CHAPTER

4 Estimation of Time-Varying Channels – A Block Approach

Geert Leus[1], **Zijian Tang**[2], **Paolo Banelli**[3]
[1]*Delft University of Technology, Delft, The Netherlands*
[2]*TNO Defence, Security and Safety, The Hague, The Netherlands*
[3]*University of Perugia, Perugia, Italy*

4.1 INTRODUCTION

For coherent detection in a wireless communication system, channel state information (CSI) is indispensable. Channel estimation has drawn tremendous attention in the literature (see Tong, Sadler, & Dong, 2004 and references therein), where the pilot-aided method is one of the most intensively studied approaches. This method is especially attractive for time-varying channels because of their short coherence time.

In this chapter, we will address pilot-aided channel estimation for both orthogonal frequency division multiplexing (OFDM) and single-carrier systems, where pilots are inserted in the frequency domain and time domain, respectively. We study these two systems under one framework because in the context of channel estimation, both systems can be characterized by data models of the same form. More specifically, the received samples can be expressed as the joint effect of the information part (due to the pilots), the interference part (due to the unknown data symbols), and the noise. Consequently, our task is to design a channel estimator that can combat both the interference and the noise. Such a data model is typical for OFDM over time-varying channels, where due to the Doppler effect, the orthogonality between the subcarriers is destroyed, and the channel matrix in the frequency domain becomes effectively a diagonally dominant yet full matrix instead of a diagonal matrix. As a result, the received frequency-domain samples depend on both the pilots and the unknown data symbols. For single-carrier systems, the channel matrix in the time domain is a strictly banded matrix if a finite impulse response (FIR) assumption for the channel is applied, and therefore, we can, in practice, find some received samples that solely depend on the pilots. However, it is sometimes beneficial to also consider received samples that depend on the unknown data symbols as well, to better suppress the interference and the noise. In any case, the resulting data model for single-carrier systems looks very similar to the data model for OFDM systems, and similar channel estimation techniques can be applied. Note that the considered data model can also account for superimposed pilot schemes (Ghogho & Swami, 2006; He & Tugnait, 2007), where the pilots and the data symbols coexist on the same subcarriers or time instants.

Whether we are dealing with OFDM or single-carrier systems, estimating a time-varying channel implies estimating a large number of unknowns, making the channel estimation problem much more

Wireless Communications Over Rapidly Time-Varying Channels. DOI: 10.1016/B978-0-12-374483-8.00004-2

difficult than in the time-invariant case. As a remedy, we adopt in this chapter a parsimonious model, referred to as the basis expansion model (BEM), to approximate the time variation of the channel (see Section 1.6.1). If the BEM is accurate with negligible approximation error, channel estimation can be achieved by just estimating the BEM coefficients, which are much smaller in number than the actual unknowns, i.e., the channel tap values at different time instants.

In the remainder of the chapter, we first discuss the system and channel model in Section 4.2. In Section 4.3, we then present channel estimation algorithms within a single OFDM symbol/time block. We indicate how to position the pilots, where to select observation samples, and what is the best channel estimation strategy. In Section 4.4, we extend these methods to situations where multiple OFDM symbols/time blocks are utilized simultaneously. In this case, the position of the pilots plays an important role in the performance. Extensions to multiple-antenna systems are considered in Section 4.5 and adaptive channel estimation is briefly discussed in Section 4.6. We conclude this chapter in Section 4.7.

4.2 SYSTEM AND CHANNEL MODEL

4.2.1 System Model

Let us start with the channel input/output relationship in discrete form. The channel is assumed to be a doubly selective (doubly dispersive) channel that can be modeled by an FIR filter, which takes the effect of the transmitter filter, the propagation paths, and the receiver filter into account. Let us use $h[n,m]$ to denote the mth channel tap at the nth time instant and let us assume that the maximal channel order is $M-1$. For such a channel, if we use $s[n]$ and $r[n]$ to represent the transmitted and received signal at the nth time instant, respectively, they are related to each other through the channel as

$$r[n] = \sum_{m=0}^{M-1} h[n,m]s[n-m] + w[n], \tag{4.1}$$

where $w[n]$ stands for the additive noise.

Let us consider a block-wise transmission scheme, where we handle the transmitted and received sequences in blocks. Suppose the jth data block $\mathbf{a}[j]$ groups K data symbols and can be expressed as

$$\mathbf{a}[j] \triangleq (a[jK] \cdots a[(j+1)K-1])^T.$$

Before transmission, $\mathbf{a}[j]$ is first transformed by a linear precoding matrix $\mathbf{T}$ of size $(K+M-1)\times K$. The resulting $(K+M-1)\times 1$ vector is given by

$$\mathbf{s}[j] \triangleq (s[j(K+M-1)] \cdots s[(j+1)(K+M-1)-1])^T \triangleq \mathbf{T}\mathbf{a}[j].$$

Note that $\mathbf{T}$ introduces a redundancy equal to $M-1$ at the transmitter. Such a redundancy is useful in many applications, e.g., to exploit the channel diversity, or to better combat the interference coming from adjacent blocks due to the channel memory, as well as for many other purposes (Scaglione, Giannakis, & Barbarossa, 1999a,b). We will come back to this issue later on.

