1)}$, we are able to show that

$$\mathbf{\Pi}_{P(Q+1)}\mathbf{U} = \mathbf{I}_{Q+1} \otimes \begin{bmatrix} \mathcal{D}\{\bar{\mathbf{t}}(i_1)\} & & \\ & \ddots & \\ & & \mathcal{D}\{\bar{\mathbf{t}}(i_V)\} \end{bmatrix} \underbrace{\begin{bmatrix} \mathbf{G}_L^{\{\bar{T}(i_1)\}} \\ \vdots \\ \mathbf{G}_L^{\{\bar{T}(i_V)\}} \end{bmatrix}}_{\bar{\mathbf{G}}_L}. \quad (5.89)$$

Because the positions of the VP non-zero pilots in each OFDM symbol are completely different to each other, accordingly, the $VP \times (L + 1)$ matrix $\bar{\mathbf{G}}_L$ consists of VP different rows of the Vandermonde matrix \mathbf{G}_L , it has a full rank $L + 1$ if $VP \geq (L + 1)$, which concludes the proof.

Chapter 6

Channel Equalization in a Single-Carrier System

6.1 Introduction

In a single-carrier transmission system over a channel with a large delay spread, it is more efficient to equalize the channel in the frequency domain utilizing a simple one-tap equalizer [30]. The underlying consideration is that the frequency-domain (FD) channel is diagonal if

1. the channel stays invariant during at least one block;
2. the inter-block interference (IBI) can be completely annihilated by inserting a sufficiently long guard interval (GI).

However, these two conditions are not always satisfied in practice, and as a result, the orthogonality among the subcarriers will be destroyed and the FD channel becomes actually a full matrix. A reliable FD equalizer for such a channel will be much more expensive, which is the key issue considered in this chapter.

To restore the orthogonality among the subcarriers, pre-processing at the receiver is indispensable. For instance, a channel shortening technique, in the form of an FIR filter, is proposed in [3] for time-invariant channels, with the aim of fitting the effective channel within the given guard interval. In [7], an FIR filter is adopted to “flatten” the channel fluctuation, which can be considered as the dual of channel shortening. In [8], both schemes are combined. These techniques work well for channels that are moderately spread in delay and Doppler domains.

Often, a strictly diagonal FD channel matrix is too difficult to achieve. In a realistic transmission system, the Doppler-induced channel has most of its power concentrated in the vicinity of the diagonal in a circular sense, with those entries that are far away from the diagonal reducing fast [96, 14]. This implies that it is more practical to assume a banded FD channel matrix. Many equalizers exploit this banded (rather than diagonal) structure to lower the complexity for single-carrier systems

The results of this chapter appeared in [99, 100].

as well as OFDM, e.g., the block zero-forcing (ZF) equalizer in [46], the block minimum mean square error (MMSE) equalizer in [81, 82], the serial iterative MMSE equalizer in [90, 92], the maximum likelihood (ML) equalizer in [71, 49], etc. It can be imagined that to enhance the equalization performance, especially at a moderate to high SNR, the band approximation error must be reduced as much as possible. One solution can be the FIR filter of [8], but it generally requires a multiple antenna assumption and can still be too complicated. Since we need not to enforce a diagonal channel matrix but a banded one, a reduced-order FIR filter with just a single tap could be adequate. Such a filter is referred to as a receiver window in [92, 90, 82, 44]. The windowing operation on the received signals is thus a time-domain point-wise multiplication with the time-varying coefficients of the window.

In this chapter, we will present two receiver architectures in combination with the windowing technique to counteract the channel time-variation as well as the IBI for a single-carrier transmission system. The first receiver will be based on the original data model (ODM), which describes the actual channel input/output (I/O) relationship. The second receiver will be based on the so-called extended data model (EDM), which extends the ODM to a larger scale without compromising its validity. The advantage of the EDM is that by introducing redundancy at the receiver, we are endowed with some extra design freedom to explore the Doppler resolution better. In both data models, the full FD channel matrix will be approximated by a strictly banded matrix for the sake of complexity.

Note that the band approximation procedure taken in this chapter will be similar to that in Chapter 3, where a strictly banded matrix is devised that is close to its full counterpart only in terms of the Frobenius norm. These two matrices will therefore not necessarily share common diagonals, which is the traditional approach in [92, 90, 82, 44]. This is especially true in the EDM case. A more profound rationale is that by this means we have translated the band approximation error in the frequency domain into a basis expansion modeling error in the time domain. More specifically, we can show that the band approximation error in the ODM corresponds to a modeling error resulting from the (C)CE-BEM*, while the band approximation error in the EDM corresponds to a modeling error resulting from the (O)CE-BEM. This idea will be reflected in our window design. Since the (O)CE-BEM in general can yield a much tighter fit to a realistic TV channel than the (C)CE-BEM, it is not hard to understand that the equalizer for the EDM will be subject to a much smaller band approximation error than for the ODM, and is thus able to yield a better performance.

*As a matter of fact, such a link also underlies the equalizer design in [92], but is not straightforward to observe.

6.2 System Model

Let us consider a discrete communication system, where the data symbols are directly transmitted over the channel. We assume that the channel to be an FIR filter with order L , i.e., if we use $h_{p,l}$ to denote the l th channel tap at the p th time instance then $h_{p,l} = 0$ if $l < 0$ or $l > L$. At the receiver, a window is deployed. Conform the FIR assumption of the channel, the observation sample at the p th time-instance y_p can be expressed as

$$y_p = w_p \sum_{l=0}^L h_{p,l} s_{p-l} + w_p v_p, \quad (6.1)$$

where w_p stands for the p th element of the window; s_p denotes the p th data symbol, and v_p the corresponding noise prior to windowing.

Like in Chapter 3, the following assumptions will be used throughout the whole chapter.

Assumption 6.1. We assume the channel to be a wide-sense stationary uncorrelated scattering (WSSUS) process, for which

$$\mathcal{E}_h \{h_{p,l} h_{p-m,l-n}\} = \sigma_l^2 \gamma_m \delta_n. \quad (6.2)$$

Here, δ_n denotes the Kronecker delta, σ_l^2 the variance of the l th channel tap, and γ_m the normalized temporal correlation.

Assumption 6.2. We assume the data symbols to be zero-mean white with unit variance, and the noise to be zero-mean white with variance σ^2 i.e.,

$$\begin{aligned} \mathcal{E}_s \{s_p s_{p-m}^*\} &= \delta_m, \\ \mathcal{E}_v \{v_{t,p} v_{t,p-m}^*\} &= \sigma^2 \delta_m. \end{aligned} \quad (6.3)$$

6.3 FD Equalization Based on the ODM

6.3.1 Equalization Scheme

Suppose that we are interested in estimating the $N - L$ data symbols, which are stacked in the vector

$$\mathbf{s}_{N-L} := [s_0, \dots, s_{N-L-1}]^T. \quad (6.4)$$

Their information can be found at the receiver in the observation sample vector $\mathbf{y}_{t,N}$, which is collected at the 0th until $(N - 1)$ st time-instance,

$$\mathbf{y}_{t,N} := [y_0, \dots, y_{N-1}]^T. \quad (6.5)$$

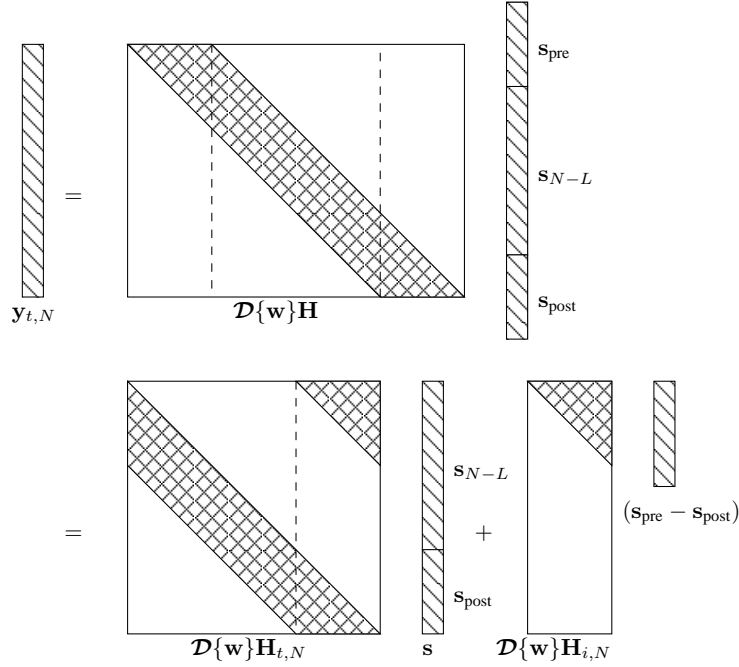


Figure 6.1: The noiseless original data model.

For these samples, the I/O relationship in (6.1) can be expressed in a block form as

$$\mathbf{y}_{t,N} = \mathcal{D}\{\mathbf{w}\}\mathbf{H}[\mathbf{s}_{\text{pre}}^T, \mathbf{s}_{N-L}^T, \mathbf{s}_{\text{post}}^T]^T + \mathcal{D}\{\mathbf{w}\}\mathbf{v}_{t,N}, \quad (6.6)$$

where

$$\begin{aligned} \mathbf{w} &:= [w_0, \dots, w_{N-1}]^T, \\ \mathbf{v}_{t,N} &:= [v_0, \dots, v_{N-1}]^T. \end{aligned} \quad (6.7)$$

\mathbf{s}_{pre} and \mathbf{s}_{post} represent the neighboring L data symbols, which come from the previous and next data blocks, respectively, and whose information is also carried in $\mathbf{y}_{t,N}$:

$$\begin{aligned} \mathbf{s}_{\text{pre}} &:= [s_{-L}, \dots, s_{-1}]^T, \\ \mathbf{s}_{\text{post}} &:= [s_{N-L}, \dots, s_{N-1}]^T. \end{aligned} \quad (6.8)$$

The $N \times (N + L)$ matrix \mathbf{H} stands for the convolutive channel with

$$[\mathbf{H}]_{p,n} := h_{p,p-n+L}. \quad (6.9)$$

The I/O relationship in (6.6) in the noiseless case is illustrated in the upper panel of Fig. 6.1, where one can observe that the channel matrix $\mathcal{D}\{\mathbf{w}\}\mathbf{H}$ is horizontally divided in three parts. The first part corresponds to the data symbols from the previous block \mathbf{s}_{pre} , which are put aside in a separate term in the lower panel of Fig. 6.1. The remaining parts of $\mathcal{D}\{\mathbf{w}\}\mathbf{H}$, which are square, are made “circulant” in the lower panel of Fig. 6.1 by adding a triangular from the first part. In mathematics, this means that we can rewrite (6.6) in the following form

$$\mathbf{y}_{t,N} = \mathcal{D}\{\mathbf{w}\}\mathbf{H}_{t,N}\mathbf{s} + \boldsymbol{\epsilon}_{t,N} + \mathcal{D}\{\mathbf{w}\}\mathbf{v}_{t,N}, \quad (6.10)$$

where

$$\mathbf{s} := [\mathbf{s}_{N-L}^T, \mathbf{s}_{\text{post}}^T]^T, \quad (6.11)$$

and the $N \times N$ matrix $\mathbf{H}_{t,N}$ has entries

$$[\mathbf{H}_{t,N}]_{p,n} := h_{p, \text{mod}(p-n, N)}, \quad (6.12)$$

where the operation $\text{mod}(a, b)$ in the subscript corresponds to the “circulant” nature of $\mathbf{H}_{t,N}$. In light of (6.10), we can easily observe that $\boldsymbol{\epsilon}_N$ represents the IBI, which is formed by

$$\boldsymbol{\epsilon}_{t,N} := \mathcal{D}\{\mathbf{w}\}\mathbf{H}_{i,N}(\mathbf{s}_{\text{pre}} - \mathbf{s}_{\text{post}}). \quad (6.13)$$

Here, $\mathbf{H}_{i,N}$ is an $N \times L$ matrix with entries

$$[\mathbf{H}_{i,N}]_{p,n} := h_{p, p-n+L}. \quad (6.14)$$

The reshaped I/O relationship in the noiseless case is illustrated in the lower panel of Fig. 6.1. Transformed into the frequency domain, (6.10) becomes

$$\begin{aligned} \mathbf{y}_{f,N} &:= \mathbf{W}_N \mathbf{y}_{t,N}, \\ &= \mathbf{H}_{f,N} \mathbf{W}_N \mathbf{s} + \boldsymbol{\epsilon}_{f,N} + \mathbf{v}_{f,N}, \end{aligned} \quad (6.15)$$

where

$$\begin{aligned} \boldsymbol{\epsilon}_{f,N} &:= \mathbf{W}_N \boldsymbol{\epsilon}_{t,N}, \\ \mathbf{v}_{f,N} &:= \mathbf{W}_N \mathcal{D}\{\mathbf{w}\}\mathbf{v}_{t,N}. \end{aligned} \quad (6.16)$$

Further,

$$\mathbf{H}_{f,N} := \mathbf{W}_N \mathcal{D}\{\mathbf{w}\}\mathbf{H}_{t,N} \mathbf{W}_N^H \quad (6.17)$$

stands for the FD channel matrix. Since $\mathcal{D}\{\mathbf{w}\}\mathbf{H}_{t,N}$ is not really circulant due to the channel time-variation, $\mathbf{H}_{f,N}$ is not diagonal.

In (6.15), except for the DFT, we did not apply other processing on the received samples, and the data model is identical to what has actually happened in reality.

We call such a data model the original data model (ODM). Typical to the ODM is that the size of the DFT N equals the number of observation samples. This will be in contrast with the EDM scheme discussed in the next section, where the size of the DFT is larger than the number of observation samples.

The IBI in the frequency domain $\epsilon_{f,N}$ can be mitigated by the utility of a guard interval of length L_z , e.g., a cyclic-prefix (CP), a zero-prefix (ZP)[118] or a non-zero prefix (NZP) [25]. In the CP case, we let

$$[s_{-L_z}, \dots, s_{-1}] = [s_{N-L}, \dots, s_{N-L+L_z-1}], \quad (6.18)$$

while in the ZP and NZP case, we let

$$[s_{-L_z}, \dots, s_{-1}]^T = [s_{N-L}, \dots, s_{N-L+L_z-1}]^T = \mathbf{p}, \quad (6.19)$$

with \mathbf{p} being a zero or non-zero pilot vector, respectively. For $L_z \geq L$, the IBI can be completely removed. In this chapter, we will focus on the scenario where L_z assumes an arbitrary value, which implies that the IBI could be still present. In that case, we need to design the window to minimize its impact. We will come back to this later on.

For the moment, let us ignore the IBI in the data model, and apply a block LMMSE equalizer on (6.15), which produces the data estimates as

$$\hat{\mathbf{s}}_{N-L} = \mathbf{Z}_{N-L}^T \mathbf{W}_N^H \mathbf{H}_{f,N}^H (\mathbf{H}_{f,N} \mathbf{H}_{f,N}^H + \mathbf{R}_{v,N})^{-1} \mathbf{y}_{f,N}, \quad (6.20)$$

with

$$\begin{aligned} \mathbf{Z}_{N-L} &:= [\mathbf{I}_{N-L}, \mathbf{0}_{(N-L) \times L}], \\ \mathbf{R}_{v,N} &:= \mathcal{E}_v\{\mathbf{v}_{f,N} \mathbf{v}_{f,N}^H\} = \sigma^2 \mathbf{W}_N \mathcal{D}\{\mathbf{w} \odot \mathbf{w}^*\} \mathbf{W}_N^H. \end{aligned} \quad (6.21)$$

Here, we come across the same problem as in OFDM: now that the FD channel matrix $\mathbf{H}_{f,N}$ is full, most of the computational effort in (6.20) is invested in inverting the covariance matrix. Like in OFDM, we will approximate $\mathbf{H}_{f,N}$ with a strictly banded matrix $\hat{\mathbf{H}}_{f,N}$, which has only non-zero entries on the main diagonal, the $Q/2$ super- and $Q/2$ sub-diagonals in a circular sense. As a result, the LMMSE equalizer becomes

$$\hat{\mathbf{s}}_{N-L} = \mathbf{Z}_{N-L}^T \mathbf{W}_N^H \hat{\mathbf{H}}_{f,N}^H (\hat{\mathbf{H}}_{f,N} \hat{\mathbf{H}}_{f,N}^H + \mathbf{R}_{v,N})^{-1} \mathbf{y}_{f,N}, \quad (6.22)$$

In the above, Q is a design parameter that can be chosen as a trade-off between complexity and performance. The resulting equalizer can be made cheaper with a smaller Q , but accordingly suffers from more out-of-band interference. In reality, its minimum value is often related with the maximal Doppler spread.

It is noteworthy that the banded matrix $\hat{\mathbf{H}}_{f,N}$ is not simply obtained by taking the $Q+1$ most significant diagonals of $\mathbf{H}_{f,N}$. Like in Chapter 3, we use the following cost function to design the non-zero diagonals of $\hat{\mathbf{H}}_{f,N}$,

$$\begin{aligned} & \arg \min_{\{\hat{\mathbf{H}}_{f,N}\}} \|\mathbf{H}_{f,N} - \hat{\mathbf{H}}_{f,N}\|^2, \\ & \text{s.t. } \hat{\mathbf{H}}_{f,N} = \hat{\mathbf{H}}_{f,N} \odot \mathbf{T}_Q. \end{aligned} \quad (6.23)$$

where \mathbf{T}_Q is a matrix of proper dimensions, which has ones on the main diagonal, the $Q/2$ super- and $Q/2$ sub-diagonals, and zeros on the remaining entries. It will be shown later on that the above cost function is related to the choice of the window \mathbf{w} .

Another important feature in (6.22) is that we also require the noise covariance matrix $\mathbf{R}_{v,N}$ to be strictly banded with a bandwidth $2Q+1$, just like the product $\hat{\mathbf{H}}_{f,N} \hat{\mathbf{H}}_{f,N}^H$. If this is achieved as we will see later on, we can apply the Cholesky factorization [37] on the covariance matrix in (6.20) such that

$$\hat{\mathbf{H}}_{f,N} \hat{\mathbf{H}}_{f,N}^H + \mathbf{R}_{v,N} = \mathbf{G}_N \mathbf{G}_N^H, \quad (6.24)$$

where the upper-triangular matrix \mathbf{G}_N will assume a sparse ‘‘V-shape’’ structure as in Fig. 3.2. By this means, applying the inverse of $\hat{\mathbf{H}}_{f,N} \hat{\mathbf{H}}_{f,N}^H + \mathbf{R}_{v,N}$ can be implemented by applying the inverses of \mathbf{G}_N and \mathbf{G}_N^H separately using, e.g., Gaussian elimination.

For an overview of the total invoked complexity, we make a list of each step that is involved in computing (6.22) in the left panel of Table 6.1.

Utilizing the efficient FFT, we can show that step 1 requires about $N \log 2N$ MACs [75]. If we take the band structure of $\hat{\mathbf{H}}_{f,N}$ into account, step 2 requires $N(Q+1)(1.5Q+1)$ MACs. Step 3 can be realized by applying the Cholesky factorization, which entails $N(\frac{7}{8}Q^2 + \frac{13}{4}Q + 2) - \frac{1}{6}Q^3 - \frac{3}{2}Q^2 + \frac{5}{6}Q$ MACs [44]. Due to the sparse ‘‘V-shape’’ structure of \mathbf{G}_N , the Gaussian elimination required in step 4 and step 5 need $N(2Q+1) - 2Q^2 - Q$ MACs each. Step 6 needs another $N(2Q+1)$ MACs due to the band feature. Like step 1, Step 7 uses IFFT, which requires roughly $N \log 2N$ MACs as well. In summary, the considered block LMMSE equalizer has a complexity of $\mathcal{O}(NQ^2)$, which is linear in N and square in Q .

It is worth mentioning that in the single-carrier system, the channel can also be equalized in the time domain. For instance, we can apply a block LMMSE equalizer directly on (6.6), for which the complexity can be shown to be $\mathcal{O}(NL^2)$. Roughly, it is more appealing to equalize the channel in the frequency domain if the channel delay spread is much longer than the Doppler spread $L \gg Q$.

The equalization procedures for the ODM are, to a great extend, analogous to the equalizers discussed for OFDM in Chapter 3. Indeed, the single-carrier system

Table 6.1: Block MMSE equalization in the single-carrier system.

Steps in (6.22)	MACs required per step
1. $\mathbf{y}_{f,N} = \mathbf{W}_N \mathbf{y}_{t,N}$;	$N \log 2N$
2. $\mathbf{M}_N = \hat{\mathbf{H}}_{f,N} \hat{\mathbf{H}}_{f,N}^H + \mathbf{R}_{v,N}$;	$N(Q+1)(1.5Q+1)$
3. $\mathbf{M}_N = \mathbf{G}_N \mathbf{G}_N^H$;	$N(\frac{7}{8}Q^2 + \frac{13}{4}Q + 2) - \frac{1}{6}Q^3 - \frac{3}{2}Q^2 + \frac{5}{6}Q$
4. $\boldsymbol{\rho}_N = \mathbf{G}_N^{-1} \mathbf{y}_{f,K}$;	$N(2Q+1) - 2Q^2 - Q$
5. $\mathbf{d}_N = \mathbf{G}_N^{-H} \boldsymbol{\rho}_N$;	$N(2Q+1) - 2Q^2 - Q$
6. $\mathbf{x}_N := \hat{\mathbf{H}}_{f,N}^H \mathbf{d}_N$;	$N(2Q+1)$
7. $\hat{\mathbf{s}}_{N-L} = \mathbf{Z}_{N-L}^T \mathbf{W}_N^H \mathbf{x}_N$.	$N \log 2N$

can be viewed as a precoded OFDM system in the frequency domain. The existence of the precoder, which is the DFT matrix in this case, brings only a little increase in complexity. The other difference here is the presence of IBI.

Hence, in order to enhance the equalization performance of the ODM, we need to design the window and the banded matrix $\hat{\mathbf{H}}_{f,N}$ such that the IBI $\|\boldsymbol{\epsilon}_{f,N}\|$ as well as the band approximation error $\|\mathbf{H}_{f,N} - \hat{\mathbf{H}}_{f,N}\|$ will be minimized in an average sense. In addition, the window should also be able to make the noise covariance matrix $\mathbf{R}_{v,N}$ strictly banded. These will be discussed next.

6.3.2 Window Design for the ODM

Due to the analogy with OFDM, the window design for the ODM will in the most part follow the same procedures described in Section 3.4. For the sake of self-completeness, the major results will be briefly reiterated here.

First, regarding the noise-shaping behavior, we follow the guideline of Proposition 3.1, and express the window \mathbf{w} as a weighted sum of $Q+1$ complex exponential series:

$$\mathbf{w} = \mathbf{B}_N \mathbf{d}, \quad (6.25)$$

where \mathbf{B}_N corresponds to the (C)CE-BEM matrix with scale $Q+1$, which is comprised of the first $Q/2+1$ and the last $Q/2$ columns of \mathbf{W}_N , and \mathbf{d} is a $(Q+1)$ -long vector containing all the weighting coefficients.

Second, we have established in Theorem 3.1 the fact that the band approximation error between $\mathbf{H}_{f,N}$ and $\hat{\mathbf{H}}_{f,N}$ can be transformed into the (C)CE-BEM modeling error. In mathematics,

$$\|\mathbf{H}_{f,N} - \hat{\mathbf{H}}_{f,N}\| = \|\mathcal{D}\{\mathbf{w}\}\mathcal{H} - \mathbf{B}_N\mathcal{C}\|, \quad (6.26)$$

where \mathcal{H} stands for an $N \times (L + 1)$ matrix with entries $[\mathcal{H}]_{n,l} = h_{n,l}$, and \mathcal{C} for a $(Q + 1) \times (L + 1)$ matrix with entries $[\mathcal{C}]_{q,l} = c_{q,l}$. Here, $c_{q,l}$ represents some coefficient, whose value needs to be determined.

Except for the above results that have already been established in Section 3.4, we introduce the following theorem with regard to the minimization of IBI (see Appendix 6A for a proof).

Theorem 6.1. *The average power of the IBI $\epsilon_{f,N}$ in the ODM is a function of the window as*

$$\mathcal{E}_{h,s}\{\|\epsilon_{f,N}\|^2\} = 2\mathbf{w}^T \mathbf{R}_{\epsilon,N} \mathbf{w}^*, \quad (6.27)$$

where $\mathbf{R}_{\epsilon,N}$ denotes a diagonal matrix with the n th diagonal entry given by

$$[\mathbf{R}_{\epsilon,N}]_{n,n} = \begin{cases} \sum_{l=n+L_z+1}^L \sigma_l^2 & \text{if } n \leq L - L_z - 1, \\ 0 & \text{otherwise.} \end{cases} \quad (6.28)$$

To minimize both the band approximation error in (6.26) and IBI in (6.27), we formulate the following minimization problem

$$\begin{aligned} \min_{\{\mathbf{w}\}} \mathcal{E}_h \left\{ \min_{\{\mathcal{C}\}} \|\mathcal{D}\{\mathbf{w}\}\mathcal{H} - \mathbf{B}_N\mathcal{C}\|^2 \right\} + \theta \mathbf{w}^T \mathbf{R}_{\epsilon,N} \mathbf{w}^*, \\ \text{s.t. } \mathbf{w} = \mathbf{B}_N \mathbf{d} \text{ and } \|\mathbf{w}\|^2 = N. \end{aligned} \quad (6.29)$$

In the above, the first constraint is due to (6.25); the second constraint is imposed to avoid a trivial all-zero window. θ is a weight factor, whose function will be discussed later on.