At the receiver, the received sample stream $r[n]$ is partitioned in blocks accordingly: $\mathbf{r}[j] \triangleq (r[j(K+M-1)] \cdots r[(j+1)(K+M-1)-1])^T$. Based on the FIR property of the channel, we can

rewrite the channel I/O relationship of (4.1) in a block form as

$$\mathbf{r}[j] = \tilde{\mathbf{H}}[j]\left(\mathbf{s}_{\mathrm{IBI}}^T[j]\ \mathbf{s}^T[j]\right)^T + \mathbf{w}[j], \tag{4.2}$$

where $\mathbf{w}[j] \triangleq (w[j(K+M-1)] \cdots w[(j+1)(K+M-1)-1])^T$, and $\tilde{\mathbf{H}}[j]$ is the $(K+M-1)\times(K+2M-2)$ channel matrix representing the convolutive operation of the channel, which is given by $[\tilde{\mathbf{H}}[j]]_{nm} \triangleq h[j(K+M-1)+n-1, M-1+n-m]$. Because the channel has a memory of $M-1$, $\mathbf{r}[j]$ will not only depend on $\mathbf{s}[j]$, but also on the last $M-1$ symbols from the previous block $\mathbf{s}[j-1]$, which is denoted in (4.2) as $\mathbf{s}_{\mathrm{IBI}}[j] \triangleq (s[j(K+M-1)-M+1] \cdots s[j(K+M-1)-1])^T$. The effect of $\mathbf{s}_{\mathrm{IBI}}[j]$ is also known as the interblock interference (IBI) as indicated by its subscript. We can further rewrite (4.2) by splitting $\tilde{\mathbf{H}}[j]$ into its first $M-1$ columns and last $K+M-1$ columns, denoted by $\mathbf{H}_{\mathrm{IBI}}[j]$ and $\mathbf{H}[j]$, respectively, which leads to

$$\mathbf{r}[j] = \mathbf{H}_{\mathrm{IBI}}[j]\mathbf{s}_{\mathrm{IBI}}[j] + \mathbf{H}[j]\mathbf{s}[j] + \mathbf{w}[j].$$

Assuming perfect block and symbol synchronization, we first apply a linear decoder at the receiver, in the form of a $K\times(K+M-1)$ matrix $\mathbf{R}$. This results in

$$\mathbf{y}[j] \triangleq \mathbf{R}\mathbf{r}[j] = \mathbf{R}\mathbf{H}_{\mathrm{IBI}}[j]\mathbf{s}_{\mathrm{IBI}}[j] + \mathbf{R}\mathbf{H}[j]\mathbf{s}[j] + \mathbf{z}[j], \tag{4.3}$$

where $\mathbf{z}[j] \triangleq \mathbf{R}\mathbf{w}[j]$.

The decoder $\mathbf{R}$ as well as the precoder $\mathbf{T}$ depend on the specific (de)modulation scheme. Two particular cases are discussed below, namely OFDM and single-carrier modulation. For both systems, different forms of transmitter redundancy can be adopted, such as the use of a cyclic prefix (CP), zero padding (ZP), or nonzero padding (NZP). We will restrict ourselves to the CP case. Other forms of transmitter redundancy can be derived in a similar way.

4.2.1.1 OFDM System with CP

In an OFDM system where a CP is embedded at the transmitter, the precoder $\mathbf{T}$ and decoder $\mathbf{R}$ can be expressed as

$$\mathbf{T} = \mathbf{T}_{\mathrm{CP}}\mathbf{W}_K^H,$$

$$\mathbf{R} = \mathbf{W}_K\mathbf{R}_{\mathrm{CP}},$$

where $\mathbf{W}_K$ denotes the K-point unitary DFT matrix, i.e., $[\mathbf{W}_K]_{kl} \triangleq (1/\sqrt{K})\mathrm{e}^{-\sqrt{-1}2\pi(k-1)(l-1)/K}$, the $(K+M-1)\times K$ matrix $\mathbf{T}_{\mathrm{CP}} \triangleq ((\mathbf{0}_{(M-1)\times(K-M+1)}\ \mathbf{I}_{M-1})^T\ \mathbf{I}_K)^T$ appends a CP, and the $K\times(K+M-1)$ matrix $\mathbf{R}_{\mathrm{CP}} \triangleq (\mathbf{0}_{K\times(M-1)}\ \mathbf{I}_K)$ discards the part of the received block corresponding to the CP. Because the transmitter redundancy equals the maximal channel order M, the IBI disappears and (4.3) becomes

$$\mathbf{y}[j] = \mathbf{W}_K\mathbf{R}_{\mathrm{CP}}\mathbf{H}[j]\mathbf{T}_{\mathrm{CP}}\mathbf{W}_K^H\mathbf{a}[j] + \mathbf{z}[j] = \mathbf{W}_K\mathbf{H}_{\mathrm{c}}[j]\mathbf{W}_K^H\mathbf{a}[j] + \mathbf{z}[j] = \mathbf{H}_{\mathrm{d}}[j]\mathbf{a}[j] + \mathbf{z}[j], \tag{4.4}$$

where $\mathbf{H}_{\mathrm{c}}[j] \triangleq \mathbf{R}_{\mathrm{CP}}\mathbf{H}[j]\mathbf{T}_{\mathrm{CP}}$ is the $K\times K$ time-domain channel matrix, and $\mathbf{H}_{\mathrm{d}}[j] \triangleq \mathbf{W}_K\mathbf{H}_{\mathrm{c}}[j]\mathbf{W}_K^H$ is the $K\times K$ frequency-domain channel matrix. If the channel is static or only slowly changing such that the time variation of the channel within an OFDM symbol can be neglected, $\mathbf{H}_{\mathrm{c}}[j]$ will be a circulant matrix (hence the subscript "c") with $(h[0] \cdots h[M-1]\ \mathbf{0}_{1\times(K-M+1)})^T$ on its first column and as a result $\mathbf{H}_{\mathrm{d}}[j]$ will be a diagonal matrix (hence the subscript "d") with

$\sqrt{K}\mathbf{W}_K(h[0]\ \cdots\ h[M-1]\ \mathbf{0}_{1\times(K-M+1)})^T$ on its diagonal. Note that we have dropped the time index n in the channel tap $h[n,m]$ because of the time-invariance assumption.