We solve (6.29) first for \mathcal{C} :

$$\mathcal{C} = \mathbf{B}_N^\dagger \mathcal{D}\{\mathbf{w}\}\mathcal{H}. \quad (6.30)$$

Defining

$$\begin{aligned} \mathcal{P}_{\mathbf{B}_N} &:= \mathbf{I}_N - \mathbf{B}_N \mathbf{B}_N^\dagger, \\ \mathbf{R}_{\mathcal{H},N} &:= \mathcal{E}_h\{\mathcal{H}\mathcal{H}^H\}, \end{aligned} \quad (6.31)$$

we can show that the minimization problem then becomes

$$\begin{aligned} \min_{\{\mathbf{w}\}} \text{tr}(\mathcal{P}_{\mathbf{B}_N} \mathcal{D}\{\mathbf{w}\} \mathbf{R}_{\mathcal{H},N} \mathcal{D}\{\mathbf{w}^H\} \mathcal{P}_{\mathbf{B}_N}^H) + \theta \mathbf{w}^T \mathbf{R}_{\epsilon,N} \mathbf{w}^*, \\ \text{s.t. } \mathbf{w} = \mathbf{B}_N \mathbf{d} \text{ and } \|\mathbf{w}\|^2 = N. \end{aligned} \quad (6.32)$$

For the channel covariance matrix $\mathbf{R}_{\mathcal{H},N}$, we adopt Assumption 6.1, and hence its entries can be expressed as

$$[\mathbf{R}_{\mathcal{H},N}]_{m,n} = \sum_{l=0}^L \sigma_l^2 \gamma_{|m-n|}. \quad (6.33)$$

Applying the same trick as used in (3.29), we can derive that

$$\text{tr}(\mathcal{P}_{\mathbf{B}_N} \mathcal{D}\{\mathbf{w}\} \mathbf{R}_{\mathcal{H},N} \mathcal{D}\{\mathbf{w}^H\} \mathcal{P}_{\mathbf{B}_N}^H) = \mathbf{w}^T \left(\sum_{n=0}^{N-1} \mathcal{D}\{\mathbf{e}_n^T \mathcal{P}_{\mathbf{B}_N}\} \mathbf{R}_{\mathcal{H},N} \mathcal{D}\{\mathcal{P}_{\mathbf{B}_N}^H \mathbf{e}_n\} \right) \mathbf{w}^*. \quad (6.34)$$

Substituting (6.25) and (6.34) in (6.32) leads further to

$$\begin{aligned} \min_{\{\mathbf{w}\}} \mathbf{d}^T \mathcal{X}_N \mathbf{d}^*, \\ \text{s.t. } \|\mathbf{d}\|^2 = N. \end{aligned} \quad (6.35)$$

with

$$\mathcal{X}_N := \mathbf{B}_N^T \left(\sum_{n=0}^{N-1} \mathcal{D}\{\mathbf{e}_n^T \mathcal{P}_{\mathbf{B}_N}\} \mathbf{R}_{\mathcal{H},N} \mathcal{D}\{\mathcal{P}_{\mathbf{B}_N}^H \mathbf{e}_n\} + \theta \mathbf{R}_{\epsilon,N} \right) \mathbf{B}_N^*. \quad (6.36)$$

In the end, \mathbf{d} can be computed as the eigenvector corresponding to the least significant eigenvalue of \mathcal{X}_N^* .

Remarks 3.1, 3.2 and 3.3 that are made in Section 3.4 remain relevant here. Besides, the following remarks will be instructive for the next section.

Remark 6.1.

In Remark 3.1, we have established the equivalence between the band approximation error in the frequency domain with the (C)CE-BEM modeling error in the time domain. As a result, it is not difficult to understand that the bottleneck of the ODM, and hence that of the approach in Section 3.4 as well, is associated with a relatively poor modeling performance of the (C)CE-BEM as reported in [122]. Intuitively, this is because the (C)CE-BEM has a period equal to the block length N , and this implies that the (C)CE-BEM channel is N -periodic [111], which is not the case to the true TV channels. As a result, the (C)CE-BEM, which is in essence a truncated DFT, will oscillate at the edges and degrade the modeling performance (the Gibbs phenomenon). To counteract this effect, one solution is to extend the data model, which will be adopted in the next section. Another straightforward solution is to use windowing. Its impact on the BEM modeling performance will be better enlightened in the next remark.

Remark 6.2.

We have not addressed the weight factor θ in (6.29) yet. First of all, we realize from Fig. 3.3 (as well as from [90, 82]) that the optimal window that minimizes only the BEM modeling error ($\theta = 0$) is a curve that smoothly reduces at the edges. A similar idea is echoed in [51, p.p. 217], which reduces the channel at the edges to zero by subtraction. By this means, we are able to achieve a smoother transition at the edges of the channel and therefore reduce the (C)CE-BEM modeling error as proved in [51]. At the same time, such a window with very low power at the edges is also beneficial to suppress the IBI, which co-exists in the same area. The simulation will show that with the given channel parameters, choosing θ in the range of $[0, 1]$ will produce similar windows, and hence similar BER performances.

6.4 FD Equalization Based on the EDM

In the previous section, we have shown that the band approximation error can be translated into the modeling error between the realistic windowed channel and a corresponding (C)CE-BEM channel. While the (C)CE-BEM suffers from a relatively large modeling error, it is reported in [107, 53] that a more generalized form, the (O)CE-BEM, can yield a much better modeling performance [122]. This is achieved by simply enlarging the exponential period of the CE-BEM from N to K with $K > N$ [c.f. (2.33)]. However, the (O)CE-BEM cannot be straightforwardly applied to the ODM because the (O)CE-BEM channel matrix, if transformed to the N -grid frequency domain, will not be strictly banded.

Essentially, the (C)CE-BEM and (O)CE-BEM both use almost the same complex exponentials as basis expansion functions, except for the different exponential period. The (C)CE-BEM, with a period N , leads to a strictly banded matrix for the ODM that is of the same size N . Intuitively, one can imagine that the (O)CE-BEM, with a period K , would also lead to a strictly banded matrix if the size of the data model were also K . This means that we need to enlarge the data model.

6.4.1 Equalization Scheme

To derive a larger data model, let us first rewrite the ODM given in (6.10) in the form as

$$\mathbf{y}_{t,N} = \mathcal{D}\{\mathbf{w}\} \bar{\mathbf{H}}_N \mathbf{s}_{N-L} + \mathcal{D}\{\mathbf{w}\} \bar{\mathbf{H}}_{i,N} \begin{bmatrix} \mathbf{s}_{\text{pre}} \\ \mathbf{s}_{\text{post}} \end{bmatrix} + \mathcal{D}\{\mathbf{w}\} \mathbf{v}_{t,N}, \quad (6.37)$$

where $\bar{\mathbf{H}}_N$ is an $N \times (N - L)$ matrix with entries $[\bar{\mathbf{H}}_N]_{p,n} := h_{p,p-n}$, and $\bar{\mathbf{H}}_{i,N}$ stands for an $N \times 2L$ matrix constructed as

$$\bar{\mathbf{H}}_{i,N} := \begin{bmatrix} \mathbf{A} & \mathbf{0}_{L \times L} \\ \mathbf{0}_{(N-2L) \times 2L} & \\ \mathbf{0}_{L \times L} & \mathbf{B} \end{bmatrix}, \quad (6.38)$$

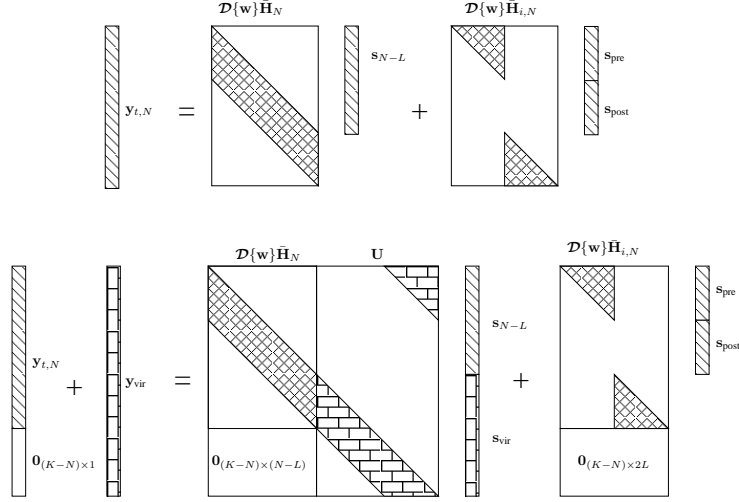


Figure 6.2: Upper: the noiseless original data model; below: the noiseless extended data model.

where the $L \times L$ matrix \mathbf{A} has entries $[\mathbf{A}]_{m,n} = h_{m,L-n+m}$, and the $L \times L$ matrix \mathbf{B} has entries $[\mathbf{B}]_{m,n} = h_{N-L+m,m-n}$.

The expression of (6.37) in the noiseless case is illustrated in the upper panel of Fig. 6.2. Comparing it with that of Fig. 6.1, we can observe that the channel matrix $\mathcal{D}\{\mathbf{w}\}\mathbf{H}$ in Fig. 6.1 is horizontally divided into three parts. The first and the last parts correspond to the IBI from \mathbf{s}_{pre} and \mathbf{s}_{post} , which are put together in Fig. 6.2 as a separate term $\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_{i,N}$. The remainder of $\mathcal{D}\{\mathbf{w}\}\mathbf{H}$ corresponding to \mathbf{s}_{N-L} is represented in Fig. 6.2 as $\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_N$.

We want to extend the data model in (6.37) to a scale K with $K \geq N + L$. First, let us append $K - N$ zeros to the end of $\mathbf{y}_{t,N}$. At the same time, we desire to create a $K \times K$ matrix $\bar{\mathbf{H}}_{t,K}$ out of the $N \times (N - L)$ matrix $\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_N$. To this end, we add $K - N$ all-zero rows at the bottom of $\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_N$, and then add a $K \times (K - N + L)$ matrix \mathbf{U} to its right side. The entries of \mathbf{U} remain to be determined. Because the channel matrix now has $K - N + L$ extra columns, accordingly, we append a $(K - N + L) \times 1$ vector \mathbf{s}_{vir} behind \mathbf{s}_{N-L} , whose values are subject to design as well. Further, we add $K - N$ all-zero rows at the bottom of the $N \times 2L$ matrix $\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_{i,N}$, and $K - N$ zeros at the bottom of the noise vector $\mathcal{D}\{\mathbf{w}\}\mathbf{v}_{t,N}$. The above steps can

be summarized in the following mathematical expression:

$$\underbrace{\begin{bmatrix} \mathbf{y}_{t,N} \\ \mathbf{0}_{(K-N) \times 1} \end{bmatrix}}_{\mathbf{y}_{t,K}} + \mathbf{y}_{\text{vir}} = \underbrace{\begin{bmatrix} \mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_N \\ \mathbf{0}_{(K-N) \times (N-L)} \end{bmatrix}}_{\mathbf{H}_{t,K}} \underbrace{\mathbf{U}}_{\mathbf{s}_K} \begin{bmatrix} \mathbf{s}_{N-L} \\ \mathbf{s}_{\text{vir}} \end{bmatrix} + \underbrace{\begin{bmatrix} \mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_{i,N} \\ \mathbf{0}_{(K-N) \times 2L} \end{bmatrix}}_{\boldsymbol{\epsilon}_{t,K}} \begin{bmatrix} \mathbf{s}_{\text{pre}} \\ \mathbf{s}_{\text{post}} \end{bmatrix} + \underbrace{\begin{bmatrix} \mathcal{D}\{\mathbf{w}\}\mathbf{v}_{t,N} \\ \mathbf{0}_{(K-N) \times 1} \end{bmatrix}}_{\mathbf{v}_{t,K}}, \quad (6.39)$$

where a $K \times 1$ vector \mathbf{y}_{vir} is added on the left side, which is set equal to

$$\mathbf{y}_{\text{vir}} = \mathbf{U}\mathbf{s}_{\text{vir}}, \quad (6.40)$$

To enable a clearer view, we plot the noiseless relationship in (6.39) in the lower panel of Fig. 6.2.

At this stage, it is worth underlining that we have introduced some redundancy only at the receiver! This operation is completely transparent to the transmitter, and hence the data rate is not compromised. A direct implication is that the ODM in (6.37) remains valid: it simply becomes a part of the larger data model in (6.39). For this reason, we will refer to the resulting channel I/O relationship as the extended data model (EDM).

In the EDM, besides the extra zeros, \mathbf{s}_{vir} and \mathbf{U} are added. \mathbf{s}_{vir} can be interpreted as some “virtual” data symbols, which are not really transmitted but included in (6.39) just because we want to make $\mathbf{H}_{t,K}$ to be a square “circulant” matrix. For the same reason, we desire that the elements of the $K \times (K - N + L)$ matrix \mathbf{U} be designed as

$$[\mathbf{U}]_{m,n} = \hat{h}_{m,\text{mod}(n-m-N+L,K)}, \quad (6.41)$$

where $\hat{h}_{m,l}$ stands for a virtual channel tap. We will discuss the specific value of $\hat{h}_{m,l}$ later on, but for the moment only assume that $\hat{h}_{m,l}$ equals zero if $l < 0$ or $l > L$. As a result, we can see that $\mathbf{H}_{t,K}$ indeed takes on a “circulant” form.

The second term on the RHS of (6.39), $\boldsymbol{\epsilon}_{t,K}$, is due to the IBI, whose first N non-zero elements expressed as

$$\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_{i,N} \begin{bmatrix} \mathbf{s}_{\text{pre}} \\ \mathbf{s}_{\text{post}} \end{bmatrix} \quad (6.42)$$

can only be eliminated by the ZP or the NZP. In the latter case, we can remove the IBI by explicitly subtracting its effect from the observation samples.

To implement the block MMSE equalizer, let us choose \mathbf{s}_{vir} to contain some ran-

dom symbols that have the same distribution as \mathbf{s}_{N-L} such that[†]

$$\mathcal{E}_s\{\mathbf{s}_K \mathbf{s}_K^H\} = \mathbf{I}_K. \quad (6.43)$$

Transformed into the frequency domain, the EDM in (6.39) becomes:

$$\mathbf{y}_{f,K} = \mathbf{H}_{f,K} \mathbf{W}_K \mathbf{s}_K + \boldsymbol{\epsilon}_{f,K} + \mathbf{v}_{f,K}, \quad (6.44)$$

where

$$\begin{aligned} \mathbf{y}_{f,K} &:= \mathbf{W}_K \mathbf{y}_K, \\ \boldsymbol{\epsilon}_{f,K} &:= \mathbf{W}_K \boldsymbol{\epsilon}_{t,K}, \\ \mathbf{v}_{f,K} &:= \mathbf{W}_K \mathbf{v}_{t,K}. \end{aligned} \quad (6.45)$$

$\mathbf{H}_{f,K}$ stands for the FD channel matrix

$$\mathbf{H}_{f,K} := \mathbf{W}_K \mathbf{H}_{t,K} \mathbf{W}_K^H, \quad (6.46)$$

which is full. Like in the previous section, we use a strictly banded matrix $\hat{\mathbf{H}}_{f,K}$ to replace $\mathbf{H}_{f,K}$ with $\hat{\mathbf{H}}_{f,K}$ having non-zero entries only on the main diagonal, the $Q/2$ super- and the $Q/2$ sub-diagonals. Neglecting the IBI $\boldsymbol{\epsilon}_{f,K}$, the block MMSE equalizer can be expressed as

$$\hat{\mathbf{s}}_{N-L} = \mathbf{Z}_{N-L}^H \mathbf{W}_K^H \hat{\mathbf{H}}_{f,K}^H (\hat{\mathbf{H}}_{f,K} \hat{\mathbf{H}}_{f,K}^H + \mathbf{R}_{v,K})^{-1} \mathbf{y}_{f,K}, \quad (6.47)$$

where

$$\begin{aligned} \mathbf{Z}_{N-L} &:= [\mathbf{I}_{N-L}, \mathbf{0}_{(N-L) \times (K-N+L)}]^T, \\ \mathbf{Z} &:= [\mathbf{I}_N, \mathbf{0}_{N \times (K-N)}]^T, \end{aligned} \quad (6.48)$$

and

$$\mathbf{R}_{v,K} := \sigma^2 \mathbf{W}_K \mathbf{Z} \mathcal{D}\{\mathbf{w} \odot \mathbf{w}^*\} \mathbf{Z}^H \mathbf{W}_K^H, \quad (6.49)$$

denotes the covariance matrix of the noise in the EDM. Similar to the ODM, if $\mathbf{R}_{v,K}$ is banded with bandwidth $2Q+1$, we can show that the thereby invoked complexity in (6.47) is linear in K and square in Q .

Hence, to enhance the equalization performance, the window of the EDM should take a three-fold task: 1) to make the noise covariance matrix $\mathbf{R}_{v,K}$ strictly banded; 2) to minimize the IBI $\|\boldsymbol{\epsilon}_{f,K}\|$ in an average sense; and 3) to minimize the band approximation error $\|\mathbf{H}_{f,K} - \hat{\mathbf{H}}_{f,K}\|$ in an average sense.

[†]To reduce the complexity, we can also set $\mathbf{s}_{\text{vir}} = \mathbf{0}$ (and thus $\mathbf{y}_{\text{vir}} = \mathbf{0}$), which in practice does not degrade the equalization performance too much. After all, we are not interested in their specific values.

6.4.2 Window Design for the EDM

Regarding the noise-shaping behavior of the window, unfortunately the following lemma holds.

Lemma 6.1. *It is impossible for $\mathbf{R}_{v,K}$ to be strictly banded.*

Proof. We prove this by contradiction. Observe that $\mathbf{R}_{v,K}$ defined in (6.49) must be a circulant matrix with its first row equal to $\frac{\sigma^2}{\sqrt{N}}(\mathbf{w} \odot \mathbf{w}^*)^T \mathbf{Z}^H \mathbf{W}_K^H$. Should $\mathbf{R}_{v,K}$ be strictly banded with bandwidth $2Q + 1$, the first row should have zeros on the entries indexed from $Q + 2$ until $K - Q$. In other words, we need

$$(\mathbf{w} \odot \mathbf{w}^*)^T \mathbf{Z}^H \mathbf{W}_K^H \mathbf{P} = \mathbf{0}_{1 \times (K-2Q-1)} \quad (6.50)$$

with

$$\mathbf{P} := [\mathbf{0}_{(K-2Q-1) \times (Q+1)}, \mathbf{I}_{K-2Q-1}, \mathbf{0}_{(K-2Q-1) \times Q}]^T. \quad (6.51)$$

However, $\mathbf{Z}^H \mathbf{W}_K^H \mathbf{P}$ is an $N \times (K - 2Q - 1)$ Vandermonde matrix, and thus has full row-rank if $N \leq K - 2Q - 1$. In that case, there exists no non-zero vector $\mathbf{w} \odot \mathbf{w}^*$ that lies in the noise subspace of $\mathbf{Z}^H \mathbf{W}_K^H \mathbf{P}$, and $\mathbf{R}_{v,K}$ can therefore never be banded. \square

As a compromise, we have to use an approximation of $\mathbf{R}_{v,K}$. This idea is reflected in the following proposition.

Proposition 6.1. *The noise covariance matrix will be approximately banded if the window is constructed as*

$$\mathbf{w} = \bar{\mathbf{B}}_N^{(0)} \mathbf{d}, \quad (6.52)$$

where $\bar{\mathbf{B}}_N^{(0)}$ is an $N \times (Q + 1)$ matrix with entries

$$[\bar{\mathbf{B}}_N^{(0)}]_{p,q} = \frac{1}{\sqrt{K}} e^{j \frac{2\pi}{K} p(q - \frac{Q}{2})}. \quad (6.53)$$

Proof. From Proposition 3.1, we understand that the noise covariance matrix $\mathbf{R}_{v,K}$ could be banded if it looked like

$$\mathbf{R}_{v,K} = \sigma^2 \mathbf{W}_K \mathcal{D}\{\mathbf{w}_K \odot \mathbf{w}_K^*\} \mathbf{W}_K^H, \quad (6.54)$$

with the $K \times 1$ vector \mathbf{w}_K structured as

$$\mathbf{w}_K = \mathbf{B}_K \mathbf{d}, \quad (6.55)$$

where \mathbf{d} stands for a $(Q+1) \times 1$ vector, and \mathbf{B}_K a $K \times (Q+1)$ matrix comprised of the first $Q/2+1$ and the last $Q/2$ columns of a K -point DFT matrix \mathbf{W}_K . Therefore, if we adopt the window

$$\mathbf{w}_K = [\mathbf{w}^T, \mathbf{0}_{1 \times (K-N)}^T]^T, \quad (6.56)$$

by letting \mathbf{w} taking the first N values of \mathbf{w}_K in (6.55), $\mathbf{R}_{v,K}$ will be approximately banded. This is equivalent to (6.52), because $\bar{\mathbf{B}}_N^{(0)}$ defined in (6.53) corresponds to the first N rows of \mathbf{B}_K . \square

Stemming from the definition of $\bar{\mathbf{B}}_N^{(0)}$ in (6.53), we introduce a more general notation $\bar{\mathbf{B}}_M^{(l)}$, which denotes an $M \times (Q+1)$ matrix consisting of the l th until $(l+M-1)$ st row of \mathbf{B}_K . In mathematics, this means that $\bar{\mathbf{B}}_M^{(l)}$ has entries

$$[\bar{\mathbf{B}}_M^{(l)}]_{n,q} = \frac{1}{\sqrt{K}} e^{j \frac{2\pi}{K} (n+l)(q-\frac{Q}{2})}, \quad (6.57)$$

for $n = l, \dots, l+M-1$, and $q = 0, \dots, Q$. If we compare the above definition with (2.33), one can directly see that $\bar{\mathbf{B}}_M^{(l)}$ tallies with the definition of the (O)CE-BEM, which uses an exponential period K and a BEM size M with $K > M$.

The (O)CE-BEM plays an important role in minimizing the band approximation error $\|\mathbf{H}_{f,K} - \hat{\mathbf{H}}_{f,K}\|$ as is evident from the following theorem (see Appendix 6B for a proof).

Theorem 6.2. *If we design the non-zero entries of the matrix \mathbf{U} in (6.41), $\hat{h}_{p,l}$, to be*

$$\hat{h}_{p,l} = \frac{1}{\sqrt{K}} \sum_{q=0}^Q e^{-j \frac{2\pi}{K} p(q-\frac{Q}{2})} c_{q,l}, \quad (6.58)$$

with $c_{q,l}$ denoting some non-zero coefficient, then the resulting band approximation error can be transformed in the time domain as the error resulting from the (O)CE-BEM. In mathematics, this can be expressed as

$$\|\mathbf{H}_{f,K} - \hat{\mathbf{H}}_{f,K}\| = \sum_{l=0}^L \|\mathcal{D}\{\Upsilon_l \mathbf{w}\} \mathbf{h}_l - \bar{\mathbf{B}}_{N-L}^{(l)} \mathbf{c}_l\|^2, \quad (6.59)$$

where Υ_l stands for an $(N-L) \times N$ selection matrix

$$\Upsilon_l := [\mathbf{0}_{(N-L) \times l}, \mathbf{I}_{(N-L)}, \mathbf{0}_{(N-L) \times (L-l)}]; \quad (6.60)$$

\mathbf{h}_l is an $(N-L) \times 1$ vector collecting the l th channel tap in the range

$$\mathbf{h}_l := [h_{l,l}, \dots, h_{N-L+l-1,l}]^T, \quad (6.61)$$

and \mathbf{c}_l is a $(Q+1) \times 1$ vector,

$$\mathbf{c}_l := [c_{0,l}, \dots, c_{Q,l}]^T. \quad (6.62)$$

In Theorem 6.2, we use for each channel tap a slightly different (O)CE-BEM matrix $\bar{\mathbf{B}}_{N-L}^{(l)}$ to approximate the time variation contained in

$$\mathcal{D}\{\Upsilon_l \mathbf{w}\} \mathbf{h}_l = [w_l h_{l,l}, \dots, w_{N-L+l} h_{N-L+l,l}]^T.$$

The resulting BEM modeling error added up for each l accounts for the band approximation error.

Next, to minimize the IBI, the following theorem is useful (see Appendix 6C for a proof).