However, when the channel is varying faster, the circularity of $\mathbf{H}_c[j]$ is destroyed, and thus $\mathbf{H}_c[j]$ is not diagonalizable by (I)DFT operations. In principle, $\mathbf{H}_d[j]$ becomes a full matrix, where the nonzero off-diagonal entries induce intercarrier interference (ICI). The entries along the antidiagonal direction basically indicate how much the bandwidth will spread due to mobility-induced Doppler shifts. Because, in practice, this Doppler spread is limited, we can assume that most of the power in $\mathbf{H}_d[j]$ is concentrated on and close to the main diagonal, and it will gradually reduce in the antidiagonal direction. Hence, we may assume that $\mathbf{H}_d[j]$ is approximately circularly banded. This assumption has also been advocated in Stamoulis, Diggavi, & Al-Dhahir (2002) and Cai & Giannakis (2003).

4.2.1.2 *Single-Carrier System with CP*

In a single-carrier system with a CP, the (I)DFT operations are omitted at the transmitter and receiver, and the precoder $\mathbf{T}$ and decoder $\mathbf{R}$ simply become

$$\mathbf{T} = \mathbf{T}_{\text{CP}},$$
$$\mathbf{R} = \mathbf{R}_{\text{CP}}.$$

As in the OFDM case, the IBI is suppressed and (4.3) can be expressed as

$$\mathbf{y}[j] = \mathbf{R}_{\text{CP}}\mathbf{H}[j]\mathbf{T}_{\text{CP}}\mathbf{a}[j] + \mathbf{z}[j] = \mathbf{H}_c[j]\mathbf{a}[j] + \mathbf{z}[j]. \tag{4.5}$$

It is, in many applications, common to transform the single-carrier system model in (4.5) from the time domain to the frequency domain just like OFDM. For time-invariant channels, this operation will enable a simple one-tap channel equalizer (Falconer, Ariyavisitakul, Benyamin-Seeyar, & Eidson, 2002) because the channel matrix in the frequency domain $\mathbf{H}_d[j]$ will become diagonal, as explained earlier. Even if the channel is time-varying, it can still be attractive to consider a frequency-domain equalizer because of the circularly banded assumption on $\mathbf{H}_d[j]$, an idea that has been explored in Schniter & Liu (2003) and Tang & Leus (2008) (see also Chapter 6).

For channel estimation, however, it is not necessary to transform the single-carrier data model from the time domain to the frequency domain. The reason for this is that we will always work in the domain where the pilots will be embedded, which is in the frequency domain for the OFDM system and in the time domain for the single-carrier system.

4.2.2 BEM Channel Model

In the OFDM system as well as in the single-carrier system, estimating $h[n,m]$ requires the estimation of KM unknowns per OFDM symbol/time block, which is quite a lot. However, due to the fact that there is some correlation among these unknowns, the problem can be reduced. One approach to reduce the number of unknowns is to use a BEM to model the time variation of the channel (see also Section 1.6.1). To explain the BEM, let us assume that the time variation of the mth channel tap is smooth, and thus $h[n,m]$ is correlated in the time index n. Then, we can accurately model K consecutive samples of the mth channel tap,

e.g., $h[j(K+M-1)+M-1,m],\ldots,h[(j+1)(K+M-1)-1,m]$, as

$$\begin{pmatrix} h[j(K+M-1)+M-1,m] \\ \vdots \\ h[(j+1)(K+M-1)-1,m] \end{pmatrix} \approx \underbrace{\begin{pmatrix} \mathbf{u}_0 & \cdots & \mathbf{u}_{I-1} \end{pmatrix}}_{\mathbf{U}} \begin{pmatrix} c_{0,j}[m] \\ \vdots \\ c_{I-1,j}[m] \end{pmatrix}, \tag{4.6}$$

where $I-1$ is the BEM order, the $K \times 1$ vector $\mathbf{u}_i$ is the ith BEM function, and $c_{i,j}[m]$ is the ith BEM coefficient of the mth channel tap within the jth block. Note that the BEM matrix $\mathbf{U}$ is a tall matrix because I is generally much smaller than K. It is furthermore predetermined and independent of the channel. The accuracy of (4.6) will depend on the design of the BEM matrix $\mathbf{U}$ and the choice of the BEM order $I-1$.

Different BEM designs are documented in various articles such as the discrete Karhunen-Loève BEM (DKL-BEM) (Visintin, 1996; Yip & Ng, 1997; Haykin, 1996; Teo & Ohno, 2005), the discrete prolate spheroidal BEM (DPS-BEM) (Zemen & Mecklenbräuker, 2005), the complex exponential BEM (CE-BEM) (Tsatsanis & Giannakis, 1996; Cirpan & Tsatsanis, 1999), and the polynomial BEM (P-BEM) (Borah & Hart, 1999a,b; Tomasin, Gorokhov, Yang, & Linnartz, 2005). Note that the CE-BEM can further be categorized into two types: the critically sampled CE-BEM (CCE-BEM) (Guillaud & Slock, 2003; Ma & Giannakis, 2003; Ma, Giannakis, & Ohno, 2003; Leus & Moonen, 2003; Kannu & Schniter, 2005) and the oversampled CE-BEM (OCE-BEM) (Thomas & Vook, 2000; Leus, 2004; Cui, Tellambura, & Wu, 2005). BEM designs other than the above are also reported, e.g., Zakharov, Tozer, & Adlard (2004) use a spline approach. Besides, it is also possible to combine the above BEMs for different purposes (Stamoulis, Diggavi, & Al-Dhahir, 2002; Nicoli, Simeone, & Spagnolini, 2003; Gorokhov & Linnartz, 2004). A comparison of the modeling performances of some BEMs is given in Zemen and Mecklenbräuker (2005) and in Tang (2007). It is noteworthy that all the above BEMs can, in principle, be utilized in this chapter with one minor adaptation: if $\tilde{\mathbf{U}}$ is one of the original BEM designs, we will construct a new BEM out of it by adopting

$$\mathbf{U} = \tilde{\mathbf{U}}\mathbf{Q}, \tag{4.7}$$

where $\mathbf{Q}$ is an $I \times I$ matrix that makes the columns of $\mathbf{U}$ orthonormal, i.e., $\mathbf{U}^H\mathbf{U} = \mathbf{I}_I$ (note that some original BEM designs already satisfy this property but not all of them). Such matrices will improve the numerical stability of the proposed algorithms.

Repeating the BEM approximation procedure of (4.6) for all channel taps, we obtain

$$\mathbf{h}[j] \approx (\mathbf{U} \otimes \mathbf{I}_M)\mathbf{c}[j], \tag{4.8}$$

where $\mathbf{h}[j] \triangleq (\mathbf{h}_0^T[j] \cdots \mathbf{h}_{K-1}^T[j])^T$, with $\mathbf{h}_k[j] \triangleq (h[j(K+M-1)+M-1+k,0] \cdots h[j(K+M-1)+M-1+k,M-1])^T$, stacks all channel taps in the jth block and where $\mathbf{c}[j] \triangleq (\mathbf{c}_0^T[j] \cdots \mathbf{c}_{I-1}^T[j])^T$, with $\mathbf{c}_i[j] \triangleq (c_{i,j}[0] \cdots c_{i,j}[M-1])^T$, stacks the BEM coefficients of all channel taps in the jth block. The notation $\otimes$ stands for the Kronecker product. By means of (4.8), we are able to reduce the total number of unknown channel parameters from KM to IM with $I \ll K$.

We can now also express the time-domain channel matrix $\mathbf{H}_c[j]$ and the frequency-domain channel matrix $\mathbf{H}_d[j]$ using the BEM. Observing that each diagonal in $\mathbf{H}_c[j]$ corresponds to one channel tap,

we can derive after some algebra that

$$\mathbf{H}_{\mathrm{c}}[j] = \sum_{i=0}^{I-1} \mathbf{U}_{\mathrm{d},i}\mathbf{C}_{\mathrm{c},i}[j], \tag{4.9}$$

where $\mathbf{U}_{\mathrm{d},i}$ is a diagonal matrix with $\mathbf{u}_i$ on its diagonal, and $\mathbf{C}_{\mathrm{c},i}[j]$ is a circulant matrix with $(\mathbf{c}_i^T[j]\,\mathbf{0}_{1\times(K-M)})^T$ on its first column. The frequency-domain channel matrix $\mathbf{H}_{\mathrm{d}}[j]$ can then be expressed as

$$\mathbf{H}_{\mathrm{d}}[j] = \mathbf{W}_K\mathbf{H}_{\mathrm{c}}[j]\mathbf{W}_K^H = \sum_{i=0}^{I-1} \mathbf{W}_K\mathbf{U}_{\mathrm{d},i}\mathbf{W}_K^H\mathbf{W}_K\mathbf{C}_{\mathrm{c},i}[j]\mathbf{W}_K^H = \sum_{i=0}^{I-1} \mathbf{U}_{\mathrm{c},i}\mathbf{C}_{\mathrm{d},i}[j], \tag{4.10}$$

where $\mathbf{U}_{\mathrm{c},i} \triangleq \mathbf{W}_K\mathbf{U}_{\mathrm{d},i}\mathbf{W}_K^H$ is a circulant matrix with $(1/\sqrt{K})\mathbf{u}_i^T\mathbf{W}_K^H$ on its first row and $\mathbf{C}_{\mathrm{d},i}[j] \triangleq \mathbf{W}_K\mathbf{C}_{\mathrm{c},i}[j]\mathbf{W}_K^H$ is a diagonal matrix with $\sqrt{K}\mathbf{W}_K\left(\mathbf{c}_i^T[j]\,\mathbf{0}_{1\times(K-M)}\right)^T$ on its diagonal.

For the channel estimation schemes to be discussed in the ensuing sections, we will use the BEM channel matrices defined in (4.9) and (4.10) instead of the true channel matrices in the system model, and we will estimate the channel by estimating the BEM coefficients. The BEM modeling error will not be taken into account in the design of the channel estimators. This is motivated by the fact that if the BEM is accurate, the modeling error is usually on the order of 10^{-4} as reported in Zemen and Mecklenbräuker (2005) and Tang (2007), which is much smaller than the typical channel noise level. However, if the BEM is not accurate enough, then we will never obtain a reliable estimate, even if we design a channel estimator that takes the BEM modeling error into account, simply because the BEM itself is not accurate. Finally note that although the BEM modeling error is not taken into account in our channel estimator design, its effect will be considered in the simulations, where we will compare the estimated channel with the true channel instead of the BEM channel.

4.3 CHANNEL ESTIMATION BASED ON A SINGLE BLOCK

4.3.1 Introduction

In the previous section, we have shown how a communications scheme can be split into temporal blocks. In this section, we will discuss block-based channel estimation, or in other words, the channel is estimated every time a block of samples is received. This strategy is especially attractive for time-varying channels, where, due to a short coherence time, it is not possible to gather a large amount of received samples for which the channel is correlated. Because we only focus on a single received block in this section, we will drop the block index j for the sake of simplicity.