Theorem 6.3. *The average power of the IBI $\epsilon_{f,K}$ for the EDM is related to the window as*

$$\mathcal{E}_{h,s}\{\|\epsilon_{f,K}\|^2\} = \mathbf{w}^T \bar{\mathbf{R}}_{\epsilon,N} \mathbf{w}^*, \quad (6.63)$$

with $\bar{\mathbf{R}}_{\epsilon,N}$ denoting an $N \times N$ diagonal matrix with the n th diagonal entry equal to

$$[\bar{\mathbf{R}}_{\epsilon,N}]_{n,n} = \begin{cases} \sum_{l=n+L_z+1}^L \sigma_l^2 & \text{if } n \leq L - L_z - 1, \\ \sum_{l=0}^{n-N+L-L_z} \sigma_l^2 & \text{if } N - L + L_z \leq n \leq N - 1, \\ 0 & \text{otherwise.} \end{cases} \quad (6.64)$$

Eventually, to minimize the IBI and band approximation error jointly in the EDM, we come up with the following design problem:

$$\begin{aligned} \min_{\{\mathbf{w}\}} \mathcal{E}_h \left\{ \min_{\{\mathbf{c}_l\}} \sum_{l=0}^L \|\mathcal{D}\{\Upsilon_l \mathbf{w}\} \mathbf{h}_l - \bar{\mathbf{B}}_{N-L}^{(l)} \mathbf{c}_l\|^2 \right\} + \theta \mathbf{w}^T \bar{\mathbf{R}}_{\epsilon,N} \mathbf{w}^*, \\ \text{s.t. } \mathbf{w} = \bar{\mathbf{B}}_N^{(0)} \mathbf{d} \text{ and } \|\mathbf{w}\|^2 = N, \end{aligned} \quad (6.65)$$

where a weight factor θ is again utilized. The first constraint above is due to Proposition 6.1. We solve the above design problem first for \mathbf{c}_l . Under Assumption 6.1, different channel taps are uncorrelated, i.e.,

$$\mathcal{E}\{\mathbf{h}_l \mathbf{h}_{l'}^H\} = 0, \quad (6.66)$$

for $l \neq l'$. Therefore, we can compute \mathbf{c}_l as

$$\mathbf{c}_l = \bar{\mathbf{B}}_{N-L}^{(l)\dagger} \mathcal{D}\{\Upsilon_l \mathbf{w}\} \mathbf{h}_l, \quad (6.67)$$

after which (6.65) becomes

$$\begin{aligned} \min_{\{\mathbf{w}\}} \sum_{l=0}^L \mathcal{E}_h \left\{ \|\mathcal{P}_{\bar{\mathbf{B}}_{N-L}^{(l)}}^{(l)} \mathcal{D}\{\Upsilon_l \mathbf{w}\} \mathbf{h}_l\|^2 \right\} + \theta \mathbf{w}^T \bar{\mathbf{R}}_{\epsilon,N} \mathbf{w}^*, \\ \text{s.t. } \mathbf{w} = \bar{\mathbf{B}}_N^{(0)} \mathbf{d} \text{ and } \|\mathbf{w}\|^2 = N, \end{aligned} \quad (6.68)$$

where

$$\mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)} := \mathbf{I}_{N-L} - \bar{\mathbf{B}}_{N-L}^{(l)} \bar{\mathbf{B}}_{N-L}^{(l)\dagger}. \quad (6.69)$$

The first term in (6.68) can be worked out as

$$\begin{aligned} & \mathcal{E}_h \{ \|\mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)} \mathcal{D}\{\Upsilon_l \mathbf{w}\} \mathbf{h}_l\|^2 \} \\ &= \text{tr} \left(\mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)} \Upsilon_l \mathcal{D}\{\mathbf{w}\} \Upsilon_l^H \mathcal{E}\{\mathbf{h}_l \mathbf{h}_l^H\} \Upsilon_l \mathcal{D}\{\mathbf{w}^H\} \Upsilon_l^H \mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)H} \right). \end{aligned} \quad (6.70)$$

The above equality holds because

$$\mathcal{D}\{\Upsilon_l \mathbf{w}\} = \Upsilon_l \mathcal{D}\{\mathbf{w}\} \Upsilon_l^H. \quad (6.71)$$

By introducing the following notation

$$\begin{aligned} \mathbf{R}_h^{(l)} &:= \Upsilon_l^H \mathcal{E}\{\mathbf{h}_l \mathbf{h}_l^H\} \Upsilon_l \\ &= \mathcal{E}_h \{ [\mathbf{0}_{1 \times l}, \mathbf{h}_l^T, \mathbf{0}_{1 \times (L-l)}]^T [\mathbf{0}_{1 \times l}, \mathbf{h}_l^T, \mathbf{0}_{1 \times (L-l)}]^* \}, \end{aligned} \quad (6.72)$$

we can rewrite (6.70) further as

$$\begin{aligned} \mathcal{E}_h \{ \|\mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)} \mathcal{D}\{\Upsilon_l \mathbf{w}\} \mathbf{h}_l\|^2 \} &= \text{tr} \left(\mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)} \Upsilon_l \mathcal{D}\{\mathbf{w}\} \mathbf{R}_h^{(l)} \mathcal{D}\{\mathbf{w}^H\} \Upsilon_l^H \mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)H} \right) \\ &= \mathbf{w}^T \bar{\mathcal{X}}_N^{(l)} \mathbf{w}^*, \end{aligned} \quad (6.73)$$

with

$$\bar{\mathcal{X}}_N^{(l)} := \sum_{n=0}^{N-L-1} \mathcal{D}\{[\mathbf{0}_{1 \times l}, \mathbf{e}_n^T \mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)}, \mathbf{0}_{1 \times (L-l)}]^T\} \mathbf{R}_h^{(l)} \mathcal{D}\{[\mathbf{0}_{1 \times l}, \mathbf{e}_n^T \mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)}, \mathbf{0}_{1 \times (L-l)}]^H\}. \quad (6.74)$$

Obviously, the vector $[\mathbf{0}_{1 \times l}, \mathbf{e}_n^T \mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)}, \mathbf{0}_{1 \times (L-l)}]^T$ in the above corresponds to the n th row of $\mathcal{P}_{\bar{\mathbf{B}}_{N-L}}^{(l)} \Upsilon_l$.

Substituting (6.73) in (6.68) and taking (6.52) into account, we can simplify the minimization problem further to

$$\begin{aligned} \mathbf{d} &= \arg \min \mathbf{d}^T \bar{\mathcal{X}}_N \mathbf{d}^*, \\ \text{s.t. } & \|\bar{\mathbf{B}}_N^{(0)} \mathbf{d}\|^2 = N, \end{aligned} \quad (6.75)$$

with

$$\bar{\mathcal{X}}_N := \bar{\mathbf{B}}_N^{(0)T} \left(\sum_{l=0}^L \bar{\mathcal{X}}_N^{(l)} + \theta \bar{\mathbf{R}}_{\epsilon, N} \right) \bar{\mathbf{B}}_N^{(0)*}. \quad (6.76)$$

Unlike the ODM case, here the columns of $\bar{\mathbf{B}}_N^{(0)}$ are not orthonormal to each other. As a result, we have to compute \mathbf{d} as the generalized eigenvector that corresponds to the least significant generalized eigenvalue of the matrix pair $(\bar{\mathcal{X}}_N^*, \bar{\mathbf{B}}_N^{(0)H} \bar{\mathbf{B}}_N^{(0)})$ [37].

The following remarks are in order.

Remark 6.3.

In the EDM case, we transform the band approximation error in the frequency domain to the (O)CE-BEM modeling error in the time domain in light of (6.59). Equipped with a larger exponential period, the (O)CE-BEM channel is reported to fit to a realistic TV channel much tighter than the (C)CE-BEM [122].

Remark 6.4.

It is unique to the EDM that the band approximation error $\|\mathbf{H}_{f,K} - \hat{\mathbf{H}}_{f,K}\|$ does not correspond to the out-of-band interference of $\mathbf{H}_{f,K}$ as in the ODM. In other words,

$$\hat{\mathbf{H}}_{f,K} \neq \mathbf{H}_{f,K} \odot \mathbf{T}_Q. \quad (6.77)$$

This is because in the ODM, the columns of the (C)CE-BEM are also the columns of the N -point DFT matrix, and hence the modeling error is orthogonal to the (C)CE-BEM channel. However, this fact does not hold for the EDM that uses the (O)CE-BEM. As a result, $\hat{\mathbf{H}}_{f,K}$ and $\mathbf{H}_{f,K}$ will not share the same diagonals: they are analogous to each other only in terms of the Frobenius norm.

Remark 6.5.

The weight factor θ in (6.65) plays a more significant role in the EDM than in the ODM. First of all, it is more difficult to find a window that is able to minimize the BEM modeling error and the IBI jointly. Minimizing them separately will often yield opposing window solutions. For example, the optimal window that only minimizes the IBI ($\theta = \infty$) should have L zeros at both edges. The channel resulting from such a window will exhibit discontinuities and all-zero areas, both of which make it extremely hard for the (O)CE-BEM to yield a tight fit. On the other hand, the window that only minimizes the BEM modeling error ($\theta = 0$) takes on usually a near-to-rectangular shape (this is due to the superior modeling performance of the (O)CE-BEM), which is thus incapable of suppressing the IBI. The tuning of θ will depend on the actual transmission parameters, such as the block size, the delay/Doppler spread of the channel, etc. For a practical time-varying system, the BEM modeling error usually plays a more demanding role than the IBI. Indeed, the IBI can be diminished by inserting a guard interval, though not necessarily longer than the channel delay spread. By adopting a sliding window approach in the equalization [92], the impact of the IBI can be further confined. In this sense, a smaller θ tends to be preferred as shown in the simulation part.

6.5 Numerical results

We test the proposed algorithms over a TV channel following Jakes' Doppler profile [45] using the TV channel generator given in [124]. The channel is assumed to have

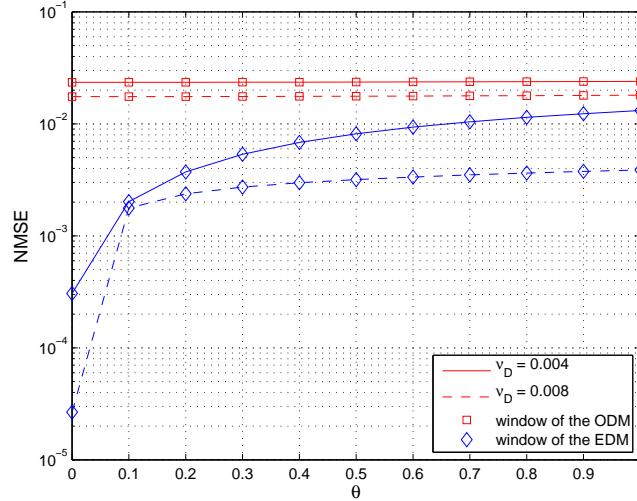


Figure 6.3: NMSE vs. θ .

$L + 1 = 31$ channel taps with the l th tap having a variance $\sigma_l^2 = c e^{-\frac{l}{10}}$ with c standing for a normalization constant.

For the transmitter setup, we use QPSK to modulate the data symbols and employ for the moment no guard interval, i.e., $L_z = 0$. Further, we set $N = 256$ for the ODM case, and set $N = 158$ and $K = 256$ for the EDM case. In this way, we can construct a block MMSE equalizer that works for both cases on the same frequency grid and inflicts an analogous complexity. To mitigate the absence of the guard interval, we will consider the BER only of the 32 data symbols that lie in the middle of the block, i.e., we assume a sliding window approach (for more details see [92]).

Study case 1. Window design as a function of θ .

As we understand from the earlier analysis, a larger θ corresponds to a window that will suppress the IBI more. We will study whether this will influence the band approximation error (BEM modeling error). To this end, we define the normalized mean square error (NMSE), which is defined as the MSE of (6.26) and (6.59) for the ODM and EDM, respectively, normalized to the block size N .

Fig. 6.3 depicts the relationship between the NMSE and θ for a normalized Doppler spread $\nu_D = 0.004$ using $Q = 2$, and a normalized Doppler spread $\nu_D = 0.008$ using $Q = 4$. Fig. 6.4 depicts the resulting BER performance at an SNR = 30dB. Fig. 6.5 and Fig. 6.6 show window examples resulting from some different θ 's.

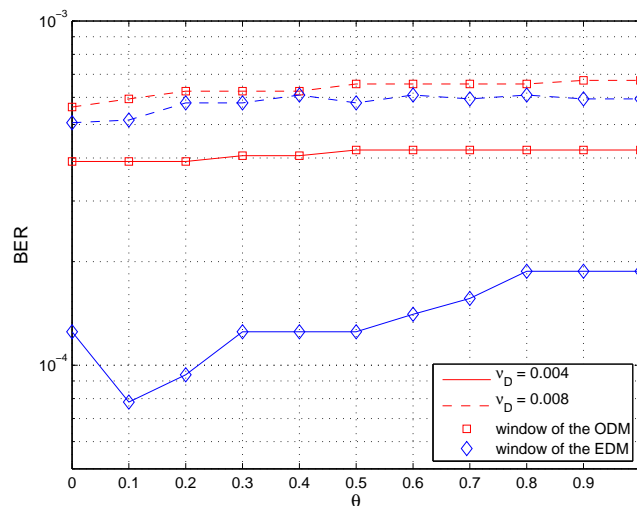


Figure 6.4: BER vs. θ .

From these figures, we can observe that in the ODM case, despite the different values of θ , the resulting windows are similar to each other, and so are the NMSE and BER performances. This is due to the window shape, which tends to zero at the edges, and is thus beneficial for suppression of both the BEM modeling error and the IBI. The windows in the EDM case exhibit a totally different behavior: the window corresponding to $\theta = 0$ is almost flat, which suggests that the (O)CE-BEM alone is already adequate to reduce the BEM modeling error. However, the IBI is virtually invariant to such a flat window. By enlarging θ , we can expect the edges of the window to tend more to zero, but the BEM modeling performance degrades rapidly as can be observed in Fig. 6.3. Fig. 6.4 suggests that a smaller θ favors the BEM modeling performance because the high Doppler spread has a relatively large impact on the overall BER whereas the IBI affects mainly the first and last data symbols of the block, and is thus less serious thanks to the utility of the sliding-window approach.

In reality, it is difficult to find a BER-optimal θ in closed form. A similar problem occurred in [64, 54] though in a different context. In the following simulations, we will just adopt $\theta = 0$ for both the ODM and EDM.

Study case 2: BER performance.

We here compare the BER performance for a range of SNRs. Also included is

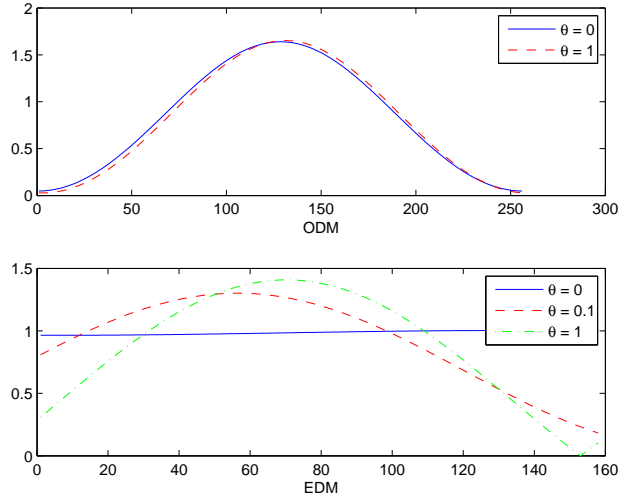


Figure 6.5: The obtained windows corresponding to different θ 's. $\nu_D = 0.004$.

the window of [92] applied to the ODM. Because this window can be longer than the data block, we choose the window length to be $N + L$. The BER comparison is made for two different kinds of TV channels $\nu_D = 0.004$ and $\nu_D = 0.008$, where we set $Q = 2$ for the former and $Q = 4$ for the latter to account for the channel time-variation properly.

Fig. 6.7 shows that the window of the ODM and of [92] render similar performances as we expected. At high SNR, the window of the EDM exhibits a remarkable performance lead, which is, however, less prominent when the channel varies faster.

In the above cases, we have assumed $L_z = 0$, which can be too harsh for a practical system. It will be more usual if we allow for a short guard interval, e.g., ZP or NZP, though not necessarily longer than the channel. In these cases, we can expect the BER performance to improve especially for the EDM. This is justified in Fig. 6.8 for $\nu_D = 0.004$, and in Fig. 6.9 for $\nu_D = 0.008$.

6.6 Summary

In this chapter, we discussed how to equalize the channel in a single-carrier system, which is achieved in the frequency domain. Because the channel is plagued by fast

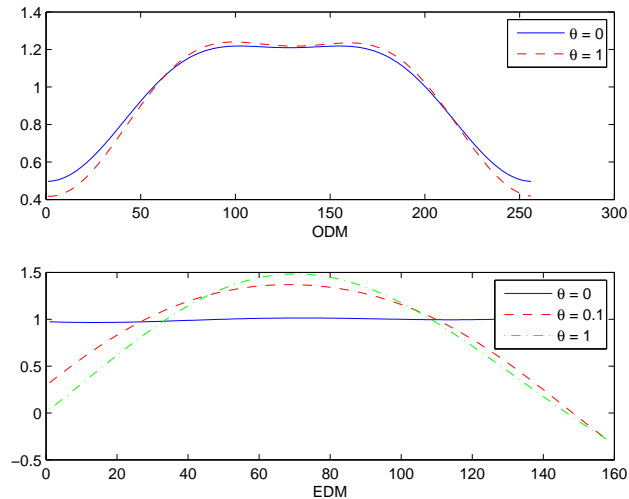


Figure 6.6: The obtained windows corresponding to different θ 's. $\nu_D = 0.008$.

fading and IBI, its matrix expression in the frequency domain is not diagonal but full. For the sake of low-complexity, we have to approximate it as banded, inducing thereby out-of-band interference, which degrades the performance.

In spite of the utility of a window, we have shown that the band approximation in the ODM is equivalent to modeling the TV channel with the (C)CE-BEM in the time domain. The (C)CE-BEM usually suffers from a large modeling error especially for a fast fading channel.

As a remedy, we introduced redundancy at the receiver in the EDM. By this means, we are able to exploit a better resolution of the Doppler spectrum in the frequency domain. It was shown that the band approximation in the EDM is equivalent to modeling the TV channel with the (O)CE-BEM in the time domain, which is typically known to have a much smaller modeling error than the (C)CE-BEM. The noise floor in the EDM is therefore considerably reduced.

The proposed channel equalizer relies on the CSI, which is usually not available at the receiver. Channel estimation for a single-carrier system is discussed in the next chapters.

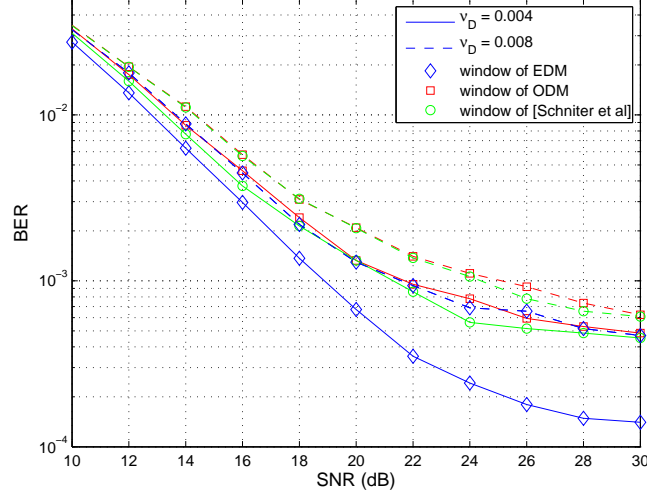


Figure 6.7: BER performance vs. SNR.

Appendix 6A: Proof of Theorem 6.1

By the definition of $\epsilon_{f,N}$ in (6.13), we understand that

$$\mathcal{E}_{h,s}\{\|\epsilon_{f,N}\|^2\} = \mathcal{E}_{h,s}\{\|\mathcal{D}\{\mathbf{w}\}\mathbf{H}_{i,N}(\mathbf{s}_{\text{pre}} - \mathbf{s}_{\text{post}})\|^2\}, \quad (6.78)$$

which, by adopting the zero-mean, unit-variance white assumption on the data symbols in Assumption 6.2, becomes

$$\mathcal{E}_{h,s}\{\|\epsilon_{f,N}\|^2\} = 2\text{tr}(\mathcal{D}\{\mathbf{w}\}\mathcal{E}_h\{\mathbf{H}_{i,N}\Phi_L\mathbf{H}_{i,N}\mathcal{D}\{\mathbf{w}^H\}\}), \quad (6.79)$$

where an $L \times L$ diagonal matrix Φ_L is introduced to account for a possible guard interval in \mathbf{s}_{pre} and \mathbf{s}_{post} :

$$\Phi_L := \mathcal{D}\{[\mathbf{1}_{1 \times (L-L_z)}, \mathbf{0}_{1 \times L_z}]^T\}. \quad (6.80)$$

It can be shown that under the WSSUS channel assumption in Assumption 6.1,

$$\mathcal{E}_h\{\mathbf{H}_{i,N}\Phi_L\mathbf{H}_{i,N}^H\} = \mathbf{R}_{\epsilon,N}. \quad (6.81)$$

Substituting the above in (6.79) concludes the proof.

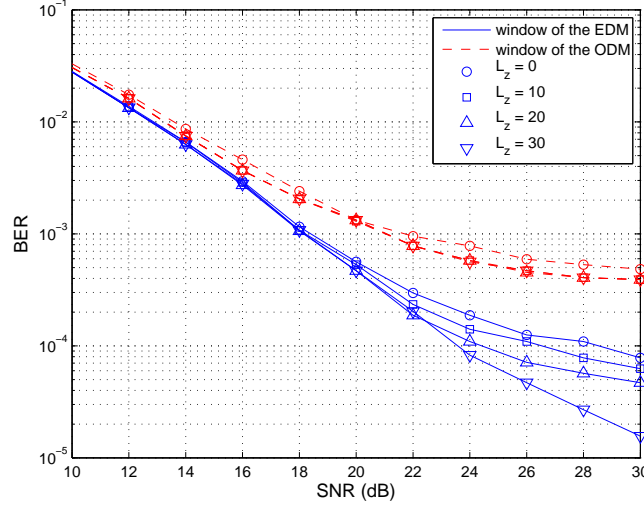


Figure 6.8: BER performance vs. SNR corresponding to different L_z . $\nu_D = 0.004$.

Appendix 6B: Proof of Theorem 6.2

From Lemma 3.1, the banded matrix $\hat{\mathbf{H}}_{f,K}$ in the frequency domain can be expressed in the time domain as

$$\begin{aligned}\hat{\mathbf{H}}_{t,K} &:= \mathbf{W}_K^H \hat{\mathbf{H}}_{f,K} \mathbf{W}_K \\ &= \sum_q \mathcal{D}\{\mathbf{B}_K \mathbf{e}_q\} \mathbf{C}_K^q,\end{aligned}\quad (6.82)$$

where \mathbf{C}_K^q is a circulant matrix with $[c_{q,0}, \dots, c_{q,L}, \mathbf{0}_{1 \times (K-L-1)}]^T$ as its first column, and $c_{q,l}$ standing for some design parameter. As a result, the band approximation error in the frequency domain can be transformed in the time domain as

$$\|\mathbf{H}_{f,K} - \hat{\mathbf{H}}_{f,K}\| = \|\mathbf{H}_{t,K} - \hat{\mathbf{H}}_{t,K}\|. \quad (6.83)$$

Recall from (6.39) that $\mathbf{H}_{t,K}$ is comprised of three parts

$$\mathbf{H}_{t,K} := \begin{bmatrix} \mathcal{D}\{\mathbf{w}\} \bar{\mathbf{H}}_N \\ \mathbf{0}_{(K-N) \times (N-L)} \end{bmatrix} \mathbf{U}, \quad (6.84)$$

where the $K \times (K - N + L)$ matrix \mathbf{U} is assigned in (6.41) with virtual channel taps $\hat{h}_{p,n}$. Because the value of $\hat{h}_{p,n}$ is subject to design, and our target is to minimize

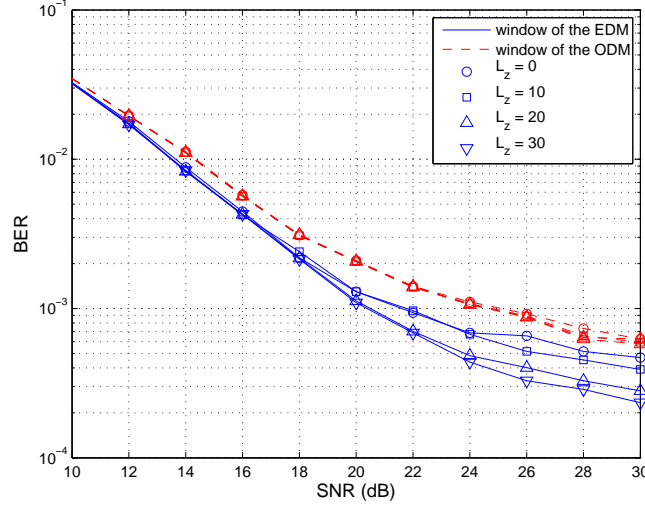


Figure 6.9: BER performance vs. SNR corresponding to different L_z . $\nu_D = 0.008$.

the difference $\|\mathbf{H}_{t,K} - \hat{\mathbf{H}}_{t,K}\|$, we can simply let \mathbf{U} take the last $K - N + L$ columns of $\hat{\mathbf{H}}_{t,K}$. In mathematics, this is equivalent to adopting (6.58). In this way, $\mathbf{H}_{t,K}$ and $\hat{\mathbf{H}}_{t,K}$ will be different only in their first $N - L$ columns. More specifically, they are different only in the $L + 1$ most significant diagonals of $\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_N$ since the remaining diagonals are all zero as can be observed in Fig. 6.2.