We will confine ourselves to pilot-aided channel estimators. For one transmitted symbol block $\mathbf{a}$, let $\mathbf{p}$ denote the pilot symbols and $\mathbf{d}$ denote the unknown data symbols. Defining $\mathscr{P}$ as the set of positions of the pilot symbols and $\mathscr{D}$ as the set of positions of the unknown data symbols, $\mathbf{p}$ has length $|\mathscr{P}|$ and $\mathbf{d}$ has length $|\mathscr{D}|$, where $|\mathscr{S}|$ denotes the cardinality of the set $\mathscr{S}$. Such a notation can account for various pilot insertion schemes. One such scheme is a multiplexed pilot scheme, where the pilot and unknown data symbols occupy different locations, i.e., $\mathscr{P} \cap \mathscr{D} = \emptyset$ and $\mathscr{P} \cup \mathscr{D} = \{0,\ldots,K-1\}$. But it also allows us to model a superimposed pilot scheme, where the pilot and unknown data symbols coexist on the same positions, i.e., $\mathscr{P} = \mathscr{D} = \{0,\ldots,K-1\}$. Let us at this point also introduce the

notation $\mathbf{x}^{\{\mathscr{S}\}}$ to represent the subvector that collects the elements of the vector $\mathbf{x}$ with indices in the set $\mathscr{S}$, as well as $\mathbf{X}^{\{\mathscr{S}_1,\mathscr{S}_2\}}$ to represent the submatrix of $\mathbf{X}$ containing the rows with indices in the set $\mathscr{S}_1$ and the columns with indices in the set $\mathscr{S}_2$, where we use a colon instead of an index set if all elements are considered. This notation will be used throughout this chapter.

The multiplexed and superimposed pilot schemes will be unified under the same framework. Although no extra bandwidth is consumed with the superimposed pilot scheme, it is heavily influenced by the interference from the unknown data symbols. To some extent this also happens with the multiplexed pilot scheme because the delay (for the single-carrier system) or Doppler (for the OFDM system) spread of the channel will introduce some mixing between the pilot symbols and the unknown data symbols, but this mixing is much smaller than with the superimposed pilot scheme. For this reason, we will mainly focus on the multiplexed pilot scheme in this section, and we will write a remark on the superimposed pilot scheme from time to time. However, the expressions are basically the same for both pilot insertion schemes.

In the ensuing part, we will first rewrite the earlier data models as a function of the pilot symbols. Based on these data models, we will then discuss various channel estimation techniques and their characteristics. Some simulation results will be given at the end of this section.

4.3.2 Channel Estimation Data Model

As already indicated, we focus mainly on multiplexed pilot schemes in this chapter. This means that we consider time-multiplexed training for the single-carrier system and frequency-multiplexed training for the OFDM system. Let us assume that the pilot symbols are grouped in G clusters, each of length $P+1$. For the gth pilot cluster $\mathbf{p}_g$, suppose $\mathscr{P}_g \triangleq \{P_g,\ldots,P_g+P\}$ denotes the set of indices that contains all the pilot positions, with P_g standing for its starting position. Hence, $\mathbf{p}_g$ is related with the transmitted symbol block $\mathbf{a}$ as $\mathbf{p}_g \triangleq \mathbf{a}^{\{\mathscr{P}_g\}}$. Below, we will describe how to proceed for the two considered systems.

4.3.2.1 *Single-Carrier System with CP*

Because the time-domain channel matrix $\mathbf{H}_{\mathrm{c}}$ is circularly banded with a band of length M (the discrete delay spread) on and below the main diagonal, we can assign to the gth pilot cluster $\mathbf{p}_g$ the observation samples $\mathbf{y}^{\{\mathscr{O}_g\}}$ whose indices are collected in the set

$$\mathscr{O}_g \triangleq \{P_g+\ell,\ldots,P_g+P+M-1-\ell\}, \tag{4.11}$$

where ℓ is an integer design parameter determining the number of observation samples that we want to take into account in our channel estimation. We will come back to this issue later on.

According to the input–output relationship in (4.5) and assuming that there is enough space in between two nonzero pilot symbols from different pilot clusters, i.e., the minimal number of samples in between two nonzero pilot symbols from different pilot clusters, denoted as Δ, satisfies $\Delta \geq M-\ell-1$, we can express $\mathbf{y}^{\{\mathscr{O}_g\}}$ as

$$\mathbf{y}^{\{\mathscr{O}_g\}} = \mathbf{H}_{\mathrm{c}}^{\{\mathscr{O}_g,\mathscr{P}_g\}}\mathbf{p}_g + \mathbf{H}_{\mathrm{c}}^{\{\mathscr{O}_g,\mathscr{D}_g\}}\mathbf{d}_g + \mathbf{z}^{\{\mathscr{O}_g\}}, \tag{4.12}$$

where $\mathscr{D}_g$ represents the set of indices of the unknown data symbols that are present in $\mathbf{y}^{\{\mathscr{O}_g\}}$, and $\mathbf{d}_g$ collects those unknown data symbols present in $\mathbf{y}^{\{\mathscr{O}_g\}}$, i.e., $\mathbf{d}_g \triangleq \mathbf{a}^{\{\mathscr{D}_g\}}$. The structure of the channel matrices in (4.12) is illustrated in Fig. 4.1.