The l th diagonal of $\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_N$ corresponds to the l th channel tap, spanning from the l th until the $(N - L + l)$ th time-instance, which can thus be expressed as

$$[w_l h_{l,l}, \dots, w_{N-L+l} h_{N-L+l,L}]^T = \mathcal{D}\{\boldsymbol{\Upsilon}_l \mathbf{w}\} \mathbf{h}_l. \quad (6.85)$$

for $l = 0, \dots, L$.

From the definition in (6.82), we understand that the (m, n) th entry of $\hat{\mathbf{H}}_{t,K}$, if $0 \leq n - m \leq L$, can be expressed as

$$[\hat{\mathbf{H}}_{t,K}]_{n,m} = \sum_q \frac{1}{\sqrt{K}} e^{j \frac{2\pi}{K} n(q - \frac{q}{2})} c_{q,n-m}. \quad (6.86)$$

As a result, the corresponding part of the l th diagonal of $\hat{\mathbf{H}}_{t,K}$ can be expressed after some algebra as

$$[[\hat{\mathbf{H}}_{t,K}]_{l,l-l}, \dots, [\hat{\mathbf{H}}_{t,K}]_{N-L-1, N-L-l-1}]^T = \bar{\mathbf{B}}_{N-L}^{(l)} \mathbf{c}_l. \quad (6.87)$$

Substituting (6.85) and (6.87) in the RHS of (6.83) concludes the proof.

Appendix 6C: Proof of Theorem 6.3

By the definition of $\epsilon_{f,K}$ in (6.42), we understand that

$$\mathcal{E}_{h,s}\{\|\epsilon_{f,K}\|^2\} = \mathcal{E}_{h,s}\|\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{H}}_{i,N} \begin{bmatrix} \mathbf{s}_{\text{pre}} \\ \mathbf{s}_{\text{post}} \end{bmatrix}\|^2. \quad (6.88)$$

which, by adopting the zero-mean, unit-variance white assumption on the data symbols in Assumption 6.2, becomes

$$\mathcal{E}_{h,s}\{\|\epsilon_{f,K}\|^2\} = \text{tr}(\mathcal{D}\{\mathbf{w}\}\mathcal{E}_h\{\bar{\mathbf{H}}_{i,N}\bar{\Phi}_N\bar{\mathbf{H}}_{i,N}^H\}\mathcal{D}\{\mathbf{w}^H\}), \quad (6.89)$$

with

$$\bar{\Phi}_N := \mathcal{D}\{[\mathbf{1}_{1 \times (L-L_z)}, \mathbf{0}_{1 \times 2L_z}, \mathbf{1}_{1 \times (L-L_z)}]^T\}, \quad (6.90)$$

which is introduced to account for the presence of a guard interval (ZP). Let us use $\bar{\mathbf{R}}_{\epsilon,N}$ to denote

$$\bar{\mathbf{R}}_{\epsilon,N} := \mathcal{E}_h\{\bar{\mathbf{H}}_{i,N}\bar{\Phi}_N\bar{\mathbf{H}}_{i,N}^H\}. \quad (6.91)$$

Under the WSSUS channel assumption in Assumption 6.1, $\bar{\mathbf{R}}_{\epsilon,N}$ will be a diagonal matrix with its diagonals defined as in (6.64). As a result,

$$\begin{aligned} \mathcal{E}_{h,s}\{\|\epsilon_{f,K}\|^2\} &= \text{tr}(\mathcal{D}\{\mathbf{w}\}\bar{\mathbf{R}}_{\epsilon,N}\mathcal{D}\{\mathbf{w}^H\}) \\ &= \mathbf{w}^T \bar{\mathbf{R}}_{\epsilon,N} \mathbf{w}^*. \end{aligned}$$

Chapter 7

Channel Estimation in a Single-Carrier System

7.1 Introduction

The equalizer described in the previous chapter requires perfect knowledge of the CSI, which is usually not available at the receiver in practice. We will discuss in this chapter how to estimate the channel for a single-carrier system. Compared with the approaches described in Chapter 4, which are proposed for an OFDM system, the channel estimation methods proposed in this chapter bear some analogies: first, the channel time-variation will be also modeled by an arbitrary but tight BEM and the channel estimation will hence be achieved by estimating the BEM coefficients. Second, the channel matrix will be assumed to be banded as well: for the OFDM case in Chapter 4, this assumption is valid since for a realistic Doppler spread, most of the channel power in the frequency domain is located in the entries that are close to the diagonal. For the single-carrier case, the band assumption is due to the FIR feature of the channel in the time domain.

The works in [55, 19, 111, 112] belong to the few that focus on blind BEM channel estimation, which is beneficial for bandwidth efficiency, but usually results into a complicated channel estimator. In this aspect, pilot-aided methods might be more useful. To save bandwidth, the channel estimator in [42] uses superimposed pilots, but the disadvantage is that at each symbol time, a portion of the total power must be allocated to the pilots. In addition, the data-induced interference can give rise to a high noise floor, which can only be remedied by some additional post-processing (data-aided channel estimation). As one solution, the superimposed pilots in [34] are designed to be orthogonal to the data such that the channel estimation can be achieved free from interference. However, the equalizer resulting from such a transmitter design entails more complexity. Other than superimposed training, the pilots can also be interleaved with the data in the time domain, a scheme that is considered in [10, 94, 59, 122, 78] as well as in this chapter. Of course, the bandwidth efficiency will be consequently sacrificed.

At a first glance, pilot-aided channel estimation seems to be easier in the single-

The results of this chapter appeared in [102, Chapter 3].

carrier system than in OFDM. This might be attributed to the fact that the TD channel matrix in the single-carrier system is strictly banded due to the FIR channel assumption, while the channel matrix in OFDM is in principle full, and thus the channel estimator is constantly plagued by the data-related interference as we have shown in Chapter 4. Indeed, for a strictly banded channel matrix, if the pilots can be grouped in a sufficiently long cluster, we are able to find observation samples that are solely dependent on pilots. This scheme is adopted in most of the pilot-aided channel estimators and is best summarized in [57]. However, in other cases where the pilot cluster is short [78], data-related interference still exists. In addition, it is sometimes even beneficial to deliberately introduce some interference. For instance, the observation samples at the edges of the cluster are not only contributed by pilots, but also contaminated by the unknown data. If we design the channel estimator properly, then including these samples in the channel estimation will be useful to average the noise better, which could compensate the loss due to the thereby induced interference. The overall performance can be still improved.

7.2 System Model

In a single-carrier transmission scheme, the I/O relationship (without windowing) can be expressed as

$$y_n = \sum_{l=0}^L h_{n,l} s_{n-l} + v_n, \quad (7.1)$$

where y_n , s_n and v_n represent the received sample, transmitted symbol and noise at the n th time instance, respectively. $h_{n,l}$ stands for the realization of the l channel tap at the n th time instance. In (7.1), we have thus modeled the channel to be an FIR filter with $L + 1$ taps, where L is assumed to be known to the receiver.

For the time-multiplexed pilot scheme, the pilots are interleaved with data symbols in the time domain. For the sake of notational ease, we will in the sequel assume that the pilots are grouped in clusters, each of the same length $L_t + 1$ *. Suppose \mathbf{t}_k stands for the k th pilot cluster whose starting position is n_k , i.e.,

$$\mathbf{t}_k := [s_{n_k}, \dots, s_{n_k+L_t}]^T. \quad (7.2)$$

The contribution of \mathbf{t}_k can be found in the received samples $[y_{n_k}, \dots, y_{n_k+L_t+L}]^T$. Let us choose \mathbf{y}_k as the k th observation sample cluster

$$\mathbf{y}_k := [y_{n_k+\Delta}, \dots, y_{n_k+L_t+L-\Delta}]^T, \quad (7.3)$$

*The analysis for unequally-long pilot clusters follows a similar path.

where the integer Δ is included to determine which of the received samples will be eventually used for channel estimation. This makes sense since the choice of Δ is related to the amount of pilot information as well as the interference due to the unknown data symbols that is fed to the channel estimator. For instance, in [59], $\Delta = L$ is chosen such that \mathbf{y}_k contains the maximum number of received samples that are exclusively dependent on the pilots and thus free from interference. In [79, 78], the situation $\Delta = 0$ is considered such that \mathbf{y}_k contains the maximum number of received samples that are dependent on the pilots and therefore the interference has to be taken into account. In the sequel, we will follow an approach similar to [57], where Δ is assumed to be an arbitrary positive integer. As a result, we can find the expression for \mathbf{y}_k as

$$\begin{aligned}\mathbf{y}_k &= \mathbf{H}_k [s_{n_k-L+\Delta}, \dots, s_{n_k+L_t+L-\Delta}]^T + \mathbf{v}_k \\ &= \mathbf{H}_k^{(p)} \mathbf{t}_k + \underbrace{\mathbf{H}_k^{(d)} \mathbf{d}_k}_{\boldsymbol{\xi}_k} + \mathbf{v}_k,\end{aligned}\quad (7.4)$$

where the noise term \mathbf{v}_k is similarly defined as \mathbf{y}_k ; \mathbf{d}_k stands for the unknown data symbols

$$\mathbf{d}_k := [s_{n_k-L+\Delta}, \dots, s_{n_k-1}, s_{n_k+L_t+1}, \dots, s_{n_k+L_t+L-\Delta}]^T; \quad (7.5)$$

and \mathbf{H}_k is an $(L_t + L - 2\Delta + 1) \times (L_t + 2L - 2\Delta + 1)$ “Toeplitz” matrix[†]

$$\mathbf{H}_k := \begin{bmatrix} h_{n_k+\Delta,L} & \cdots & h_{n_k+\Delta,0} & & \\ & \ddots & \vdots & \ddots & \\ & & h_{n_k+L_t+L-\Delta,L} & \cdots & h_{n_k+L_t+L-\Delta,0} \end{bmatrix}. \quad (7.6)$$

The $(L_t + L - 2\Delta + 1) \times (L_t + 1)$ matrix $\mathbf{H}_k^{(p)}$ consists of the columns of \mathbf{H}_k that are indexed from $L - \Delta + 1$ until $L - \Delta + L_t + 1$, which correspond to the positions of the pilots \mathbf{t}_k ; The $(L_t + L - 2\Delta + 1) \times (2L - 2\Delta)$ matrix $\mathbf{H}_k^{(d)}$ consists of the remaining columns of \mathbf{H}_k , which correspond to the positions of the data symbols. The partitioning of \mathbf{H}_k is illustrated in Fig. 7.1. From the definitions above, it is not difficult to see that in order for \mathbf{y}_k to be free from interference, Δ must be at least L such that $\mathbf{H}_k^{(d)}$ vanishes (or has a non-positive number of columns). However, it is not wise to choose $\Delta > L$, which leads to information loss. In short, we require that

$$0 \leq \Delta \leq L. \quad (7.7)$$

Suppose there are in total $K + 1$ pilot clusters that are cast for channel estimation, whose starting positions are n_0, \dots, n_K , respectively. For each pilot cluster, we

[†]It is not strictly Toeplitz because it is not constant along the diagonals.

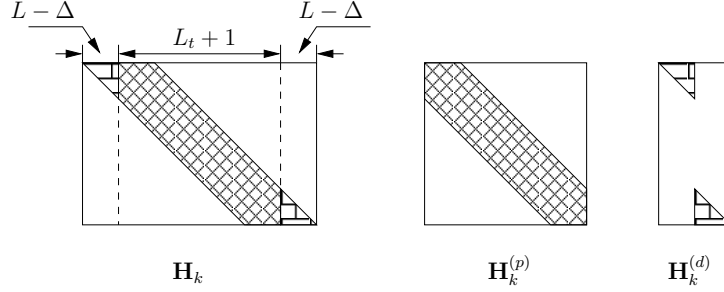


Figure 7.1: Partitioning of the channel matrix \mathbf{H}_k . Its rows corresponds to the positions of \mathbf{y}_k and its columns to the positions of \mathbf{t}_k and \mathbf{d}_k

designate the corresponding observation samples as in (7.3) and establish the I/O relationship as in (7.4). Stacking these results vertically, we obtain

$$\mathbf{y} = \mathbf{H}^{(p)}\mathbf{t} + \underbrace{\mathbf{H}^{(d)}\mathbf{d}}_{\boldsymbol{\xi}} + \mathbf{v}, \quad (7.8)$$

with

$$\mathbf{y} := [\mathbf{y}_0^T, \dots, \mathbf{y}_K^T]^T, \quad (7.9)$$

and \mathbf{t} , \mathbf{d} and \mathbf{v} similarly defined as \mathbf{y} ; further

$$\mathbf{H}^{(p)} := \begin{bmatrix} \mathbf{H}_0^{(p)} & & \\ & \ddots & \\ & & \mathbf{H}_K^{(p)} \end{bmatrix}, \quad (7.10)$$

and $\mathbf{H}^{(d)}$ similarly defined as $\mathbf{H}^{(p)}$.

7.3 Data Model for Channel Estimation

If we assume that the TV channel within the interval $n = 0, \dots, N - 1$ can be tightly approximated by a BEM, then the l th channel tap will admit the following equality

$$[h_{l,0}, \dots, h_{l,N-1}]^T = \underbrace{[\mathbf{b}_0, \dots, \mathbf{b}_Q]}_{\mathbf{B}} [c_{0,l}, \dots, c_{Q,l}]^T, \quad (7.11)$$

where \mathbf{b}_q stands for the q th expansion basis as we have introduced in Chapter 2, and $c_{q,l}$ for the q th BEM coefficient of the l th channel tap. With the aid of the BEM, the CSI can be attained via estimating the BEM coefficients. To this end, let us substitute

(7.11) in (7.4). It can be shown after some algebra that the k th observation sample cluster is related with the BEM coefficients as

$$\begin{aligned} \mathbf{y}_k &= \sum_{q=0}^Q \mathcal{D}\{\mathbf{b}_{k,q}\} \mathbf{C}_q [s_{n_k-L+\Delta}, \dots, s_{n_k+L_t+L-\Delta}]^T + \mathbf{v}_k \\ &= \sum_{q=0}^Q \mathcal{D}\{\mathbf{b}_{k,q}\} \mathbf{C}_q^{(p)} \mathbf{t}_k + \underbrace{\sum_{q=0}^Q \mathcal{D}\{\mathbf{b}_{k,q}\} \mathbf{C}_q^{(d)} \mathbf{d}_k}_{\boldsymbol{\xi}_k} + \mathbf{v}_k, \end{aligned} \quad (7.12)$$

where

$$\mathbf{b}_{k,q} := [[\mathbf{b}_q]_{n_k+\Delta}, \dots, [\mathbf{b}_q]_{n_k+L_t+L-\Delta}]^T; \quad (7.13)$$

and \mathbf{C}_q is an $(L_t + L - 2\Delta + 1) \times (L_t + 2L - 2\Delta + 1)$ Toeplitz matrix

$$\mathbf{C}_q = \begin{bmatrix} c_{q,L} & \cdots & c_{q,0} & \mathbf{0} \\ & \ddots & & \\ \mathbf{0} & & c_{q,L} & \cdots & c_{q,0} \end{bmatrix}. \quad (7.14)$$

Accordingly, the matrix $\mathbf{C}_q^{(p)}$ consists of the columns of \mathbf{C}_q that are indexed from $L - \Delta + 1$ until $L - \Delta + L_t + 1$, and the remaining columns comprise $\mathbf{C}_q^{(d)}$. It is more convenient to rewrite the last equality in (7.12) as

$$\mathbf{y}_k = \mathcal{P}_k \mathbf{c} + \boldsymbol{\xi}_k + \mathbf{v}_k, \quad (7.15)$$

where

$$\mathbf{c} := [c_{0,0}, \dots, c_{0,L}, \dots, c_{Q,0}, \dots, c_{Q,L}]^T, \quad (7.16)$$

and

$$\mathcal{P}_k := \mathbf{Z}_k \mathbf{U}_k, \quad (7.17)$$

with

$$\begin{aligned} \mathbf{Z}_k &:= [\mathcal{D}\{\mathbf{b}_{k,0}\} \quad \cdots \quad \mathcal{D}\{\mathbf{b}_{k,Q}\}], \\ \mathbf{U}_k &:= \mathbf{I}_{Q+1} \otimes \mathbf{T}_k, \\ \mathbf{T}_k &:= \begin{bmatrix} [\mathbf{t}_k]_{\Delta} & \cdots & [\mathbf{t}_k]_0 & \mathbf{0} \\ \vdots & & \ddots & \\ [\mathbf{t}_k]_L & \cdots & & [\mathbf{t}_k]_0 \\ \vdots & \ddots & & \vdots \\ [\mathbf{t}_k]_{L_t} & \cdots & & [\mathbf{t}_k]_{L_t-L} \\ & \ddots & & \vdots \\ \mathbf{0} & & [\mathbf{t}_k]_{L_t} & \cdots & [\mathbf{t}_k]_{L_t-\Delta} \end{bmatrix}. \end{aligned} \quad (7.18)$$

Stacking the results in (7.15) for $k = 0, \dots, K$, we obtain

$$\mathbf{y} = \mathcal{P}\mathbf{c} + \boldsymbol{\xi} + \mathbf{v}, \quad (7.19)$$

where

$$\mathcal{P} := \mathbf{Z}\mathbf{U}, \quad (7.20)$$

with

$$\mathbf{Z} := \begin{bmatrix} \mathbf{Z}_0 & & \\ & \ddots & \\ & & \mathbf{Z}_K \end{bmatrix},$$

$$\mathbf{U} = [\mathbf{U}_0^T \quad \dots \quad \mathbf{U}_K^T]^T.$$

7.4 Channel Estimation

In the following, we will use three different channel estimators to attain the knowledge of \mathbf{c} : the LMMSE estimator is a Bayesian approach, which relies on the statistics of \mathbf{c} , while the LS estimator and the BLUE treat \mathbf{c} as a deterministic variable. Due to the analogies to channel estimation for OFDM as mentioned in the introduction, we will see that the channel estimators for a single-carrier system will admit the same expressions as those presented in Chapter 4, though some of the symbols will have a different implication here.

Like in Chapter 3, the following assumptions are adopted

Assumption 7.1. We assume the channel to be a wide-sense stationary uncorrelated scattering (WSSUS) process, i.e.,

$$\mathcal{E}_h\{h_{p,l}h_{p-m,l-n}\} = \sigma_l^2\gamma_m\delta_n, \quad (7.21)$$

where δ_n denotes the Kronecker delta, σ_l^2 the variance of the l th channel tap, and γ_m the normalized time correlation.

Assumption 7.2. The noise \mathbf{v} is assumed to be zero-mean white with variance σ^2 , i.e.,

$$\mathbf{R}_v := \mathcal{E}_v\{\mathbf{v}\mathbf{v}^H\} = \sigma^2\mathbf{I}. \quad (7.22)$$

Assumption 7.3. The data symbols \mathbf{d} are assumed to be zero-mean white with variance σ_s^2 , and uncorrelated with the noise \mathbf{v} , i.e.,

$$\mathcal{E}\{\mathbf{d}\mathbf{d}^H\} = \sigma_s^2\mathbf{I}, \quad (7.23)$$

and

$$\mathcal{E}\{\mathbf{d}\mathbf{v}^H\} = \mathbf{0}. \quad (7.24)$$

7.4.1 The LMMSE Estimator

The LMMSE estimator treats \mathbf{c} as a stochastic variable. To be more specific, we introduce the following assumption:

Assumption 7.4. *The channel vector \mathbf{c} is assumed to be uncorrelated with the noise \mathbf{v} and the data symbols \mathbf{d} , i.e.,*

$$\mathcal{E}\{\mathbf{c}\mathbf{v}^H\} = \mathbf{0}, \quad (7.25)$$

and

$$\mathcal{E}\{\mathbf{c}\mathbf{d}^H\} = \mathbf{0}. \quad (7.26)$$

Its covariance matrix

$$\mathbf{R}_c := \mathcal{E}\{\mathbf{c}\mathbf{c}^H\} \quad (7.27)$$

is assumed to be known (see (4.68) for its computation).

We seek a linear filter \mathbf{F} such that the MSE between the estimated BEM coefficients

$$\hat{\mathbf{c}} = \mathbf{F}\mathbf{y} \quad (7.28)$$

and the true BEM coefficients \mathbf{c} is minimal. In other words, we solve

$$\mathbf{F}_{\text{LMMSE}} = \arg \min_{\{\mathbf{F}\}} \text{trace}\{\mathcal{E}_{\mathbf{c},\mathbf{d},\mathbf{v}}\{(\mathbf{F}\mathbf{y} - \mathbf{c})(\mathbf{F}\mathbf{y} - \mathbf{c})^H\}\}. \quad (7.29)$$

Similarly to (4.36), it can be shown that

$$\mathcal{E}_{\mathbf{c},\mathbf{d},\mathbf{v}}\{(\mathbf{F}\mathbf{y} - \mathbf{c})(\mathbf{F}\mathbf{y} - \mathbf{c})^H\} = \mathbf{F}(\mathcal{P}\mathbf{R}_c\mathcal{P}^H + \mathbf{R}_\xi + \mathbf{R}_v)\mathbf{F}^H - 2\Re\{\mathbf{R}_c\mathcal{P}^H\mathbf{F}^H\} + \mathbf{R}_c. \quad (7.30)$$

In deriving the above, we make use of Assumptions 7.3 and 7.4, and introduce the covariance matrix for the interference term ξ :

$$\mathbf{R}_\xi := \mathcal{E}_{\mathbf{c},\mathbf{d}}\{\xi\xi^H\}. \quad (7.31)$$

The computation of \mathbf{R}_ξ is different from that in Chapter 4, as we can observe that due to Assumption 7.3, \mathbf{R}_ξ becomes here a block-wise diagonal matrix

$$\mathbf{R}_\xi = \mathbf{I}_{K+1} \otimes \mathbf{R}_{\xi_k}. \quad (7.32)$$

Conform Assumption 7.1, each diagonal block \mathbf{R}_{ξ_k} is again diagonal with the n th diagonal entry given by

$$[\mathbf{R}_{\xi_k}]_{n,n} = \begin{cases} \sum_{l=\Delta+n}^L \sigma_l^2 & \text{if } n \leq L - \Delta - 1, \\ \sum_{l=0}^{n-L_t+\Delta-1} \sigma_l^2 & \text{if } n \geq L_t - \Delta + 1, \\ 0 & \text{otherwise.} \end{cases} \quad (7.33)$$

Substituting (7.30) in (7.29), and forcing its derivative with respect to \mathbf{F} to zero, we obtain

$$\mathbf{F}_{\text{LMMSE}} = \mathbf{R}_c\mathcal{P}^H(\mathcal{P}\mathbf{R}_c\mathcal{P}^H + \mathbf{R}_\xi + \mathbf{R}_v)^{-1}. \quad (7.34)$$

7.4.2 The Least Squares Estimator

The Least Squares (LS) estimator \mathbf{F}_{LS} treats \mathbf{c} as a deterministic variable. It is straightforward to obtain

$$\mathbf{F}_{\text{LS}} := \mathcal{P}^\dagger \quad (7.35)$$

such that

$$\hat{\mathbf{c}}_{\text{LS}} = \mathbf{c} + \mathcal{P}^\dagger(\boldsymbol{\xi} + \mathbf{v}). \quad (7.36)$$

The LS estimator is the most robust estimator, in the sense that it is not sensitive to a mismatch of the channel and noise statistics. However, it performs inferior when the interference (and noise) is prominent, especially in the case if \mathcal{P} is ill-conditioned or even rank-deficient.

7.4.3 An Iterative BLUE

From (7.19), we can find an expression for the BLUE following similar steps as in [48, Appendix 6B] by treating the interference $\boldsymbol{\xi}$ and noise \mathbf{v} as a single disturbance term such that

$$\begin{aligned} \hat{\mathbf{c}}_{\text{BLUE}} &= \mathbf{F}_{\text{BLUE}} \mathbf{y}, \\ \mathbf{F}_{\text{BLUE}} &= (\mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{c}) \mathcal{P})^{-1} \mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{c}), \end{aligned} \quad (7.37)$$

where $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$ denotes the covariance matrix of the disturbance with \mathbf{c} again viewed as a deterministic variable:

$$\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c}) := \mathcal{E}_{\mathbf{d}, \mathbf{v}} \{ (\boldsymbol{\xi} + \mathbf{v})(\boldsymbol{\xi} + \mathbf{v})^H \}. \quad (7.38)$$

The computation of $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$ is different than in Chapter 4: due to Assumption 7.2 and Assumption 7.3, we understand that $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$ is a block-wise diagonal matrix

$$\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c}) = \begin{bmatrix} \tilde{\mathbf{R}}_{\xi_0}(\mathbf{c}) & & \\ & \ddots & \\ & & \tilde{\mathbf{R}}_{\xi_K}(\mathbf{c}) \end{bmatrix} + \sigma^2 \mathbf{I}, \quad (7.39)$$

with $\tilde{\mathbf{R}}_{\xi_k}(\mathbf{c})$ standing for the covariance matrix of $\boldsymbol{\xi}_k$, which treats only the data symbol \mathbf{d}_k as random [c.f. (7.33)]

$$\begin{aligned} \tilde{\mathbf{R}}_{\xi_k}(\mathbf{c}) &:= \mathcal{E}_{\mathbf{d}_k} \{ \boldsymbol{\xi}_k \boldsymbol{\xi}_k^H \}, \\ &= \sigma_s^2 \left(\sum_{q=0}^Q \mathcal{D}\{\mathbf{b}_{p,q}\} \mathbf{C}_q^{(d)} \right) \left(\sum_{q=0}^Q \mathcal{D}\{\mathbf{b}_{p,q}\} \mathbf{C}_q^{(d)} \right)^H. \end{aligned} \quad (7.40)$$

The equation in (7.37) is not solvable in closed form since its computation entails the knowledge of \mathbf{c} itself. A recursive approach can be therefore applied: suppose at the i th iteration, an estimate for \mathbf{c} is available denoted as $\hat{\mathbf{c}}_{\text{BLUE}}^{(i)}$. Next, we use this estimate to update the covariance matrix $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{c})$ using (7.40). The update of $\hat{\mathbf{c}}$ in turn is used to produce the BLUE for the next iteration and so on:

$$\begin{aligned}\mathbf{F}_{\text{BLUE}}^{(i+1)} &= (\mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\hat{\mathbf{c}}_{\text{BLUE}}^{(i)}) \mathcal{P})^{-1} \mathcal{P}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\hat{\mathbf{c}}_{\text{BLUE}}^{(i)}), \\ \hat{\mathbf{c}}_{\text{BLUE}}^{(i+1)} &= \mathbf{F}_{\text{BLUE}}^{(i+1)} \mathbf{y}.\end{aligned}$$

To initialize, we can simply let

$$\hat{\mathbf{c}}_{\text{BLUE}}^{(0)} = \mathbf{0}, \quad (7.41)$$

which results in the following expression for the first iteration:

$$\mathbf{F}_{\text{BLUE}}^{(1)} = \mathcal{P}^\dagger. \quad (7.42)$$

The resulting $\hat{\mathbf{c}}_{\text{BLUE}}^{(1)}$ is actually the least-squares fit as obtained in the previous section.

Note that the BLUE in the single-carrier system resembles to the “weighted LS estimator” in [78].

7.5 Channel Identifiability

Like in Chapter 5, we define channel identifiability in a single-carrier system as the full column-rank condition of the matrix \mathcal{P} . Such a full rank \mathcal{P} leads to a unique channel estimator for the LS estimator and the BLUE. It is also significant to the performance of the LMMSE estimator in the moderate-to-high SNR regime. We will illustrate this effect in the simulation part with some numerical examples.

To begin with, let us first adopt the assumption on the length of each pilot cluster.

Assumption 7.5. *The length of each pilot cluster L_t is assumed to be*

$$L_t \geq 2\Delta. \quad (7.43)$$

Assumption 7.5 implies that the $(L_t + L - 2\Delta + 1) \times (L + 1)$ matrix \mathbf{T}_k defined in (7.18) is ‘tall’. Based on this assumption, we can establish the following fact.

Theorem 7.1. *Under Assumption 7.5 and an arbitrary BEM that is defined in Chapter 2, \mathcal{P} will be of full rank if the following conditions on the pilots are satisfied.*

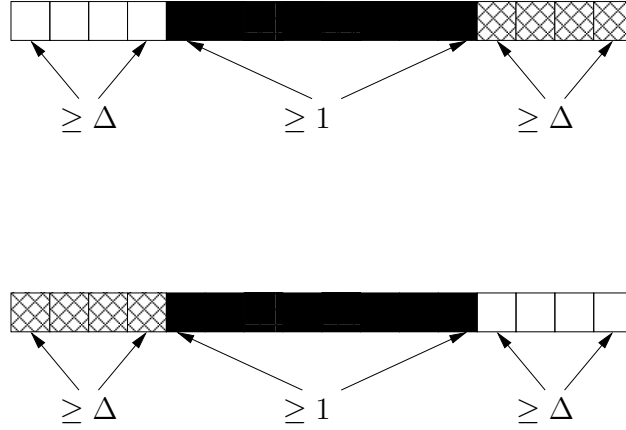


Figure 7.2: Examples of the proposed pilot structure: the white boxes stands for the area where zero pilots are located; the black boxes for the area where non-zero pilots are located; the hatched boxes for the area where the arbitrary pilots (zero or non-zero) are located.

(C1.) Inside each pilot cluster, at least either the first or the last Δ pilots are zeros. In addition, we suppose that each pilot cluster has at least one non-zero pilot that is located between the Δ th and $(L_t - \Delta)$ th position.

We refer to Fig. 7.2 for an illustration of such a pilot structure.

(C2.) The total number of such pilot clusters is greater or equal to the BEM scale, i.e.,

$$K + 1 \geq Q + 1. \quad (7.44)$$

We will give here only a sketch of the proof of Theorem 7.1 (for details see Appendix 7A). Due to Assumption 7.5 and supposing that the first Δ pilots of each pilot cluster are zeros, \mathbf{T}_k becomes a “tall” banded matrix with zeros above its main diagonal[‡]. Obviously, the matrix \mathcal{P} is a stack of $K + 1$ by $Q + 1$ of such banded submatrices (each weighted by a diagonal matrix). By changing the order of the rows and columns of \mathcal{P} , we can obtain a new matrix which has a block-wise banded structure. This process is illustrated in Fig. 7.3, where the blocks with the same pattern are put together. It can be shown that the full-rank condition of \mathcal{P} relies on the full-rank condition of the blocks (circled in Fig. 7.3) on the main diagonal of the permuted matrix.

Remark 7.1.

[‡]For an $M \times N$ tall matrix \mathbf{A} , its main diagonal is defined as the vector $[[\mathbf{A}]_{0,0}, \dots, [\mathbf{A}]_{N-1,N-1}]^T$.

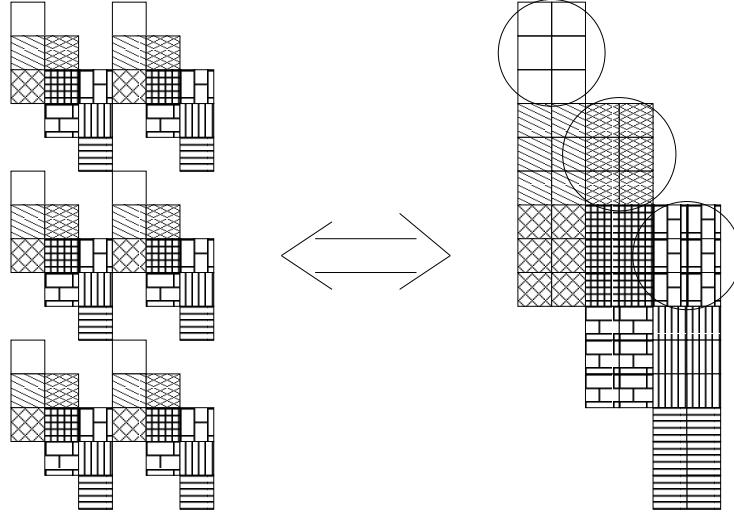


Figure 7.3: Changing a stack of banded matrices to a block-wise banded matrix by row- and column-interleaving.

Assumption 7.5 is key to Theorem 7.1. For the case other than Assumption 7.5 where $L_t < 2\Delta$, the special structure of \mathcal{P} depicted in Fig. 7.3 is ruined. To be more specific, $\mathbf{T}(k)$ will have not only zeros above its main diagonal. Consequently, the channel identifiability will not be guaranteed, and will depend on the specific value of the pilots and the choice of a particular BEM.

Remark 7.2.

In [59], so-called “TDKD” pilot clusters are considered (its frequency-domain variant “FDKD” is discussed in Chapter 5). In each cluster of the “TDKD” pilots, a non-zero pilot is guarded by zero pilots on both sides. This is a special case of the pilot structure defined in **(C1)** with $\Delta = L$ and $L_t \geq 2L$. In that case, we know from **(C2)** that the channel is identifiable for $K \geq Q$ as also argued in [59].

Remark 7.3.

Note that the parameter Δ is the choice of the receiver. We will show in the simulation that $\Delta = 0$ is optimal for the performance of the LMMSE estimator and the BLUE. In that case, the constraint on the pilot structure is very loose: the channel is always identifiable as long as there are more than Q non-zero pilot clusters. The length of such pilot clusters is not even important.

7.6 Simulation Results

In this section, we examine the performance of the channel estimators that are discussed in the previous section. We let the TV channel follow the Jakes' Doppler spectrum [45], which can be simulated by the channel generator proposed in [124].

The channel's time-variation is characterized by the normalized Doppler spread $\nu_D = 0.002$. We use a DPS-BEM to approximate the TV channel, which spans a window size $N = 256$ and has $Q + 1 = 3$ basis expansion functions. Further, we assume the channel to be an FIR filter with $L + 1$ taps. We will examine two cases, $L = 4$ and $L = 10$. The channel taps are generated independently with a uniform power delay profile, i.e., each tap has the same power $\frac{1}{L+1}$.

Pilots are inserted in the time domain for channel estimation. Within a block of 256 transmitted symbols, 176 are QPSK data symbols. The remaining 80 symbols are reserved for pilots, which are grouped in $K + 1 = 5$ clusters, each cluster containing $L_t + 1 = 16$ pilots. In the case where Assumption 7.5 is applicable, we construct the pilots following Theorem 7.1, where the non-zero pilots are randomly picked constant-modulus symbols[§]. They are further amplified such that the average pilot power equals the data symbol power. In the case where Assumption 7.5 is not applicable, we let all the pilots take randomly picked constant-modulus symbols.

The LMMSE estimator is equipped with the perfect knowledge of the channel statistics. Its performance is depicted in Fig. 7.4, where the results are attained for different channel lengths. As argued in the previous section, the number of observation samples could lead to different channel estimates. This is reflected by the choice of Δ . As one can observe in Fig. 7.4, a smaller Δ can improve the performance, especially for a longer channel. In that case, the noise is better averaged out by processing more observation samples in the channel estimator. Although more interference due to the unknown data is thereby induced, it is efficiently counteracted by the LMMSE estimator.

A different behavior is exhibited by the LS estimator, whose performance is plotted in Fig. 7.5. The estimator based on a smaller Δ performs only better at low-to-moderate SNR. In that area, since the noise plays a more dominant role, it is beneficial to employ more observation samples. However, when the SNR increases, the interference becomes more pronounced, and the LS estimator suffers from a high noise floor. The noise floor can only be reduced by adopting a larger Δ .

The behavior of the BLUE is analogous to that of the LMMSE estimator as can be seen in Fig. 7.6. Note that the case $\Delta = L$ is not shown, because the BLUE in that case just reduces to the LS estimator.

[§]We did not use "TDKD" pilots because they perform better for short channels where $L_t \geq 2L$ but worse for long channels.

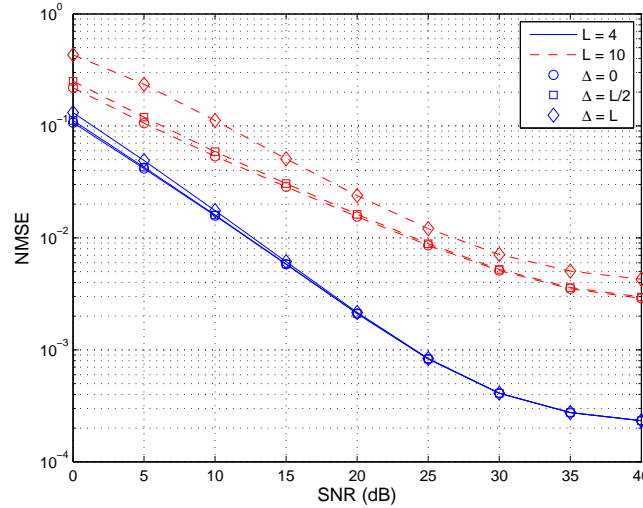


Figure 7.4: The performance of the LMMSE estimator. Solid curves $L = 4$; Dashed curves $L = 10$.

In Fig. 7.7, we compare the best performance of the LMMSE estimator, the LS estimator and the BLUE to each other. We find that for a short channel (or equivalently, if the pilot cluster is long enough), the three estimators perform similarly, whereas for a long channel, the LMMSE estimator and the BLUE are much more accurate than the LS estimator. Compared to the LMMSE estimator, the BLUE is more appealing because except for the statistical knowledge of the noise and data symbols, it does not require the channel statistics, which are in practice hard to attain especially in the case of a TV channel [106]. On the other hand, the BLUE is computationally more complicated due to the iterations.

7.7 Summary

We discussed in this chapter how to acquire the channel state information in a single-carrier system by means of time-domain multiplexed pilots. Analogous to the OFDM case, the channel is approximated by an arbitrary BEM that is tight with respect to the realistic TV channel.

The pilots are grouped in clusters. For such clustered pilots, we first formulated their relationship with the observation samples in terms of the BEM coefficients. The number of observation samples can actually be selected: the more observation

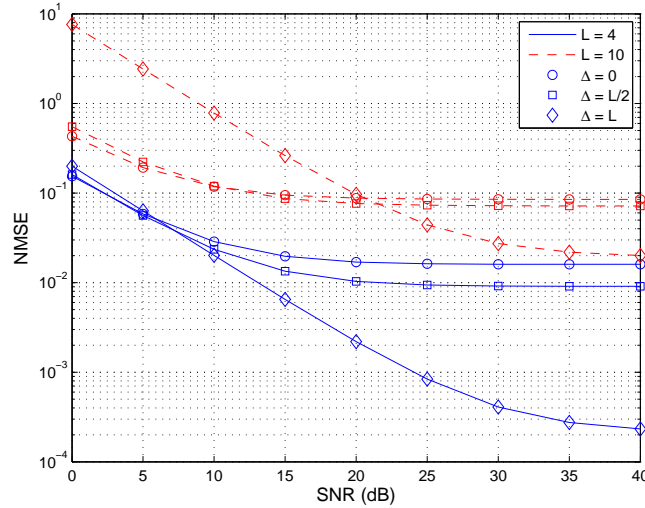


Figure 7.5: The performance of the LS estimator. Solid curves $L = 4$; Dashed curves $L = 10$.

samples are considered, the more pilot-related information can be included. On the other hand, the more data-related interference will be induced. Three types of channel estimators were discussed, the LMMSE estimator, the LS estimator and the BLUE. The interference is explicitly taken into account in the design of the LMMSE estimator and the BLUE.

For these channel estimators, we have established channel identifiability, which can be applied to a general BEM assumption.

We have seen from the simulation that for the LMMSE estimator and the BLUE, a higher performance is achieved by employing more observation samples. In this case, the noise can be better averaged out while the interference is efficiently counteracted. For the LS estimator, however, employing more observation samples is only beneficial at low-to-moderate SNR. At high SNR, this will lead to a higher noise floor.

The channel estimation in a single-carrier system can be viewed as a counterpart to that in OFDM. For instance, the channel estimators in both cases bear the same expression; the strategy of choosing which observation samples for respective channel estimators is similar. Except for these analogies, the channel estimation in a single-carrier system could be somewhat simpler than in OFDM. This is mainly because the channel matrix in the time domain is always strictly banded due to

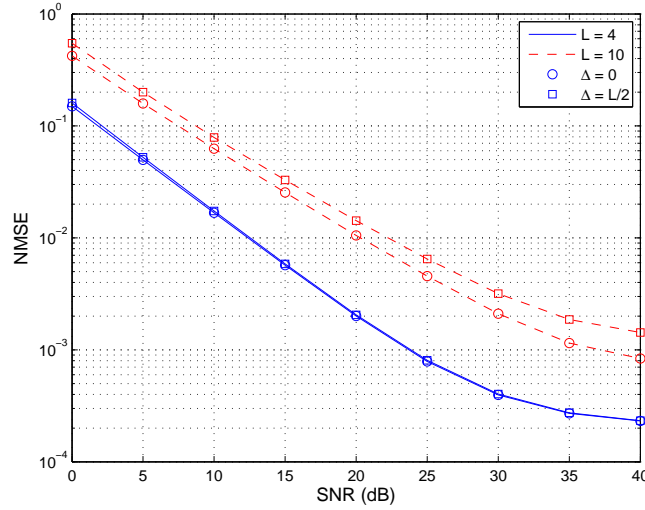


Figure 7.6: The performance of the BLUE. Solid curves $L = 4$; Dashed curves $L = 10$.

the FIR assumption while its counterpart in the frequency domain is only approximately banded. As a result, the pilot power is more dissipated in OFDM (to the whole OFDM symbol). The signal to inference ratio could therefore be lower in OFDM. For the LMMSE estimator and the BLUE, more observation samples must accordingly be taken into account in OFDM, which increases computational complexity. Also, the conditions on channel identifiability in a single-carrier system are much less stringent than in OFDM. The pilot cluster in a single-carrier system is required only to be longer than a design parameter Δ , whose value itself is at the choice of the receiver. For the LMMSE estimator and BLUE, Δ can be even chosen to be zero. From the point of view of channel identifiability, this could allow us to employ fewer pilots in a single-carrier system.

Appendix 7A: Proof of Theorem 7.1

To begin with, let us first use an $N \times (Q + 1)$ matrix \mathbf{B} to stand for a BEM matrix, which is introduced in Chapter 2; \mathcal{S} stands for a set consisting of M unique elements, which range from 0 until $N - 1$. $\mathbf{B}^{\{\mathcal{S}\}}$ is hence an $M \times (Q + 1)$ matrix comprised of rows of \mathbf{B} whose indices are collected in the set \mathcal{S} . We assume that

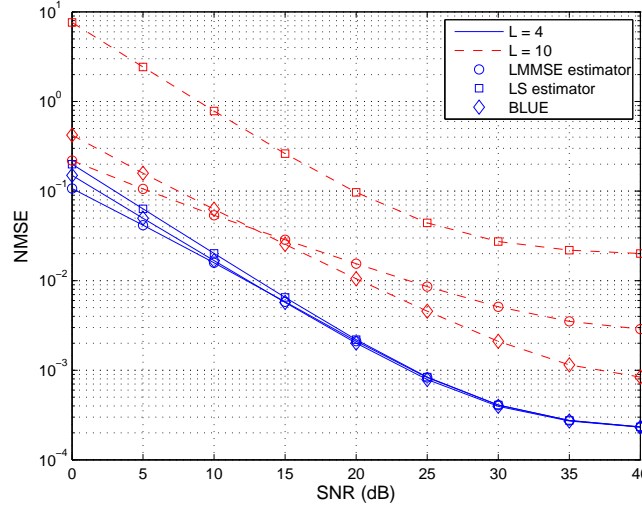


Figure 7.7: Comparison of different channel estimators. Solid curves $L = 4$; Dashed curves $L = 10$.

Assumption 7.6. $\mathbf{B}^{\{S\}}$ is of full rank $Q + 1$ if $M \geq Q + 1$.

Assumption 7.6 holds in general for the CE-BEM and P-BEM, in which case $\mathbf{B}^{\{S\}}$ is a ‘tall’ Vandermonde matrix. For the other BEMs, this is true via computer validation.

To prove Theorem 7.1, we need also the following lemma.

Lemma 7.1. Let us use \mathbf{A} to denote a matrix that has the following block structure

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{0} & \cdots & \mathbf{0} \\ \diamond & \mathbf{A}_{1,1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \mathbf{A}_{N-1,N-1} \\ \diamond & \cdots & \cdots & \diamond \end{bmatrix}, \quad (7.45)$$

where $\mathbf{A}_{n,n}$ stands for a $P \times M$ ‘tall’ submatrix ($P \geq M$); \diamond stands for some arbitrary block entries. Then \mathbf{A} will be of full rank MN if and only if each matrix $\mathbf{A}_{n,n}$ is of full rank M for $n = 0, \dots, N - 1$.

Proof. We prove by contradiction. Suppose all $\mathbf{A}_{n,n}$ ’s are of full rank but \mathbf{A} is not of

full rank. Then, there exists a non-zero vector

$$\mathbf{g} := [\mathbf{g}_0^T \quad \cdots \quad \mathbf{g}_{N-1}^T]^T, \quad (7.46)$$

with \mathbf{g}_n representing an $M \times 1$ vector, such that

$$\boldsymbol{\xi} = \mathbf{A}\mathbf{g} = \mathbf{0}. \quad (7.47)$$

Observing the structure of \mathbf{A} in (7.45), in order for the first P elements of $\boldsymbol{\xi}$ to be zeros, we require $\mathbf{g}_0 = \mathbf{0}$ for the full-rank $\mathbf{A}_{0,0}$. Similarly, in order for the next P elements of $\boldsymbol{\xi}$ to be zeros, we require $\mathbf{g}_1 = \mathbf{0}$, which corresponds to the column position of the full-rank matrix $\mathbf{A}_{1,1}$. Repeating the argument above, we obtain

$$\mathbf{g}_0 = \cdots = \mathbf{g}_{N-1} = \mathbf{0}, \quad (7.48)$$

which contradicts the assumption that \mathbf{g} is not an all-zero vector. Therefore, we complete the proof. \square

To begin with the proof of Theorem 7.1, we first, without loss of generality, assume that the pilot structure satisfies the first case of (C1)[¶], where the position of the first non-zero element in each pilot cluster μ_k is assumed to be

$$\Delta \leq \mu_k \leq L_t - \Delta; \quad (7.49)$$

The subsequent entries of the pilot cluster are allowed to take an arbitrary value. As a result, the first $\mu_k - \Delta$ rows of the matrix \mathcal{P}_k are all zeros. Hence, we need only to concentrate on the remaining rows, which are collectively denoted by $\bar{\mathcal{P}}_k$ [c.f. (7.18)]

$$\bar{\mathcal{P}}_k := [\mathcal{D}\{\bar{\mathbf{b}}_{k,0}\}\bar{\mathbf{T}}_k \quad \cdots \quad \mathcal{D}\{\bar{\mathbf{b}}_{k,Q}\}\bar{\mathbf{T}}_k], \quad (7.50)$$

where

$$\bar{\mathbf{b}}_{k,q} := [[\mathbf{b}_q]_{n_k+\mu_k}, \cdots, [\mathbf{b}_q]_{n_k+L_t+L-\Delta}]^T,$$

$$\bar{\mathbf{T}}_k := \begin{bmatrix} [\mathbf{t}_k]_{\mu_k} & & \mathbf{0} \\ \diamond & \ddots & \\ \vdots & \ddots & [\mathbf{t}_k]_{\mu_k} \\ \diamond & \cdots & \diamond \end{bmatrix}.$$

Here, \diamond denotes an entry that is irrelevant at this point. For the sake of convenience, we let all the matrices $\bar{\mathcal{P}}_k$ contain the same number of rows by appending at the end $\mu_k - \mu_{\min}$ all-zero rows with

$$\mu_{\min} = \min(\mu_0, \cdots, \mu_K). \quad (7.51)$$

[¶]The proof for the other case of (C1) follows a similar approach.