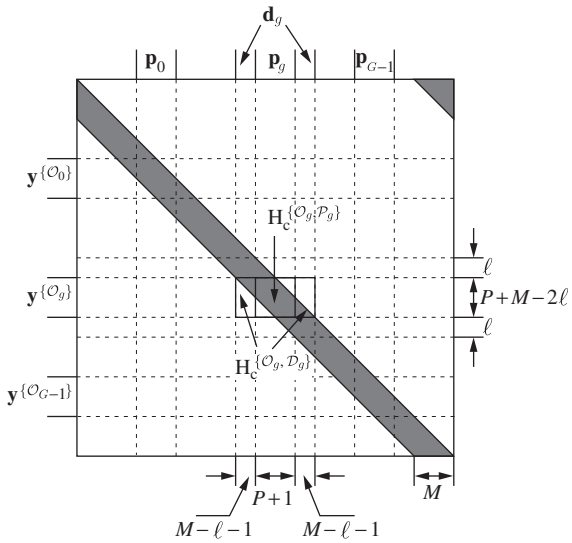

FIGURE 4.1

The partitioning of the time-domain channel matrix $\mathbf{H}_{\mathrm{c}}$ into $\mathbf{H}_{\mathrm{c}}^{\{\mathcal{O}_g,\mathcal{P}_g\}}$ and $\mathbf{H}_{\mathrm{c}}^{\{\mathcal{O}_g,\mathcal{D}_g\}}$. The dark shaded areas represent large values, whereas the white areas represent zeros.

Using the BEM expression in (4.9), we can now rewrite (4.12) as

$$\mathbf{y}^{\{\mathcal{O}_g\}} = \sum_{i=0}^{I-1} \mathbf{U}_{\mathrm{d},i}^{\{\mathcal{O}_g,\mathcal{O}_g\}} \mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,\mathcal{P}_g\}} \mathbf{p}_g + \underbrace{\sum_{i=0}^{I-1} \mathbf{U}_{\mathrm{d},i}^{\{\mathcal{O}_g,\mathcal{O}_g\}} \mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,\mathcal{D}_g\}} \mathbf{d}_g}_{\mathbf{i}_g} + \mathbf{z}^{\{\mathcal{O}_g\}}, \tag{4.13}$$

where $\mathbf{U}_{\mathrm{d},i}^{\{\mathcal{O}_g,\mathcal{O}_g\}}$ is the submatrix of the diagonal matrix $\mathbf{U}_{\mathrm{d},i}$ consisting of the diagonal elements on the positions $\mathcal{O}_g$, $\mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,\mathcal{P}_g\}}$ consists of the columns of $\mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,:\}}$ corresponding to the positions of the pilot symbols in $\mathbf{p}_g$, and $\mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,\mathcal{D}_g\}}$ consists of the columns of $\mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,:\}}$ corresponding to the positions of the unknown data symbols in $\mathbf{d}_g$. Note that $\mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,\mathcal{P}_g\}}$ and $\mathbf{C}_{\mathrm{c},i}^{\{\mathcal{O}_g,\mathcal{D}_g\}}$ have the same structure as $\mathbf{H}_{\mathrm{c}}^{\{\mathcal{O}_g,\mathcal{P}_g\}}$ and $\mathbf{H}_{\mathrm{c}}^{\{\mathcal{O}_g,\mathcal{D}_g\}}$ in Fig. 4.1.

The significance of ℓ becomes clearer from the definitions above. By adjusting the value of ℓ, the number of observation samples varies, and accordingly, the channel estimator can deal with different amounts of information regarding the pilot symbols and the unknown data symbols. For instance, in Ma, Giannakis, and Ohno (2003), ℓ is chosen as $\ell = M-1$, so that $\mathcal{D}_g = \emptyset$ and thus $\mathbf{H}_{\mathrm{c}}^{\{\mathcal{O}_g,\mathcal{D}_g\}}$ vanishes. Moreover, under the assumption that $P \geq M-1$, we then have that $\mathcal{O}_g = \{P_g+M-1,\ldots,P_g+P\}$ and thus $\mathbf{y}^{\{\mathcal{O}_k\}}$ contains the maximum number of received samples that depend on the pilot symbols without any interference from the unknown data symbols. Taking ℓ larger than $M-1$ degrades the

performance but lowers the complexity. The largest value of ℓ is given by $(P+M-1)/2$ (assume $P+M-1$ is even), which is the case where only a single observation sample per pilot cluster is selected. Rousseaux, Leus, Stoica, and Moonen (2004) and Rousseaux, Leus, and Moonen (2006) chose $\ell = 0$, which means that $\mathscr{O}_g = \{P_g, \ldots, P_g + P + M - 1\}$ and thus $\mathbf{y}^{\{\mathscr{O}_k\}}$ contains the maximum number of received samples that depend on both the pilot and the interfering unknown data symbols. Taking ℓ smaller than 0 will not change the performance and only increases the complexity. Taking the earlier condition $\Delta \geq M - \ell - 1$ into account, which is required for (4.12) to hold, the smallest value of ℓ is given by $\max\{0, M - \Delta - 1\}$. In the sequel, we will follow an approach similar to the approach of Leus & van der Veen (2005), and let ℓ assume an arbitrary integer value, which we bound by $\max\{0, M - \Delta - 1\} \leq \ell \leq (P + M - 1)/2$.

Rewriting (4.13) as a function of the BEM coefficients $\mathbf{c} = (\mathbf{c}_0^T \cdots \mathbf{c}_{I-1}^T)^T$, we obtain

$$\mathbf{y}^{\{\mathscr{O}_g\}} = \mathbf{A}_g \mathbf{c} + \mathbf{i}_g + \mathbf{z}^{\{\mathscr{O}_g\}}, \tag{4.14}$$

with

$$\mathbf{A}_g \triangleq \left(\mathbf{U}_{\mathrm{d},0}^{\{\mathscr{O}_g,\mathscr{O}_g\}} \mathbf{P}_{\mathrm{c},g}^{\{\mathscr{O}_g,\mathscr{C}_g\}} \quad \cdots \quad \mathbf{U}_{\mathrm{d},I-1}^{\{\mathscr{O}_g,\mathscr{O}_g\}} \mathbf{P}_{\mathrm{c},g}^{\{\mathscr{O}_g,\mathscr{C}_g\}} \right), \tag{4.15}$$