Equivalently, we consider the rank condition of $\bar{\mathcal{P}}$ which is defined as

$$\bar{\mathcal{P}} := \left[\bar{\mathcal{P}}_0^T \quad \mathbf{0}_{(L+1)(Q+1) \times (\mu_0 - \mu_{\min})} \quad \cdots \quad \bar{\mathcal{P}}_K^T \quad \mathbf{0}_{(L+1)(Q+1) \times (\mu_K - \mu_{\min})} \right]^T. \quad (7.52)$$

In line with the structure of $\bar{\mathbf{T}}_k$, we understand that $\bar{\mathcal{P}}$ is a stack of $(L_t + L - \Delta + \mu_{\min} + 1) \times (L + 1)$ banded matrices both horizontally and vertically, as depicted in the left of Fig. 7.3. To explore such a structure, we interleave the columns of $\bar{\mathcal{P}}$ with an interleaving depth $L + 1$; at the same time, we interleave the rows of $\bar{\mathcal{P}}$ with an interleaving depth $L_t + L - \Delta + \mu_{\min} + 1$. As a result, we obtain

$$\mathbf{\Pi}_{L_t + L - \Delta + \mu_{\min} + 1}^T \bar{\mathcal{P}} \mathbf{\Pi}_{L+1} = \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{0} & \cdots & \mathbf{0} \\ \diamond & \mathbf{A}_{1,1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \mathbf{A}_{L,L} \\ \diamond & \cdots & \cdots & \diamond \end{bmatrix}, \quad (7.53)$$

where the $\mathbf{\Pi}_d$ denotes a depth- d interleaving matrix of proper dimensions as defined in (5.68); the $(K + 1) \times (Q + 1)$ matrix $\mathbf{A}_{l,l}$ can be shown to admit an expression as

$$\begin{aligned} \mathbf{A}_{l,l} &:= \begin{bmatrix} [\mathbf{b}_{0,0}]_{n_0 + \mu_0 + l} [\mathbf{t}_0]_{\mu_0} & \cdots & [\mathbf{b}_{0,Q}]_{n_0 + \mu_0 + l} [\mathbf{t}_0]_{\mu_0} \\ \vdots & \ddots & \vdots \\ [\mathbf{b}_{K,0}]_{n_K + \mu_K + l} [\mathbf{t}_K]_{\mu_K} & \cdots & [\mathbf{b}_{K,Q}]_{n_K + \mu_K + l} [\mathbf{t}_K]_{\mu_K} \end{bmatrix} \\ &= \mathcal{D}\{[\mathbf{t}_0]_{\mu_0}, \dots, [\mathbf{t}_K]_{\mu_K}\}^T \underbrace{\begin{bmatrix} [\mathbf{b}_{0,0}]_{n_0 + \mu_0 + l} & \cdots & [\mathbf{b}_{0,Q}]_{n_0 + \mu_0 + l} \\ \vdots & \ddots & \vdots \\ [\mathbf{b}_{K,0}]_{n_K + \mu_K + l} & \cdots & [\mathbf{b}_{K,Q}]_{n_K + \mu_K + l} \end{bmatrix}}_{\mathbf{\bar{B}}_l}. \end{aligned} \quad (7.54)$$

Because of (C1), $[\mathbf{t}_k]_{\mu_k} \neq 0$ for $k = 0, \dots, K$. Further, the $(K + 1) \times (Q + 1)$ matrix $\mathbf{\bar{B}}_l$ is carved out of the larger BEM matrix \mathbf{B} , and thus of full rank $Q + 1$ according to Assumption 7.6. Applying Lemma 7.1 on (7.53), we conclude the proof.

Chapter 8

MSE-Optimal Training for Time-Selective Channels

8.1 Introduction

For pilot-aided channel estimation, an interesting topic is what is the ‘best’ pilot structure in terms of a certain optimality criterion. It is quite common in a practical communication system that the bandwidth and power consumption dedicated for pilots are constrained. A judicious design of the pilot symbols and their positions can lead to a significant “potential gain” [109]. For instance, [57] consider the case in of a single-carrier system where the pilots are optimized in terms of the channel estimator’s MSE. The OFDM case is tackled in [69]; the MIMO case for a single-carrier system is considered in [62] and the MIMO case for OFDM is considered in [6]. [26] loosens the constraint on the pilot cluster size and optimizes the pilot structure in terms of the CRB of the channel estimator. The design of CRB-optimal pilots in the presence of a precoder is treated in [73, 115]. There are also works that strike on the impact of the pilot structure on channel capacity [2, 40, 74].

Compared to the time-invariant channel case, there are, however, much fewer works that cast attention to the optimality of pilots on time-varying channels. [27] studies time-selective channels, and claims that the periodic single pilot minimizes the maximum steady-state channel MSE, whereas in the practical SNR regime, the superimposed training is more beneficial to both the MSE and BER performance. For a doubly-selective channel, [47] argues that the “Time-Domain Kronecker Delta” (TDKD) pilot structure that is also used in [59] is optimal in the MSE of the channel estimator for a single-carrier system, and its variant, the “Frequency-Domain Kronecker Delta” (FDKD) pilot structure, is optimal for an OFDM system. It is worth underlying that [27] approximates the time-varying channel with a first-order Gauss-Markov model, while [59, 47] uses the (C)CE-BEM. Both models can induce quite a large modeling error for fast-varying channels, which is nonetheless neglected in the analysis. Note that the positions of the pilots can play an important

The results of this chapter appeared in [101].

role in the modeling error: a well-known example can be encountered for the P-BEM, for which the maximum modeling error is minimized if the pilots are placed according to the roots of the Chebyshev polynomials [50].

By adopting a tighter BEM other than the (C)CE-BEM, we are allowed in principle to bypass the impact of the modeling error, but the optimal pilots for these BEMs remain unsolved. Due to the lack of insight, many existing channel estimation works just adopt the classical equi-distant/powerful pilot scheme like in e.g., [10, 122, 78] without arguing its optimality. In this chapter, we will try to cast some light on how to attain the optimal pilots to minimize the channel MSE for time-selective channels via a numerical approach. A time-selective channel is a simpler version of a doubly-selective channel, where the channel has only one time-varying tap. The time-variation of the channel tap will be modeled by a tight BEM such as reported in [122], where most of the BEMs (except the (C)CE-BEM) inflicts a BEM modeling error lower than -40dB , whose impact is indeed negligible. This means that the proposed approach can be applied for these BEMs. Focusing on an LS channel estimator, we will first formulate the problem as the minimization of an MSE-related cost function, which is next transformed into a mixed-integer non-linear programming problem [31]. The simulation results show that the resulting pilot structure can produce a lower channel MSE than the classical equi-distant/powerful pilots if a general BEM assumption is adopted.

8.2 Data Model for Time-Selective Channel Estimation

Let us consider a communication system over a time-selective channel. Because the channel considered in this chapter has only one tap in the time domain (instantaneous channels), we drop the tap index l in the channel notation, $h_{n,l}$, which is used in the previous chapters, and just use h_n to stand for the channel gain at the n th time-instance. The I/O relationship in discrete form can then be expressed as

$$y_n = h_n s_n + v_n, \quad (8.1)$$

where y_n , s_n , and v_n denote respectively the received signal, the transmitted signal and the noise at the n th time-instance.

To approximate the time-variation with a BEM, let us choose the time instants $n = 0, \dots, N-1$ as an observation window and collect all the channel gains within this observation window in the vector

$$\mathbf{h} := [h_0, \dots, h_{N-1}]^T. \quad (8.2)$$

By selecting a proper scale Q for the BEM, we can closely fit the channel time variation within the observation window with $Q + 1$ BEM coefficients:

$$\mathbf{h} = \mathbf{B}\mathbf{c} + \boldsymbol{\delta}, \quad (8.3)$$

where \mathbf{B} is an $N \times (Q + 1)$ BEM matrix, and \mathbf{c} a $(Q + 1)$ -long vector containing the related BEM coefficients. $\boldsymbol{\delta}$ represents the BEM modeling error. In the sequel, we will assume that by choosing an appropriate BEM, the modeling error can be kept sufficiently small [122]. As a result, its impact in a practical SNR region can be neglected. It is also notable that the idea of a BEM is reflected in some other applications without inducing a modeling error, e.g., the frequency-domain channel in an OFDM system can be model by a (C)CE-BEM [69]. In this case, \mathbf{h} corresponds to the channel vector in the frequency domain, and \mathbf{B} corresponds to the (C)CE-BEM,

$$[\mathbf{B}]_{n,q} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}nq}, \quad (8.4)$$

with N denoting the total number of subcarriers. If the frequency-domain channel must abide with some virtual subcarriers in an OFDM system [32], then \mathbf{h} can be viewed as a stack of all the non-virtual subcarriers, and \mathbf{B} will correspond to an (O)CE-BEM:

$$[\mathbf{B}]_{n,q} = \frac{1}{\sqrt{N'}} e^{j\frac{2\pi}{N'}nq}, \quad (8.5)$$

with N' denoting the total number of subcarriers, which is larger than N . For these cases, the approach discussed in this chapter can be also applied.

To estimate the BEM coefficients \mathbf{c} , we resort to a time-multiplexed training scheme, where the pilots are interleaved with data symbols in the time domain. For time-selective channels, the channel estimation and equalization can thus be completely decoupled. Suppose there are K pilots transmitted during the observation window, whose positions are collected in the set

$$\mathcal{G} := \{g_0, g_1, \dots, g_{K-1}\}. \quad (8.6)$$

Furthermore, we assume that the k th pilot s_{g_k} has phase ϕ_k and power P_k , i.e.,

$$s_{g_k} = \sqrt{P_k} e^{j\phi_k}, \quad (8.7)$$

with the total power assigned for training

$$P = \sum_k P_k. \quad (8.8)$$

With the aid of these notations and neglecting the modeling error, we can express the I/O relationship that results from the pilots in matrix/vector form as

$$\mathbf{y}_p = \mathbf{\Phi}_p \sqrt{\mathcal{P}_p} \mathbf{B}_p \mathbf{c} + \mathbf{v}_p, \quad (8.9)$$

where \mathbf{y}_p is a $K \times 1$ vector collecting the received samples at the pilot-related positions, i.e.,

$$\mathbf{y}_p := [y_{g_0}, \dots, y_{g_{K-1}}]^T; \quad (8.10)$$

\mathbf{v}_p is similarly defined as \mathbf{y}_p ; \mathcal{P}_p and $\mathbf{\Phi}_p$ are both $K \times K$ diagonal matrices

$$\mathcal{P}_p := \mathcal{D}\{[P_0, \dots, P_{K-1}]^T\}, \quad (8.11)$$

$$\mathbf{\Phi}_p := \mathcal{D}\{[e^{j\phi_0}, \dots, e^{j\phi_{K-1}}]^T\}. \quad (8.12)$$

\mathbf{B}_p is a $K \times (Q+1)$ matrix consisting of K rows carved out of \mathbf{B} corresponding to the pilot positions

$$\mathbf{B}_p := \mathbf{B}^{\{\mathcal{G}\}}. \quad (8.13)$$

An LS estimate from (8.9) can be obtained as

$$\hat{\mathbf{c}} = (\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1} \mathbf{B}_p^H \sqrt{\mathcal{P}_p} \mathbf{\Phi}_p^H \mathbf{y}_p, \quad (8.14)$$

which results in a channel MSE equal to

$$\text{MSE} = \mathcal{E}\{\|(\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1} \mathbf{B}_p^H \sqrt{\mathcal{P}_p} \mathbf{\Phi}_p^H \mathbf{v}_p\|^2\}. \quad (8.15)$$

If we assume that the noise \mathbf{v}_p is zero-mean white with variance σ^2 , i.e.,

$$\mathcal{E}\{\mathbf{v}_p \mathbf{v}_p^H\} = \sigma^2 \mathbf{I}_K, \quad (8.16)$$

the channel MSE will be independent of the phase of the pilots as can be observed from

$$\text{MSE} = \sigma^2 \text{tr}\{(\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1}\}. \quad (8.17)$$

In this way, we only need to focus on the powers and positions of the pilots. The optimization problem can thus be formulated as

$$\begin{aligned} \arg \min_{\{\mathcal{G}, \mathcal{P}_p\}} \text{MSE}(\mathcal{G}, \mathcal{P}_p) &= \arg \min_{\{\mathcal{G}, \mathcal{P}_p\}} \sigma^2 \text{tr}\{(\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1}\} \\ \text{s.t. } \mathbf{B}_p &= \mathbf{B}^{\{\mathcal{G}\}}, \mathcal{G} \subset \{0, \dots, N-1\} \text{ and } |\mathcal{G}| = K, \\ \mathcal{P}_p &\text{ positive diagonal with } \text{tr}\{\mathcal{P}_p\} = P. \end{aligned} \quad (8.18)$$

$$\mathcal{P}_p \text{ positive diagonal with } \text{tr}\{\mathcal{P}_p\} = P. \quad (8.19)$$

8.3 Optimization Algorithm

The above formulation is a mixed-integer non-linear optimization problem, where we have posed restrictions on the total number and power of the pilots. [31] provides an algorithm known as the Generalized Benders Decomposition (GBD). In a nutshell, the GBD iteratively projects the minimization problem onto the \mathcal{P}_p – space (primal problem) and the \mathcal{G} – space (relaxed master problem). Since the constraints in (8.18) and (8.19) are separable in \mathcal{G} and \mathcal{P}_p , the primal problem becomes simply the search for the optimal \mathcal{P}_p subject to a fixed \mathcal{G} , and the relaxed master problem becomes the search for the optimal \mathcal{G} subject to a fixed \mathcal{P}_p . The GBD algorithm is summarized in Table 8.1:

Table 8.1: *The GBD algorithm.*

-
1. We use the superscript (k) to denote the iteration index. To initialize, let us set $k = 0$, and assume $\mathcal{G}^{(0)} = \{0, \dots, K/2 - 1, N - K/2, \dots, N - 1\}$ and $\mathcal{P}_p^{(0)} = \frac{P}{K} \mathbf{I}_K$.
 2. For a fixed $\mathcal{P}_p^{(k)}$, solve $v_{\text{master}}^{(k+1)} = \min_{\{\mathcal{G}\}} \text{MSE}(\mathcal{G}, \mathcal{P}_p^{(k)})$ subject to (8.18) resulting in $\mathcal{G}^{(k+1)}$.
 3. For a fixed $\mathcal{G}^{(k+1)}$, solve $v_{\text{primal}}^{(k+1)} = \min_{\{\mathcal{P}_p\}} \text{MSE}(\mathcal{G}^{(k+1)}, \mathcal{P}_p)$ subject to (8.19) resulting in $\mathcal{P}_p^{(k+1)}$.
 4. For a predetermined ϵ , if $|v_{\text{primal}}^{(k)} - v_{\text{master}}^{(k)}| \leq \epsilon$ then terminate. Otherwise, return to Step 2.
-

8.3.1 Solving the Primal Problem

The primal problem in step 3 is a non-linear programming problem (NLP). To solve this problem, we first introduce the following lemma:

Theorem 8.1. *The function $\text{MSE}(\mathcal{G}, \mathcal{P}_p) = \sigma^2 \text{tr}\{(\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1}\}$ is strictly convex on the positive-diagonal matrix \mathcal{P}_p for a given \mathcal{G} .*

Theorem 8.1 (see Appendix 8A for a proof) ensures that there exists a unique global optimum for the pilot powers once their positions are fixed. Indeed, in the case of $K = Q + 1$, the optimal pilot powers can be sought analytically [32]. In other cases where $K > Q + 1$, we can resort to numerous convex optimization algorithms [12]. An example can be the MATLAB[®] built-in function $fmincon(\cdot)$, which will be used in the simulations.

8.3.2 Solving the Master Problem

Compared to the primal problem, the solution to the master problem in step 2 of Table 8.1, which is essentially a binary programming problem, is less straightforward to find. In the first place, the MSE cost function is not necessarily convex on the pilot positions. For instance, if the (C)CE-BEM is used, it is well-known that the solutions for all the equi-distant positions are optimal [69, 73]. The non-convexity is bypassed in, e.g., [29], by relaxing the binary problem to a non-binary problem such that the convexity can be still called upon. Unfortunately, this approach, when applied to our problem, does not facilitate a fast convergence due to a very large pilot position space. In the sequel, we will resort to the combined Genetic Algorithm (GA) [24], which has a fast convergence rate.

The working principle is illustrated by the flow chart in Fig. 8.1. For clarification, we explain the major terms that are listed in the blocks.

- “Population” represents a set of candidate (pilot position) solutions;
- “Weakness” represents the cost function value defined in (8.17) for each candidate. Those candidates that do not satisfy condition (8.18) will be penalized by a weakness value equal to infinity;
- “Reproduction” represents the operation that copies the population except for the candidates that either have the largest weakness or the smallest weakness. The former will be discarded while the latter will be copied twice since it has the lowest MSE at the moment;
- “Crossover” represents the operation applied on the candidate pool produced by the reproduction: at the l th iteration, all the candidates are randomly grouped into pairs (parents). Let us take one such pair for example: suppose “Parent1” is the m th candidate $[g_0^{(l,m)}, \dots, g_{K-1}^{(l,m)}]$ and “Parent2” the n th candidate $[g_0^{(l,n)}, \dots, g_{K-1}^{(l,n)}]$ with $m \neq n$. Here, $g_k^{(l,i)}$ stands in our context for the position of the k th pilot that corresponds to the i th candidate obtained at the l th iteration. “Mating” these two parents results in two new candidates (children) with “Child1” denoted as $[g_0^{(l+1,m)}, \dots, g_{K-1}^{(l+1,m)}]$ and “Child2” as

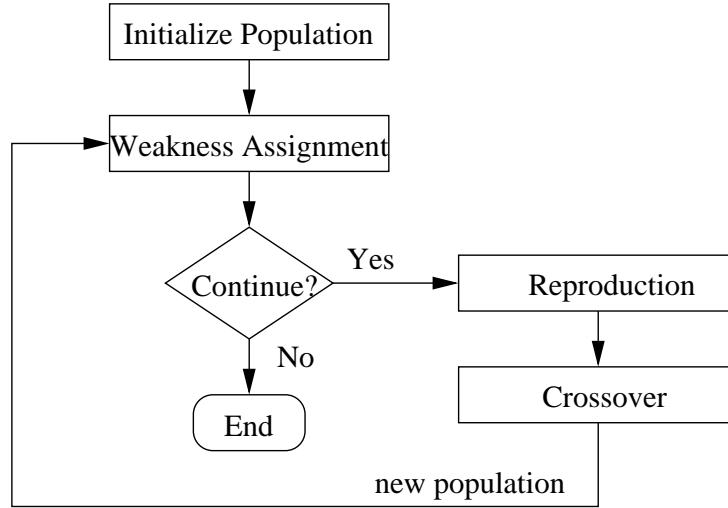


Figure 8.1: A flowchart of GA.

$[g_0^{(l+1,n)}, \dots, g_{K-1}^{(l+1,n)}]$. In generating each entry of the children, either of two operations, replication or swapping, will take place with the same probability. For the k th entry for instance, in case of replication $g_k^{(l+1,m)} = g_k^{(l,m)}$, and $g_k^{(l+1,n)} = g_k^{(l,n)}$. In case of swapping, we first express the k th entry of 'Parent1' and 'Parent2' in binary-form. These two binary strings are then divided at the same arbitrary place, from which the right-hand part of "Parent2" will be concatenated to the left-hand part of "Parent1", and likewise, the right-hand part of "Parent1" will be concatenated to the left-hand part of 'Parent2'. In this way, two new binary strings are created and converted back in decimal form. A numerical example of "crossover" is given as follows.

Example 8.1. Suppose the k th entry of "Parent1" and "Parent2" are

$$g_k^{(l,m)} = 2 \text{ and } g_k^{(l,n)} = 5,$$

respectively. In case of replication, their children stay unchanged, i.e.,

$$g_k^{(l+1,m)} = g_k^{(l,m)} \text{ and } g_k^{(l+1,n)} = g_k^{(l,n)}.$$

If swapping takes place instead, we first find the binary expression for $g_k^{(l,m)}$ and $g_k^{(l,n)}$, which are

$$g_k^{(l,m)} = [0, 1, 0] \text{ and } g_k^{(l,n)} = [1, 0, 1],$$

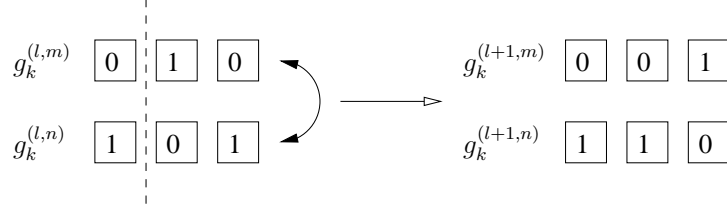


Figure 8.2: An example of swapping.

respectively. Suppose a cutting place is chosen randomly at the first bit. We obtain

$$g_k^{(l+1,m)} = [0, 0, 1], \text{ and } g_k^{(l+1,n)} = [1, 1, 0],$$

which in decimal-form equal 1 and 6, respectively. The procedure of swapping is illustrated in Fig. 8.2.

8.3.3 Imposing Symmetric Constraint

If the observation window size N or pilot number K is large, the GA must abide with a large population size. This results in a higher complexity and lower convergence rate, which in practice, often leads to a “near-optimal” solution. The same remark can be made for the algorithm to solve the primal (power) problem. It is thus helpful if we could equip the searching algorithms with some *a priori* knowledge about the solution. For the considered case (both N and K are assumed to be even), we could constrain the pilot structure to be symmetric with respect to the center of the observation window, i.e.,

$$g_i = N - g_{K-i-1} - 1, \quad (8.20)$$

$$P_i = P_{K-i-1}. \quad (8.21)$$

This constraint is introduced in light of the following properties:

Property 8.1. *The BEM matrices \mathbf{B} of the DKL-BEM, DPS-BEM, CE-BEM and the P-BEM that are introduced in this thesis all admit inherently a symmetrical structure up to a sign. If we use \mathbf{B}_u and \mathbf{B}_d to denote the first and second half of \mathbf{B} , respectively, i.e.,*

$$\mathbf{B}_u := \begin{bmatrix} \mathbf{I}_{\frac{N}{2}} & \mathbf{0}_{\frac{N}{2} \times \frac{N}{2}} \end{bmatrix} \mathbf{B},$$

$$\mathbf{B}_d := \begin{bmatrix} \mathbf{0}_{\frac{N}{2} \times \frac{N}{2}} & \mathbf{I}_{\frac{N}{2}} \end{bmatrix} \mathbf{B},$$

then they are related with each other as

$$\mathbf{B}_d = \mathbf{\Lambda} \mathbf{B}_u^* \mathbf{\Xi}, \quad (8.22)$$

with $\mathbf{\Lambda}$ being an $\frac{N}{2} \times \frac{N}{2}$ flipping matrix,

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{bmatrix}_{\frac{N}{2}}. \quad (8.23)$$

The definition of the matrix $\mathbf{\Xi}$ depends on the concerned BEM. To be more specific,

$$\mathbf{\Xi} = \begin{cases} \mathbf{I}_{Q+1} & \text{for the CE-BEM,} \\ \mathcal{D}\{[(-1)^0, \dots, (-1)^Q]^T\} & \text{for the remaining BEMs.} \end{cases} \quad (8.24)$$

Property 8.1 (see Appendix 8B for a proof) tells the interesting fact that the upper half of any BEM matrix mirrors the lower half up to a sign. Enlightened by the symmetry of the BEM, we observe further that

Property 8.2. If $(\mathcal{G}, \mathcal{P}_p)$ denotes a certain pilot structure with

$$\begin{aligned} \mathcal{P}_p &= \mathcal{D}\{[P_0, \dots, P_{K-1}]^T\}, \\ \mathcal{G} &= \{g_0, \dots, g_{K-1}\}, \end{aligned}$$

then we can always find another pilot structure $(\tilde{\mathcal{G}}, \tilde{\mathcal{P}}_p)$ with

$$\begin{aligned} \tilde{\mathcal{P}}_p &= \mathcal{D}\{[P_{K-1}, \dots, P_0]^T\}, \\ \tilde{\mathcal{G}} &= \{N-1-g_0, \dots, N-1-g_{K-1}\}, \end{aligned}$$

such that

$$MSE(\tilde{\mathcal{G}}, \tilde{\mathcal{P}}_p) = MSE(\mathcal{G}, \mathcal{P}_p). \quad (8.25)$$

Apparently, the new pilot structure $(\tilde{\mathcal{G}}, \tilde{\mathcal{P}}_p)$ is just a flipped version of $(\mathcal{G}, \mathcal{P}_p)$.

Property 8.2 (see Appendix 8C for a proof) suggests that if one specific pilot structure is optimal in an MSE sense, then there exists at least another (flipped) pilot structure that is also optimal. Should there be a unique global minimum, this would imply that the optimal pilot structure ought to be symmetric. Although we have observed that in some applications, e.g., when using the (C)CE-BEM, there might exist multiple optimal solutions [59], and thus imposing a symmetric constraint upon the search algorithm could in theory lead to an MSE degradation, in practice, the algorithm with a symmetry constraint inflicts a much lower complexity and renders a performance very close to (or even better than) the algorithms without a symmetry

Table 8.2: A Pilot Structure Comparison $K = 6$ - Pilot Positions

	symm. NLP / symm. GA	non-symm. NLP / non-symm. GA	equi-powered/ equi-distant	equi-powered / symm. GA	symm. NLP / equi-distant
g_0	1	1	1	1	1
g_1	48	51	44	44	44
g_2	126	53	87	105	87
g_3	131	138	130	152	130
g_4	209	210	173	213	173
g_5	256	253	216	256	216

Table 8.3: A Pilot Structure Comparison $K = 6$ - Pilot Powers

	symm. NLP / symm. GA	non-symm. NLP / non-symm. GA	equi-powered/ equi-distant	equi-powered / symm. GA	symm. NLP / equi-distant
p_0	0.8184	1.4643	1	1	0.6234
p_1	1.2214	0.3162	1	1	1.0457
p_2	0.9157	1.8032	1	1	1.0733
p_3	0.9157	0.3162	1	1	0.6099
p_4	1.2214	0.31622	1	1	1.2571
p_5	0.8184	0.31622	1	1	1.1890

constraint, as shown in the simulation part. This is probably due to the fact that the channel MSE is not a convex function of the pilot positions and has a large number of local minima. By enforcing symmetry, the search space is reduced and hence the global minimum might never be reached. On the other hand, a smaller search space is beneficial in avoiding local minima and thus increases the convergence rate.

8.4 Numerical Examples

We generate time-selective channels as prescribed in [124] for a normalized Doppler spread $\nu_D = 0.002$. The (O)CE-BEM assumption will be adopted to approximate the channel's time-variation, though other BEMs are also applicable but will not be examined here due to space restrictions. Following the definition given in (8.28), we set $N = 256$, $Q = 4$, and $\kappa = 4$.

We compare different solutions for $K = 6$ pilots, which are listed in Table 8.2

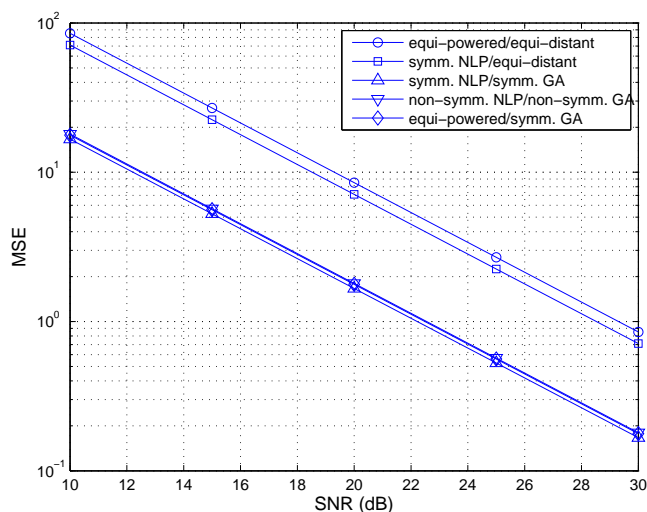


Figure 8.3: MSE vs. SNR.

and Table 8.3 for their positions and powers, respectively. The symmetric GA is equipped with a population pool size of 100 and iterates 100 times, and the non-symmetric GA is equipped with a population pool size of 250 and iterates 1000 times. The latter requires much larger parameters for the setup because more unknowns need to be sought.

The different solutions that are compared are based on the following methods:

1. The proposed method denoted as “symm. NLP/symm. GA” searches the pilot powers using the MATLAB[®] function $fmincon(\cdot)$, and positions using the GA, with both algorithms under a symmetry constraint.
2. The method denoted as “non-symm. NLP/non-symm. GA” searches the pilot powers using the MATLAB[®] function $fmincon(\cdot)$, and positions using the GA. No constraints are imposed on either algorithm.
3. The method denoted as “equi-powered/equi-distant” simply uses pilots that have the same power and a uniform distance.
4. The method denoted as “equi-powered/symm. GA” uses pilots that have the same power, but searches the positions using the symmetric GA.

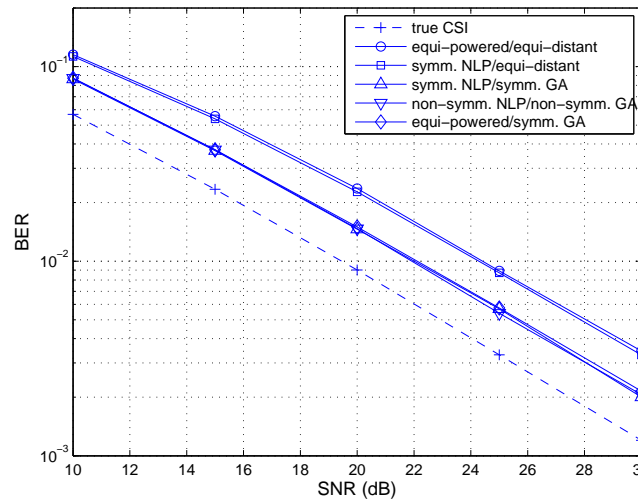


Figure 8.4: BER vs. SNR.

- The method denoted as “symm. NLP/equi-distant” places the pilots with a uniform distance, but searches the pilot powers using the MATLAB[®] function $fmincon(\cdot)$ under a symmetry constraint.

The corresponding MSE performances are plotted in Fig. 8.3, From Fig. 8.3, one can observe that the method “symm. NLP/symm. GA” produces the best channel estimates whereas the traditional method “equi-powered/equi-distant” produces the worst. Note that the method “non-symm. NLP/non-symm. GA” requires a much larger computational complexity than the method “symm. NLP/symm. GA”, but its performance is even slight worse. It is also noteworthy that only optimizing the pilot positions but assuming equal power, the method “equi-powered/symm. GA” can already render a performance comparable to that of the method “non-symm. NLP/non-symm. GA”. This might suggest that the pilot positions are more significant to channel estimation than the pilot powers, and equal pilot powers are in practice very attractive for hardware realization.

BER performances are plotted in Fig. 8.4, which result from a maximum likelihood equalizer that is based on the estimated channel knowledge from different pilot structures. Similar remarks as in Fig. 8.3 can be made here, though the BER performance discrepancy due to different pilot structures is less pronounced. Also, the BER performance based on the true CSI (denoted as “true CSI”) is listed in Fig. 8.4

for comparison. We can see that it performs 2dB better than the equalizer based on the best channel estimates (the method “symm. NLP/symm. GA”).

8.5 Summary

This chapter discussed how to search numerically for the pilot structure that is optimal in a channel MSE sense. The time-selective channel is approximated by an arbitrary BEM, which has only a negligible modeling error with respect to the realistic channel.

First, a mixed integer non-linear programming problem is formulated based on the MSE expression, which is a function of the pilot powers and pilot positions. This problem can be solved readily by the GBD algorithm, essentially a numerical approach that iteratively optimizes the pilot powers and pilot positions. We have shown that the optimization problem is convex in the pilot powers, but non-convex in the pilot positions. The latter can be solved by means of the combined GA. To accelerate the convergence rate of the optimization algorithms, we constrained the pilots with a symmetric structure such that only half of the pilots needs to be optimized.

Appendix 8A: Proof of Theorem 8.1

Proving this theorem is equivalent to proving that for two positive-diagonal matrices $\mathcal{P}_{p,1}$, $\mathcal{P}_{p,2}$, and the real number $\alpha \in [0, 1]$, there holds

$$\begin{aligned} & \text{tr}\{(\mathbf{B}_p^H(\alpha\mathcal{P}_{p,1} + (1-\alpha)\mathcal{P}_{p,2})\mathbf{B}_p)^{-1}\} \\ & < \text{tr}\{(\mathbf{B}_p^H\alpha\mathcal{P}_{p,1}\mathbf{B}_p)^{-1}\} + \text{tr}\{(\mathbf{B}_p^H(1-\alpha)\mathcal{P}_{p,2}\mathbf{B}_p)^{-1}\}. \end{aligned} \quad (8.26)$$

Obviously, the two matrices $\mathbf{B}_p^H\mathcal{P}_{p,1}\mathbf{B}_p$ and $\mathbf{B}_p^H\mathcal{P}_{p,2}\mathbf{B}_p$ are both positive-definite Hermitian, and can therefore be simultaneously diagonalized as [63]

$$\begin{aligned} \mathbf{B}_p^H\mathcal{P}_{p,1}\mathbf{B}_p &= \mathbf{M}^{-H}\mathbf{M}^{-1}, \\ \mathbf{B}_p^H\mathcal{P}_{p,2}\mathbf{B}_p &= \mathbf{M}^{-H}\mathcal{D}\{[1/\lambda_0, \dots, 1/\lambda_{K-1}]^T\}\mathbf{M}^{-1}, \end{aligned}$$

where \mathbf{M} is a $K \times K$ non-singular matrix, and $1/\lambda_k$ is denoted as the k th eigenvalue of the matrix $\mathbf{B}_p^H\mathcal{P}_{p,2}\mathbf{B}_p(\mathbf{B}_p^H\mathcal{P}_{p,1}\mathbf{B}_p)^{-1}$, which is always positive, i.e., $1/\lambda_k > 0$. As

a result, it is straightforward to see that

$$\begin{aligned}
& \text{tr}\{(\mathbf{B}_p^H \alpha \mathcal{P}_{p,1} \mathbf{B}_p)^{-1}\} + \text{tr}\{(\mathbf{B}_p^H (1-\alpha) \mathcal{P}_{p,2} \mathbf{B}_p)^{-1}\} \\
&= \frac{1}{\alpha} \text{tr}\{\mathbf{M} \mathbf{M}^H\} + \frac{1}{1-\alpha} \text{tr}\{\mathbf{M} \mathcal{D}\{[\lambda_0, \dots, \lambda_{K-1}]^T\} \mathbf{M}^H\} \\
&= \frac{1}{\alpha} \sum_{m=0}^{K-1} \sum_{k=0}^{K-1} |[\mathbf{M}]_{m,k}|^2 + \frac{1}{1-\alpha} \sum_{m=0}^{K-1} \sum_{k=0}^{K-1} \lambda_k |[\mathbf{M}]_{m,k}|^2 \\
&= \sum_{m=0}^{K-1} \sum_{k=0}^{K-1} \frac{1-\alpha(1-\lambda_k)}{\alpha(1-\alpha)} |[\mathbf{M}]_{m,k}|^2.
\end{aligned}$$

Likewise

$$\begin{aligned}
& \text{tr}\{(\mathbf{B}_p^H (\alpha \mathcal{P}_{p,1} + (1-\alpha) \mathcal{P}_{p,2}) \mathbf{B}_p)^{-1}\} \\
&= \text{tr}\{\mathbf{M} \mathcal{D}\{[\frac{\lambda_0}{1-\alpha(1-\lambda_0)}, \dots, \frac{\lambda_{K-1}}{1-\alpha(1-\lambda_{K-1})}]^T\} \mathbf{M}^H\} \\
&= \sum_{m=0}^{K-1} \sum_{k=0}^{K-1} \frac{\lambda_k}{1-\alpha(1-\lambda_k)} |[\mathbf{M}]_{m,k}|^2.
\end{aligned}$$

It can be shown after some algebra that

$$\frac{\lambda_k}{1-\alpha(1-\lambda_k)} < \frac{1-\alpha(1-\lambda_k)}{\alpha(1-\alpha)}, \quad (8.27)$$

for any positive λ_k and $\alpha \in [0, 1]$, and hence (8.26) holds, which concludes the proof.

Appendix 8B: Proof of Property 8.1

The P-BEM satisfies this property by definition [c.f. (2.34)]. The symmetry inherent to the DPS-BEM is observed in [95], though without argument. Actually, this is because the kernel matrix of the DPS-BEM, \mathbf{C} , in (2.26) is bisymmetric, i.e., \mathbf{C} is Toeplitz and $\mathbf{C} = \mathbf{C}^T$. For such a bisymmetric matrix, its singular vectors are symmetric as proved in [13]. The same applies for the DKL-BEM, whose kernel matrix is the covariance matrix or the channel, and thus is bisymmetric as well. For the CE-BEM, we make a slight modification and define

$$[\tilde{\mathbf{B}}]_{n,q} = e^{j \frac{2\pi}{\kappa N} (n - \frac{N-1}{2})(q - \frac{Q}{2})}, \quad (8.28)$$

where $\kappa = 1$ for the (C)CE-BEM, and $\kappa > 1$ for the (O)CE-BEM. Compared to the canonical definitions of the CE-BEM, \mathbf{B} , in (2.30) and (2.33), a phase shift is introduced in (8.28):

$$\mathbf{B} = \tilde{\mathbf{B}} \mathcal{D}\{[e^{-j \frac{2\pi}{\kappa N} \frac{N-1}{2} \frac{Q}{2}}, \dots, e^{j \frac{2\pi}{\kappa N} \frac{N-1}{2} \frac{Q}{2}}]^T\}, \quad (8.29)$$

whose effect can be simply absorbed in the corresponding BEM coefficients, without influencing the BEM modeling performance. Obviously, the CE-BEM defined as such admits the symmetric property.

Appendix 8C: Proof of Property 8.2

Let us absorb the effect of \mathcal{P}_p and \mathcal{G} in a larger $N \times N$ diagonal matrix \mathcal{P} , whose diagonal consists of zeros except for the pilot positions, i.e.,

$$[\mathcal{P}]_{i,i} = \begin{cases} [\mathcal{P}_p]_{k,k} & \text{if } i = g_k, \\ 0 & \text{if } i \notin \mathcal{G}. \end{cases} \quad (8.30)$$

By this means, it is equivalent to write (8.17) as

$$\text{MSE}(\mathcal{P}) = \sigma^2 \text{tr}\{(\mathbf{B}^H \mathcal{P} \mathbf{B})^{-1}\}. \quad (8.31)$$

Suppose \mathcal{P} corresponds to the optimal pilots $(\mathcal{G}, \mathcal{P}_p)$. We divide it into the left-upper half, $\mathcal{P}_{u,\text{opt}}$, and the right-bottom half, $\mathcal{P}_{d,\text{opt}}$, respectively, i.e.,

$$\begin{aligned} \mathcal{P}_u &:= \begin{bmatrix} \mathbf{I}_{\frac{N}{2}} & \mathbf{0}_{\frac{N}{2} \times \frac{N}{2}} \end{bmatrix} \mathcal{P} \begin{bmatrix} \mathbf{I}_{\frac{N}{2}} & \mathbf{0}_{\frac{N}{2} \times \frac{N}{2}} \end{bmatrix}^T, \\ \mathcal{P}_d &:= \begin{bmatrix} \mathbf{0}_{\frac{N}{2} \times \frac{N}{2}} & \mathbf{I}_{\frac{N}{2}} \end{bmatrix} \mathcal{P} \begin{bmatrix} \mathbf{0}_{\frac{N}{2} \times \frac{N}{2}} & \mathbf{I}_{\frac{N}{2}} \end{bmatrix}^T, \end{aligned}$$

such that

$$\mathcal{P} = \begin{bmatrix} \mathcal{P}_u & \\ & \mathcal{P}_d \end{bmatrix}.$$

In line with such a decomposition and by virtue of Property 8.1, we can rewrite (8.31) as

$$\begin{aligned} \text{MSE}(\mathcal{P}) &= \sigma^2 \text{tr}\{(\mathbf{B}_u^H \mathcal{P}_u \mathbf{B}_u + \mathbf{B}_d^H \mathcal{P}_d \mathbf{B}_d)^{-1}\} \\ &= \sigma^2 \text{tr}\{(\mathbf{B}_u^H \mathcal{P}_u \mathbf{B}_u + \mathbf{\Xi}^H \mathbf{B}_u^T \mathbf{\Lambda}^H \mathcal{P}_d \mathbf{\Lambda} \mathbf{B}_u^* \mathbf{\Xi})^{-1}\}. \end{aligned}$$

For the flipped pilot structure $(\tilde{\mathcal{G}}, \tilde{\mathcal{P}}_p)$, we can derive similarly a matrix $\tilde{\mathcal{P}}$ like in (8.30), which is obviously related to \mathcal{P} as

$$\tilde{\mathcal{P}} = \begin{bmatrix} \mathbf{\Lambda}^H \mathcal{P}_d \mathbf{\Lambda} & \\ & \mathbf{\Lambda}^H \mathcal{P}_u \mathbf{\Lambda} \end{bmatrix},$$

where Λ is the flipping matrix defined in (8.23). Likewise, the MSE corresponding to the pilot structure $\tilde{\mathcal{P}}$ admits the expression [c.f. (8.31)] as

$$\begin{aligned} \text{MSE}(\tilde{\mathcal{P}}) &= \sigma^2 \text{tr} \{ (\mathbf{B}_u^H \Lambda^H \mathcal{P}_d \Lambda \mathbf{B}_u + \Xi^H \mathbf{B}_u^T \mathcal{P}_u \mathbf{B}_u^* \Xi)^{-1} \} \\ &\stackrel{a}{=} \sigma^2 \text{tr} \{ (\Xi^H \mathbf{B}_u^H \Lambda^H \mathcal{P}_d \Lambda \mathbf{B}_u \Xi + \mathbf{B}_u^T \mathcal{P}_u \mathbf{B}_u^*)^{-1} \} \\ &\stackrel{b}{=} \sigma^2 \text{tr} \{ (\Xi^H \mathbf{B}_u^T \Lambda^H \mathcal{P}_d \Lambda \mathbf{B}_u^* \Xi + \mathbf{B}_u^H \mathcal{P}_u \mathbf{B}_u)^{-1} \} \\ &= \text{MSE}(\mathcal{P}), \end{aligned}$$

where $\stackrel{a}{=}$ holds because $\Xi \Xi = \Xi^H \Xi^H = \mathbf{I}_{Q+1}$, and $\stackrel{b}{=}$ holds because $\text{tr} \{ \mathbf{A}^{-1} \} = \text{tr} \{ \mathbf{A}^{*-1} \}$ for a Hermitian matrix \mathbf{A} .

9.1 Conclusions

OFDM is characterized by a diagonal channel matrix in the frequency domain if the channel is constant within the duration of an OFDM symbol. To equalize such a diagonal channel matrix requires only a linear complexity with respect to the total number of subcarriers. However, when the channel time-variation cannot be ignored within the OFDM symbol, the resulting FD channel matrix is full.

Despite this fact, we have explored the property that most of the channel power is concentrated in the entries close to the main diagonal and hence the channel matrix can be approximated as banded. We have constructed a block LMMSE equalizer which is based on a strictly banded matrix instead of on the original full channel matrix. This enables us to apply a simple Cholesky factorization such that the overall complexity is still linear in the number of subcarriers, just like for the TI channel case. To reduce the band approximation error between the strictly banded matrix and the original full matrix, we introduced a window at the receiver. At the meantime, we have shown that the corresponding band approximation error can be interpreted as the modeling error resulting from using the (C)CE-BEM in the time domain. In order to enhance the equalization performance, we optimized the value of both the window and the strictly banded matrix.

Next, we discussed how to estimate the channel, which is indispensable for the channel equalizer. Again, the full channel matrix due to the channel time-variation presents a challenge to traditional channel estimators. The pilot-carrying subcarriers are constantly corrupted by interference due to the unknown data symbols. As counteract, we proposed clustered pilots. In addition, we have found that for different types of channel estimators, the optimal number of observation samples must individually be chosen: for the LMMSE estimator and BLUE, it is optimal to employ the whole observed OFDM symbol for channel estimation because of their superior capability to suppress the interference. In contrast, it is wise for the LS estimator to only choose those observation samples that have highest SINR.

By extending channel estimation from a single OFDM symbol to multiple OFDM

symbols, we are able to exploit the temporal correlation better, and hence the estimation precision can be enhanced. This raised the question as how to distribute the pilots along both the frequency axis and the time axis. Existing pilot schemes can be categorized into the comb-type, the block-type and the mixed-type scheme. From the point of view of channel identifiability, we posed constraints on the numbers, positions and size of the pilot clusters for these pilot schemes. Further, we have compared their channel estimation performance for different channel situations. It turned out that the comb-type scheme produces the best channel estimates for a fast fading channel while the block-type scheme produces the worst. If the channel is overspread in both the Doppler and delay domain, then the mixed-type scheme yields the most robust performance.

To equalize the TV channel for a single-carrier system in the frequency domain, we came across the same problems as in OFDM. The FD channel matrix is not diagonal but approximately banded, which implies that the low-complexity equalizer based on a banded channel matrix assumption can be similarly applied here. In addition, we have found in the single-carrier system that by introducing redundancy at the receiver, we can exploit a better Doppler frequency resolution, which results in a smaller band approximation error. This band approximation error is related to the (O)CE-BEM modeling error in the time domain. Recall that in the case where no receiver redundancy is introduced just like in OFDM, the band approximation error corresponds to the (C)CE-BEM modeling error. It is well-known that the (O)CE-BEM is much better than the (C)CE-BEM in terms of the modeling error. This is why receiver redundancy yields a better performance in a single-carrier system.

To attain channel knowledge in a single-carrier system, we have proposed to use pilots which are clustered and interleaved with data symbols in the time domain. This means that channel estimation in a single-carrier system has been achieved in the time domain, where we have again used a BEM to approximate the real TV channel. Like in OFDM, the choice of observation samples plays a distinctive role for different channel estimators. It is better for the LMMSE estimator and the BLUE to employ as many observation samples as possible to improve the estimation precision. The opposite is true for the LS estimator in order to avoid unnecessary interference. We also discussed channel identifiability in a single-carrier system, for which the pilots are required to follow a certain structure if an arbitrary BEM is used.

An optimized pilot structure can improve the channel estimation performance under the same power and bandwidth constraints. We discussed a simple case where the channel is assumed to be only time-selective. A numerical method has been proposed, which optimized the pilot powers and positions iteratively. It was shown that the channel estimation performance is a convex function of the pilot powers, and hence the existing convex optimization algorithms can be readily ap-

plied. However, it is not convex with respect to the pilot positions. To solve this, we proposed a combined GA. To make the GA converge faster, a symmetry constraint was posed on the pilot structure, which was inspired by the structure of the BEM.

The contributions summarized above were all presented along with numerical simulations, through which we have shown that the proposed methods outperform other existing works, or for the same performance require a much smaller complexity. The simulations were based on realistic TV channels and tested for diverse normalized Doppler spreads up to $\nu_D = 0.008$. For existing telecommunication standards such as IEEE 802.11a, this means that our algorithms can work for a maximal vehicle velocity of 200 kilometers per hour. This can be even higher for those standards that are based on a lower carrier frequency and/or higher symbol rate.

9.2 Future Research

The window is indispensable in the low-complexity equalization for both the OFDM and single-carrier systems. In the thesis, it was assumed that the window is aligned with the received block of interest (thus the window size equals the block size), and the window takes on a form of a one-tap filter. It would be interesting to study if the window could be elongated to the neighboring received blocks like in [44], and if the window could take on a more complex structure such as an FIR filter as was proposed in [8].

The methods proposed in this thesis have been proven effective for estimating a rapidly changing channel. Estimating a long TI channel has also been intensively studied in the literature. However, when the channel is overspread in both the Doppler and delay domain, we will need more pilots, which is not favorable for bandwidth efficiency. Two solutions might be interesting to investigate. First, like the BEM used in the Doppler domain, is it possible to introduce a parsimonious model in the delay domain such that the number of channel taps can also be reduced? Second, data-aided channel estimation could be useful [123], where not only the pilots but also the estimated data can be employed to assist channel estimation.

Another interesting topic is diversity techniques. Space-time coding has proven useful to improve the robustness of the transmission link. It is not yet clear how to extend this technique to the Doppler domain. At this moment, most results are attained for a relatively slow channel where a CE-BEM assumption is valid [85, 60]. It will be of great significance if a diversity-enabled transmission strategy can be found for a faster fading channel.

Recently, the popularity of OFDM has been extended to other areas such as underwater acoustic communications (UAC) [93] or radar communications [28, 113]

where usually other modulation techniques were employed. This brings about many new challenges, especially in the context of the Doppler effects. For instance, in UAC, acoustic waves are transmitted instead of electro-magnetic waves; the symbol period is typically large, and the waves are constantly moving. All of these can lead to an enormously high Doppler spread even if the mobile is moving at a low velocity. Another complication is that this thesis considered only narrow-band signals, in which the Doppler effect manifests itself mainly as a frequency shift around the carrier frequency of the transmitted signals. In contrast, UAC or radar communications consider wide-band signals, where the bandwidth after modulation is much higher than the bandwidth of the signal itself*, e.g., the bandwidth in UAC can be as high as 10 to 20kHz whereas the acoustic signals have a bandwidth of at most 4kHz. In wide-band signals, the Doppler effect translates into a time scaling of the signal waveform, i.e., the received symbol period can be either larger or smaller than the transmitted symbol period. How to take these phenomena into account in signal processing remains to be studied.

*In telecommunications, "wide-band" has many meanings in different contexts. Here, we adopt the definition that is often used in audio frequency modulation [16].