where $\mathbf{P}_{\mathrm{c},g}$ is a circulant matrix with $(\mathbf{p}_g^T \ \mathbf{0}_{1\times(K-P-1)})^T$ on its first column, and $\mathscr{C}_g$ is the set of column indices of $\mathbf{P}_{\mathrm{c},g}$ that are hit by every $\mathbf{c}_i$, i.e., $\mathscr{C}_g \triangleq \{P_g, \ldots, P_g + M - 1\}$. Note that $\mathscr{C}_g$ has the same starting point as $\mathscr{P}_g$, but it has length M instead of $P+1$. To derive (4.15), we have made use of the commutativity of the circular convolution, meaning that $\mathbf{C}_{\mathrm{c},i}^{\{:,\mathscr{P}_g\}} \mathbf{p}_g = \mathbf{P}_{\mathrm{c},g}^{\{:,\mathscr{C}_g\}} \mathbf{c}_i$.

Stacking the results obtained in (4.14) for all G pilot clusters, $\mathbf{y}^{\{\mathscr{O}\}} \triangleq (\mathbf{y}^{\{\mathscr{O}_0\}T} \cdots \mathbf{y}^{\{\mathscr{O}_{G-1}\}T})^T$, where the index set $\mathscr{O}$ is given by $\mathscr{O} \triangleq \{\mathscr{O}_0, \ldots, \mathscr{O}_{G-1}\}$, we obtain

$$\mathbf{y}^{\{\mathscr{O}\}} = \mathbf{A}\mathbf{c} + \mathbf{i} + \mathbf{z}^{\{\mathscr{O}\}}, \tag{4.16}$$

where $\mathbf{i}$ and $\mathbf{z}^{\{\mathscr{O}\}}$ are similarly defined as $\mathbf{y}^{\{\mathscr{O}\}}$, and

$$\begin{aligned} \mathbf{A} &\triangleq (\mathbf{A}_0^T \cdots \mathbf{A}_{G-1}^T)^T \\ &= \begin{pmatrix} \mathbf{U}_{\mathrm{d},0}^{\{\mathscr{O}_0,\mathscr{O}_0\}} \mathbf{P}_{\mathrm{c},0}^{\{\mathscr{O}_0,\mathscr{C}_0\}} & \cdots & \mathbf{U}_{\mathrm{d},I-1}^{\{\mathscr{O}_0,\mathscr{O}_0\}} \mathbf{P}_{\mathrm{c},0}^{\{\mathscr{O}_0,\mathscr{C}_0\}} \\ \vdots & & \vdots \\ \mathbf{U}_{\mathrm{d},0}^{\{\mathscr{O}_{G-1},\mathscr{O}_{G-1}\}} \mathbf{P}_{\mathrm{c},G-1}^{\{\mathscr{O}_{G-1},\mathscr{C}_{G-1}\}} & \cdots & \mathbf{U}_{\mathrm{d},I-1}^{\{\mathscr{O}_{G-1},\mathscr{O}_{G-1}\}} \mathbf{P}_{\mathrm{c},G-1}^{\{\mathscr{O}_{G-1},\mathscr{C}_{G-1}\}} \end{pmatrix}. \end{aligned} \tag{4.17}$$

4.3.2.2 OFDM System with CP

In contrast to the time-domain channel matrix $\mathbf{H}_{\mathrm{c}}$, the frequency-domain channel matrix $\mathbf{H}_{\mathrm{d}}$ basically is a full matrix, but with most of its power located on or close to the main diagonal, or in other words, the matrix $\mathbf{H}_{\mathrm{d}}$ is approximately circularly banded. There is only one case where the matrix $\mathbf{H}_{\mathrm{d}}$ is exactly circularly banded, and that is when the (C)CE-BEM is used to model the channel. In that case, $\mathbf{H}_{\mathrm{d}}$ will have I nonzero diagonals, where $I-1$ is the order of the (C)CE-BEM. However, the problem with this approach is that the (C)CE-BEM does not provide a good fit to the actual channel (Zemen & Mecklenbräuker, 2005; Tang, 2007).

Due to the fact that $\mathbf{H}_{\mathrm{d}}$ is a full matrix, it is less obvious to choose which received samples to use for channel estimation. However, similar to Section 4.3.2.1, if we assume that $\mathbf{H}_{\mathrm{d}}$ is approximately circularly banded with a band of length D (the significant discrete Doppler spread) centered around the main diagonal (assume $D-1$ is even), we can assign to the gth pilot cluster $\mathbf{p}_g$ the observation samples $\mathbf{y}^{\{\mathscr{O}_g\}}$ whose indices are collected in the set

$$\mathscr{O}_g \triangleq \{P_g - (D-1)/2 + \ell, \ldots, P_g + P + (D-1)/2 - \ell\}, \tag{4.18}$$

where ℓ is again an integer design parameter that controls the number of observation samples that will be used for channel estimation. Note that in contrast to other chapters, we use D here to indicate the double-sided discrete Doppler spread instead of the single-sided one, in order to better reflect the duality between M for the single-carrier system with CP and D for the OFDM system with CP.

According to the input–output relationship in (4.4), we can express $\mathbf{y}^{\{\mathscr{O}_g\}}$ as

$$\mathbf{y}^{\{\mathscr{O}_g\}} = \mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_g,\mathscr{P}\}}\mathbf{p} + \mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_g,\mathscr{D}\}}\mathbf{d} + \mathbf{z}^{\{\mathscr{O}_g\}}. \tag{4.19}$$

The difference with (4.12) for the time-domain case is that now all the pilot and unknown data symbols are present in $\mathbf{y}^{\{\mathscr{O}_g\}}$, due to the fact that $\mathbf{H}_{\mathrm{d}}$ is basically a full matrix. Moreover, there is no special constraint on Δ, which represents the minimal number of samples in between two nonzero pilot symbols from different pilot clusters. The structure of the channel matrices in (4.19) is illustrated in Fig. 4.2.