Bibliography

- [1] K. Abed-Meraim, J.-F. Cardoso, A. Y. Gorokhov, P. Loubaton, and E. S. Moulines, "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Transactions on Signal Processing*, vol. 45, pp. 42–55, Jan. 1997.
- [2] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of known symbols for frequency-selective block-fading channels," *IEEE Transactions on Information Theory*, vol. 48, pp. 2338–2353, Aug. 2002.
- [3] N. Al-Dhahir, "FIR channel-shortening equalizers for MIMO ISI channels," *IEEE Transactions on Communications*, vol. 49, pp. 213–218, Feb 2001.
- [4] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451 – 1458, Oct 1998.
- [5] C. R. N. Athaudage and A. D. S. Jayalath, "Enhanced MMSE channel estimation using timing error statistics for wireless OFDM systems," *IEEE Transactions on Broadcasting*, vol. 50, pp. 369–376, Dec 2004.
- [6] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for mimo ofdm systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1615–1624, June 2003.
- [7] I. Barhumi, G. Leus, and M. Moonen, "Time-domain and frequency-domain per-tone equalization for OFDM in doubly-selective channels," *Elsevier Signal Processing*, vol. 84, pp. 2055–2066, Nov. 2004.
- [8] I. Barhumi, G. Leus, and M. Moonen, "Equalization for OFDM over doubly-selective channels," *IEEE Transactions on Signal Processing*, vol. 54, pp. 1445–1458, Apr. 2006.
- [9] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. Kluwer Academic/Plenum Publishers, 1999.
- [10] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Transactions on Communications*, vol. 47, pp. 862–873, June 1999.

- [11] D. K. Borah and B. D. Hart, "A robust receiver structure for time-varying, frequency-flat Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, pp. 862–873, Mar. 1999.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*. CAM, 2004.
- [13] J. A. Cadzow, "SVD representation of unitarily invariant matrices," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, pp. 512–516, June 1984.
- [14] X. Cai and G. B. Giannakis, "Bounding performance and suppressing intercarrier interference in wireless mobile OFDM," *IEEE Transactions on Communications*, vol. 51, pp. 2047–2056, Dec. 2003.
- [15] R. C. Cannizzaro, P. Banelli, and G. Leus, "Adaptive channel estimation for OFDM systems with Doppler spread," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, July 2006.
- [16] A. B. Carlson, *Communication systems, 2nd edition*. McGraw-Hill Inc., 1981.
- [17] J.-W. Choi and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2083–2088, Sept. 2005.
- [18] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 49, pp. 1375–1387, Aug. 2001.
- [19] H. A. Cirpan and M. K. Tsatsanis, "Maximum likelihood blind channel estimation in the presence of doppler shifts," *IEEE Transactions on Signal Processing*, vol. 47, pp. 1559–1569, June 1999.
- [20] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 48, pp. 223–229, Sept. 2002.
- [21] T. Cui, C. Tellambura, and Y. Wu, "Low-complexity pilot-aided channel estimation for OFDM systems over doubly-selective channels," *IEEE International Conference on Communications, ICC*, vol. 3, pp. 1980–1984, May 2005.
- [22] P. J. Davis, *Interpolation and Approximation*. Dover Publications, Inc., New York, 1975.
- [23] E. de Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 3593–3596, Apr. 1997.
- [24] K. Deb and M. Goyal, "A combined genetic adaptive search (GeneAs) for engineering design," *Computer Science and Informatics*, vol. 26, no. 4, pp. 30–45, 1996.
- [25] L. Deneire, B. Gyselinckx, and M. Engels, "Training sequence versus cyclic prefix - a new look on single carrier communication," *IEEE Communications Letters*, vol. 7, no. 5, pp. 292–294, 2001.
- [26] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 3055–3069, Dec. 2002.

- [27] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1403–1418, May 2004.
- [28] B. J. Donnet and J. D. Longstaff, "Combining MIMO radar with OFDM communications," *Proceedings of the Third European Radar Conference, Manchest, UK*, pp. 37–40, Sept. 2006.
- [29] A. Dua, K. Medepalli, and A. J. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 2353–2357, Sept. 2006.
- [30] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, pp. 58–66, Apr. 2002.
- [31] C. A. Floudas, *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Oxford University Press, 1995.
- [32] M. Ghogho, "On optimum pilot design for OFDM systems with virtual carriers," *IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, Mar. 2006.
- [33] M. Ghogho and A. Swami, "Improved channel estimation using superimposed training," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, pp. 110–114, July 2004.
- [34] M. Ghogho and A. Swami, "Estimation of doubly-selective channels in block transmissions using data-dependent superimposed training," *European Signal Processing Conference, EUSIPCO*, Sept. 2006.
- [35] G. B. Giannakis, "Filterbands for blind channel identification and equalization," *IEEE Signal Processing Letters*, vol. 4, pp. 184–187, July 1997.
- [36] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proceedings of the IEEE*, vol. 86, pp. 1969–1986, Oct. 1998.
- [37] G. H. Golub and C. F. van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1989.
- [38] A. Gorokhov and J.-P. Linnartz, "Robust OFDM receivers for dispersive time-varying channels: Equalization and channel acquisition," *IEEE Transactions on Communications*, vol. 52, pp. 572–583, Apr. 2004.
- [39] M. Guillaud and D. T. M. Slock, "Channel modeling and associated inter-carrier interference equalization for OFDM systems with high doppler spread," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. IV, pp. 237–240, Apr. 2003.
- [40] B. Hassibi and B. M. Hochwald, "How much training is needed in multi-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, pp. 951–963, Apr. 2003.

- [41] S. Haykin, *Adaptive filter theory*. Englewood Cliffs : Prentice-Hall, 1996.
- [42] S. He and J. Tugnait, "Doubly-selective multiuser channel estimation using superimposed training and discrete Prolate spheroidal basis expansion models," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, Apr. 2007.
- [43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1999.
- [44] S.-J. Hwang and P. Schniter, "Efficient sequence detection of multicarrier transmissions over doubly dispersive channels," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 93638, 17 pages, 2006.
- [45] W. C. Jakes, *Microwave Mobile Channels*. New York: Wiley, 1974.
- [46] W. G. Jeon, K. H. Chang, and Y. S. Cho, "An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels," *IEEE Transactions on Communications*, vol. 47, pp. 27–32, Jan. 1999.
- [47] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, Mar. 2005.
- [48] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey, USA, 1993.
- [49] J. Kim, J. R. W. Heath, and E. Powers, "Receiver designs for Alamouti coded OFDM systems in fast fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 550–559, Mar. 2005.
- [50] D. Kincaid and W. Cheney, *Numerical Analysis*. Brooks/Cole Publishing Company, 1991.
- [51] C. Lanczos, *Applied Analysis*. Prentice Hall, 1956.
- [52] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.
- [53] G. Leus, "On the estimation of rapidly time-varying channels," *European Signal Processing Conference, EUSIPCO*, pp. 2227–2230, Sept. 2004.
- [54] G. Leus, "Semi-blind channel estimation for rapidly time-varying channels," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 773–776, Mar. 2005.
- [55] G. Leus and M. Moonen, "Deterministic subspace based blind channel estimation for doubly-selective channels," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, June 2003.
- [56] G. Leus and M. Moonen, *Equalization Techniques for Fading Channels, Chapter in Handbook on Signal Processing for Mobile Communications*. CRC Press, 2004.
- [57] G. Leus and A.-J. van der Veen, *Channel Estimation, Chapter in Smart Antennas - State of the Art*. Hindawi, 2005.
- [58] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Transactions on Communications*, vol. 46, pp. 1146–1162, July 1998.

- [59] X. Ma, G. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective fading channels," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1351–1366, May 2003.
- [60] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1832–1840, July 2003.
- [61] X. Ma, G. Leus, and G. B. Giannakis, "Space-time-doppler coding for correlated time-selective fading channels," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2167–2181, June 2005.
- [62] X. Ma, L. Yang, and G. Giannakis, "Optimal training for mimo frequency-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 453–466, Mar. 2005.
- [63] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. Academic Press, 1979.
- [64] A. Medles and D. T. M. Slock, "Augmenting the training sequence part in semiblind estimation for MIMO channels," *Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 1825 – 1829, Nov. 2003.
- [65] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. New Jersey, USA, 2000.
- [66] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Transactions on Communications*, vol. 42, pp. 2908–2914, Oct. 1994.
- [67] Y. Mostofi and D. C. Cox, "ICI mitigation for pilot-aided OFDM mobile systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 765–774, Mar. 2005.
- [68] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.
- [69] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Transactions on Consumer Electronics*, vol. 44, pp. 1122–1128, Aug. 1998.
- [70] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of frequency-selective fast-varying channels," *IEEE Transactions on Communications*, vol. 51, pp. 1337–1347, Aug. 2003.
- [71] S. Ohno, "Maximum likelihood inter-carrier interference suppression for wireless OFDM with null subcarriers," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 3, pp. 849 – 852, Mar. 2005.
- [72] S. Ohno, "Performance of single-carrier block transmissions over multipath fading channels with linear equalization," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3678–3687, Oct. 2006.

- [73] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless ofdm," *IEEE Transactions on Communications*, vol. 50, pp. 2113–2123, Dec. 2002.
- [74] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 2138–2145, Sept. 2004.
- [75] B. Porat, *A Course in Digital Signal Processing*. John Wiley & Sons, 1997.
- [76] J. G. Proakis, *Digital Communications*. Boston: Macgraw-Hill, 2000.
- [77] T. S. Rappaport, *Wireless Communications, Principles & Practice*. Prentice-Hall PTR, 1999.
- [78] O. Rousseaux, G. Leus, and M. Moonen, "Estimation and equalization of doubly-selective channels using known symbol padding," *SP*, vol. 54, pp. 979–990, Mar. 2006.
- [79] O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "Gaussian maximum likelihood channel estimation with short training," *WC*, vol. 4, pp. 2945–2955, Nov. 2004.
- [80] L. Rugini and P. Banelli, "Windowing techniques for ICI mitigation in multicarrier systems," *European Signal Processing Conference, EUSIPCO*, 2005.
- [81] L. Rugini, P. Banelli, and G. Leus, "Simple equalization of time-varying channels for OFDM," *IEEE Communications Letters*, vol. 9, pp. 619–621, July 2005.
- [82] L. Rugini, P. Banelli, and G. Leus, "Low-complexity banded equalizers for OFDM systems in Doppler spread channels," *EURASIP Journal on Applied Signal Processing*, pp. Article ID 67404, 13 pages, 2006.
- [83] B. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Transactions on Communications*, vol. 15, pp. 805–811, Dec. 1967.
- [84] F. Sanzi, S. Jeltng, and J. Speidel, "A comparative study of iterative channel estimation for mobile OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 849–859, Sept. 2003.
- [85] A. M. Sayeed and B. Aazhang, "Joint multipath-Doppler diversity in mobile wireless communications," *IEEE Transactions on Communications*, vol. 47, pp. 123–132, Jan. 1999.
- [86] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers - part I: Unification and optimal designs," *IEEE Transactions on Signal Processing*, vol. 47, pp. 1988–2006, July 1999.
- [87] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers - part II: Blind channel estimation, synchronization and direct equalization," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2007–2022, July 1999.
- [88] D. Schafhuber and G. Matz, "MMSE and adaptive prediction of time-varying channels for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 593–602, Mar. 2005.
- [89] D. Schafhuber, G. Matz, and F. Hlawatsch, "Kalman tracking of time-varying channels in wireless MIMO-OFDM systems," *Asilomar Conference on Signals, Systems, and Computers*, pp. 1261–1265, 2003.

- [90] P. Schniter, "Low-complexity equalization of OFDM in doubly-selective channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1002–1011, Apr 2004.
- [91] P. Schniter, "On doubly dispersive channel estimation for pilot-aided pulse-shaped multicarrier modulation," *Proc. Conference on Information Sciences and Systems*, (Princeton, NJ), Mar. 2006.
- [92] P. Schniter and H. Liu, "Iterative frequency-domain equalization for single-carrier systems in doubly-dispersive channels," *Asilomar Conference on Signals, Systems, and Computers*, pp. 667–671, Nov 2004.
- [93] B. S. Sharif, J. Neasham, O. R. Hinton, and A. E. Adams, "A computationally efficient Doppler compensation system for underwater acoustic communications," *IEEE Journal of Oceanic Engineering*, vol. 25, pp. 52–61, Jan. 2000.
- [94] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading rayleigh MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 3265–3277, Nov. 2004.
- [95] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty - V: the discrete case," *Bell Labs Technical Journal*, vol. 57, pp. 1371–1429, May-Jun 1978.
- [96] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2451–2464, Oct. 2002.
- [97] Z. Tang, R. C. Cannizzaro, P. Banelli, and G. Leus, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2226–2238, May 2007.
- [98] Z. Tang and G. Leus, "Pilot schemes for time-varying channel estimation in ofdm systems," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, June 2007.
- [99] Z. Tang and G. Leus, "Receiver design for single-carrier transmission over time-varying channels," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, 2007.
- [100] Z. Tang and G. Leus, "A novel receiver architecture for single-carrier transmission over time-varying channels," *IEEE Journal on Selected Areas in Communications*, submitted.
- [101] Z. Tang and G. Leus, "Time-multiplexed training for time-selective channels," *IEEE Signal Processing Letters*, to appear.
- [102] Z. Tang, G. Leus, and P. Banelli, *Time-Varying Channel Estimation - a Block Approach*, Chapter in *Wireless Communications over Rapidly Time-Varying Channels*. approved by Academic Press.
- [103] Z. Tang, G. Leus, and P. Banelli, "Pilot-assisted time-varying OFDM channel estimation," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 4, pp. 133–136, May 2006.
- [104] Z. Tang, G. Leus, and P. Banelli, "Pilot-assisted time-varying OFDM channel estimation based on multiple OFDM symbols," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, June 2006.

- [105] K. D. Teo and S. Ohno, "Optimal MMSE finite parameter model for doubly-selective channels," *IEEE Global Telecommunications Conference, GLOBECOM*, 2005.
- [106] S. Thoen, L. V. der Perre, and M. Engles, "Modeling the channel time-variance for fixed wireless communications," *IEEE Communications Letters*, vol. 6, pp. 331–333, Aug. 2002.
- [107] T. A. Thomas and F. W. Vook, "Multi-user frequency-domain channel identification, interference suppression, and equalization for time-varying broadband wireless communications," *Proceedings of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 444–448, Mar. 2000.
- [108] S. Tomasin, A. Gorokhov, H. Yang, and J.-P. Linnartz, "Iterative interference cancellation and channel estimation for mobile OFDM," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 238–245, Jan. 2005.
- [109] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Processing Magazine*, vol. 21, pp. 12–25, Nov. 2004.
- [110] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *International Journal of Adaptive Control and Signal Processing*, vol. 10, pp. 159–176, Mar. 1996.
- [111] J. K. Tugnait and W. Luo, "Linear prediction error method for blind identification of periodically time-varying channel," *IEEE Transactions on Signal Processing*, vol. 50, pp. 3070–3082, Dec. 2002.
- [112] J. K. Tugnait and W. Luo, "Blind identification of time-varying channels using multistep linear predictors," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1739–1749, June 2004.
- [113] P. van Genderen and H. Nikookar, "Radar network communication," *Proceedings of International Conference Communications, Bucharest, Romania*, pp. 313–316, June 2006.
- [114] M. Visintin, "Karhunen-Loeve expansion of a fast Rayleigh fading process," *IEEE Electronics Letters*, vol. 32, pp. 1712–1713, Aug 1996.
- [115] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about traing: Guidelines derived using the affine precoding framework and the CRB," *IEEE Transactions on Signal Processing*, vol. 54, pp. 940–954, Mar 2006.
- [116] H. S. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 353–357, May 1996.
- [117] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Magazine*, vol. 17, pp. 29–48, May 2004.
- [118] Z. Wang, X. Ma, and G. B. Giannakis, "OFDM or single-carrier block transmissions?," *IEEE Transactions on Communications*, vol. 52, pp. 380–394, Mar. 2004.
- [119] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Transactions on Communications*, vol. 49, pp. 467–479, Mar. 2001.

-
- [120] K.-W. Yip and T.-S. Ng, "Karhunen-loève expansion of the WSSUS channel output and its application to efficient simulation," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 640–646, May 1997.
- [121] Y. V. Zakharov, T. C. Tozer, and J. F. Adlard, "Polynomial spline-approximation of Clarke's model," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1198–1208, May 2004.
- [122] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3597–3607, Sept. 2005.
- [123] T. Zemen, C. F. M. J. Wehinger, and R. R. Müller, "Iterative joint time-variant channel estimation and multi-user detection for MC-CDMA," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 1469 – 1478, June 2006.
- [124] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, pp. 920–928, June 2003.
- [125] S. Zhou, B. Muquet, and G. B. Giannakis, "Subspace-based (semi-) blind channel estimation for block precoded space-time OFDM," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1215–1228, May 2002.

Summary

This thesis is dedicated to efficiently estimate and equalize time-varying channels in OFDM and single-carrier systems.

We first explore a low-complexity channel equalizer, which is constructed based on a banded approximation of the full channel matrix in the frequency domain. Thanks to the banded structure, existing low-complexity numeric algorithms can be applied. To minimize the discrepancy between the banded approximation and the original full channel matrix, a window is designed and deployed at the receiver. Further, the strictly banded matrix is optimally sought such that its discrepancy with respect to the original (windowed) channel matrix is minimized according to the Frobenius norm.

To acquire the channel state information (CSI), which is indispensable for the equalizer design, pilots are inserted in the OFDM symbol. The channel is estimated in the time domain, where the time-variation of each channel tap is approximated utilizing a basis expansion model (BEM), and thus the time-varying channel is captured by the corresponding BEM coefficients. For a certain pilot structure, we explore different channel estimation methods, and discover that for each channel estimator, the window of observation samples must be optimized.

A question that arises when multiple pilot OFDM symbols are employed to enhance the estimation precision is, what is the strategy to distribute the pilots within as well as along the OFDM symbols. This question is answered in terms of the channel identifiability conditions. Numerical results are also used to compare different pilot distribution schemes.

By transforming the channel to the frequency domain, equalization of a single-carrier system can be similarly realized as for OFDM, where windowing and band approximation operations can be utilized. Moreover, it is possible in the single-carrier system to expand the data model by appending extra samples to the received

samples. These extra samples are free to design, which enable us to better exploit the Doppler spectrum resolution. As a result, the band approximation error can be further reduced, which eventually improves the equalization performance without inflicting too much extra complexity.

Channel estimation for single-carrier systems is also discussed. Pilots are inserted in the time domain. Like in OFDM, we need to optimize the number of observation samples for distinctive channel estimators. Channel identifiability is also discussed for the single-carrier system, which leads to constraints on the pilot structure.

As an effort to explore the optimal pilot structure, we examine the simple case of time-varying frequency-flat channels. Numerical optimization methods are applied to seek the best pilot power allocation and distribution strategy. We discover that by imposing a symmetric constraint on the pilot structure, we can accelerate the optimization speed without incurring a noticeable performance penalty.

Samenvatting

Dit proefschrift is toegewijd aan het egaliseren en schatten van tijdsvarierende kanalen in OFDM en single-carrier systemen.

Om een eenvoudige egalisator te realiseren, benaderen we de kanaalmatrix in het frequentiedomein door een bandmatrix. Zo een bandmatrix geeft aanleiding tot eenvoudige numerieke algoritmes voor kanaalinversie. Om de benaderingsfout tussen de bandmatrix en de originele volle kanaalmatrix te verkleinen, wordt een speciaal venster ontworpen dat wordt aangewend aan de ontvanger. Verder wordt de bandmatrix zo ontworpen dat het verschil ten aanzien van de originele kanaalmatrix (inclusief venster) minimaal is volgens de Frobenius norm.

De kanaalkennis die nodig is voor het opbouwen van de egalisator wordt verworven door middel van trainingsbits in het OFDM symbool. Het kanaal wordt echter geschat in het tijdsdomein, waar de tijdsvariatie van elke kanaaltap wordt benaderd door een basis expansion model (BEM). Het tijdsvarierend kanaal wordt dan volledig vertegenwoordigd door de overeenkomstige BEM coëfficiënten. Voor een bepaalde structuur van de trainingsbits worden verscheidene kanaalschatters bestudeerd en we hebben aangetoond dat het optimaal aantal observatiemonsters verschilt van kanaalschatter tot kanaalschatter.

Meerdere OFDM symbolen met trainingsbits kunnen samen worden gebruikt om een betere kanaalschatting te verkrijgen. Een vraag die men dan kan stellen is wat de beste strategie is om de trainingsbits zowel binnen als over de verschillende OFDM symbolen te verspreiden. Deze vraag wordt beantwoord aan de hand van de identificeerbaarheidscondities van het kanaal. Daarnaast worden ook verschillende trainingsstructuren vergeleken door middel van numerieke simulaties.

Door het kanaal te transformeren naar het frequentiedomein, kan een single-carrier systeem op een gelijkaardige manier ge-egaliseerd worden als een OFDM systeem, waarbij we opnieuw kunnen gebruik maken van vensters en bandmatrix-

benaderingen. Het is tevens mogelijk om het datamodel van een single-carrier systeem te vergroten door extra monsters aan de reeds ontvangen monsters toe te voegen. Deze extra monsters zijn vrij te kiezen wat ons impliciet een hogere resolutie van het Doppler-spectrum oplevert. Ten gevolge daarvan, kan de bandmatrixbenadering alsook de egalisatie worden verbeterd, zonder een al te hoge extra kost.

Kanaalschatting voor single-carrier systemen wordt ook besproken. Ditmaal worden de trainingsbits geplaatst in het tijdsdomein. Zoals voor OFDM, is het opnieuw noodzakelijk om voor de verschillende kanaalschatters het aantal observatiemonsters te optimaliseren. De identificeerbaarheid van het kanaal in een single-carrier systeem wordt ook bestudeerd. Aan de hand daarvan worden een aantal voorwaarden opgesteld waaraan de structuur van de trainingsbits moet voldoen.

In een poging om de optimale structuur voor de trainingsbits te vinden, onderzoeken we het eenvoudige geval van tijdsvarierende frequentievlakke kanalen. Numerieke optimalisatietechnieken worden aangewend om het optimale vermogen en de optimale verdeling van de trainingsbits te achterhalen. Verder tonen we aan dat als we symmetrie opleggen aan de trainingsstructuur, we de optimalisatie kunnen versnellen zonder een merkbaar performantieverlies.

Acknowledgment

Let me begin with my parade of acknowledgement first with thanking all the commission members for kindly agreeing to devote time and effort in judging and giving precious opinions to this thesis.

I want to show my great gratitude to Alle-Jan, not only just because he enrolled me as his PhD student despite all the doubts he has ever had but, more importantly, he opens for me the door to the great world of linear algebra. His textbook of the course "Signal Processing for Communications" remains to be one of few resources that I have been constantly referring to.

Also my coach Geert deserves my biggest applause. Without his mentorship, I would have hardly gone through. I really learned a lot from him, in particular, the way how he breaks a problem up with a piercing eye to quickly pinpoint where the bottleneck lies. He is a great listener: I never need to invest a lot of efforts in explaining things to him. He is a relentless critics whenever I come up with any new ideas. Afterwards, he turns into a keen buddy helping me find out what the solution can be. Thanks Geert!

Speical thanks to our Italian colleagues: Claudio, Luca and Paolo (have I already thanked you before Paolo?). The cooperation with Claudio and Paolo has led us to a very fruitful research. Luca spent a large portion of his six-months-stay in Delft in just answering me a lot of silly questions. He has also very kindly offered to be my referral when I was soliciting my present job.

I would thank Martijn for taking the trouble to help me translate the summary and propositions in Dutch. I know that this is a terribly difficult job to find a proper Dutch word for technical English. And he did this while Tycho (his son) was sick.

I also want to thank Liang and Wenqian for helping me design the cover. Actually, they exhibited even more enthusiasm than I myself in making it better. Besides, Liang was to me also the best landlord of the world.

I feel lucky that I have so many helpful and nice colleagues: Hieu, Antoon, Laura B, Laura S, Paula, Kun, Claud, Vijay, Bas, Sayit, Tao, Yiyin, Eelco, Filip, Relja, Antonio, Qiang, Kees-Jan, Haiyan, Zhifeng... It has been the best four years of my life.

Last but not least, I would also express my gratitude to my new colleagues at The Mathworks for their kind understanding and supporting in the finishing phase of the thesis.

Curriculum Vitae

Zijian Tang was born in Shanghai, China on 16 August 1973. He obtained his Bachelor degree in electrical engineering in 1998 from Shanghai Jiao Tong University. In 1999, he came to the Netherland. After one year of learning the Dutch language, he got enrolled in the department of electrical engineering at Delft University of Technology in 2000, and graduated from the Network and Systems group in 2003 with the Master of Science diploma. Since then, he was employed by the same group as a researcher leading to the PhD. degree.

In the summer of 2003, he took an internship at NATLAB (Philips R&D center) in Eindhoven with the Connectivity Systems & Networks group. Since Sept. 2007, he has been working at The Mathworks as an application engineer.

Zijian Tang has won the best student paper award (Honorable Mention) during the 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2007 for the paper "Receiver Design For Single-Carrier Transmission Over Time-Varying Channels" co-authored with Dr. Geert Leus.