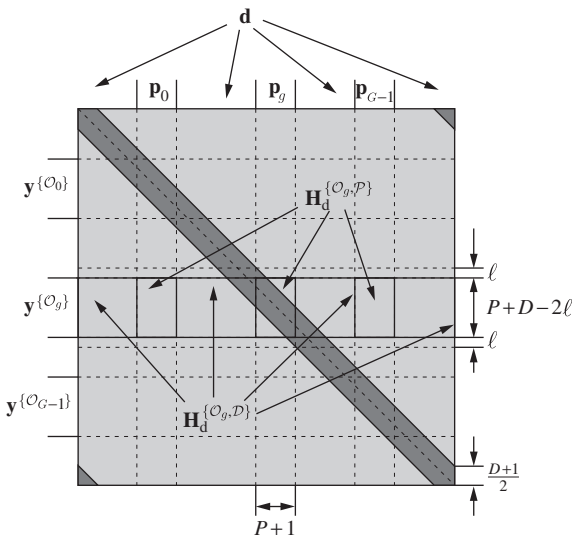

FIGURE 4.2

The partitioning of the frequency-domain channel matrix $\mathbf{H}_{\mathrm{d}}$ into $\mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_g,\mathscr{P}\}}$ and $\mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_g,\mathscr{D}\}}$. The dark shaded areas represent large values, whereas the light shaded areas represent small values.

Using the BEM expression in (4.10), we can now rewrite (4.19) as

$$\mathbf{y}^{\{\mathscr{O}_g\}} = \sum_{i=0}^{I-1} \mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,\mathscr{P}\}} \mathbf{C}_{\mathrm{d},i}^{\{\mathscr{P},\mathscr{P}\}} \mathbf{p} + \underbrace{\sum_{i=0}^{I-1} \mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,\mathscr{D}\}} \mathbf{C}_{\mathrm{d},i}^{\{\mathscr{D},\mathscr{D}\}} \mathbf{d}}_{\mathbf{i}_g} + \mathbf{z}^{\{\mathscr{O}_g\}}, \tag{4.20}$$

where $\mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,\mathscr{P}\}}$ consists of the columns of $\mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,:\}}$ corresponding to the positions of the pilot symbols in $\mathbf{p}$, $\mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,\mathscr{D}\}}$ consists of the columns of $\mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,:\}}$ corresponding to the positions of the unknown data symbols in $\mathbf{d}$, and finally $\mathbf{C}_{\mathrm{d},i}^{\{\mathscr{P},\mathscr{P}\}}$ and $\mathbf{C}_{\mathrm{d},i}^{\{\mathscr{D},\mathscr{D}\}}$ are the submatrices of the diagonal matrix $\mathbf{C}_{\mathrm{d},i}$ consisting of the diagonal elements on the positions $\mathscr{P}$ and $\mathscr{D}$, respectively. Note that $\mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,\mathscr{P}\}}$ and $\mathbf{U}_{\mathrm{c},i}^{\{\mathscr{O}_g,\mathscr{D}\}}$ have the same structure as $\mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_g,\mathscr{P}\}}$ and $\mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_g,\mathscr{D}\}}$ in Fig. 4.2.

Similarly to the time-domain case, the relation of ℓ with the significant Doppler spread D will indicate how much significant interference from the unknown data symbols we will take into account. For instance, if we take $\ell = (D-1)/2$, we take no significant interference into account. Under the assumption that $P \geq D-1$, $\mathbf{y}^{\{\mathscr{O}_g\}}$ then contains the maximum number of received samples that depend on the pilot symbols without any significant interference from the unknown data symbols. However, this generally does not mean that the interference term, i.e., the second term in (4.20), vanishes. This will only be true when a (C)CE-BEM is used and $D = I$, as considered in Kannu & Schniter (2005), but as already mentioned earlier, the (C)CE-BEM is not a very accurate model. Taking ℓ larger than $(D-1)/2$ lowers the complexity as in the time-domain case but might not necessarily degrade the performance. Anyway, the largest value of ℓ is given by $(P+D-1)/2$ (assume $P+D-1$ is even), which is the case where only a single observation sample per pilot cluster is selected. If we take $\ell = 0$, $\mathbf{y}^{\{\mathscr{O}_g\}}$ contains the maximum number of received samples that depend on both the pilot symbols and the significantly interfering unknown data symbols. Note that in the time-invariant case, taking $\ell = (D-1)/2$ or $\ell = 0$ is the same because $D = 1$, which then leads to the traditional time-invariant OFDM channel estimation procedure presented for instance in Negi & Cioffi (1998). Taking ℓ smaller than zero increases the complexity as in the time-domain case, but might change the performance. Hence, we will not put any lower bound on ℓ yet. To conclude, we let ℓ assume an arbitrary integer value, which we bound by $\ell \leq (P+D-1)/2$.

Rewriting (4.20) as a function of the BEM coefficients $\mathbf{c} = (\mathbf{c}_0^T \cdots \mathbf{c}_{I-1}^T)^T$, we can obtain

$$\mathbf{y}^{\{\mathscr{O}_g\}} = \mathbf{A}_g \mathbf{c} + \mathbf{i}_g + \mathbf{z}^{\{\mathscr{O}_g\}}, \tag{4.21}$$

with

$$\mathbf{A}_g \triangleq \left(\mathbf{U}_{\mathrm{c},0}^{\{\mathscr{O}_g,\mathscr{P}\}} \cdots \mathbf{U}_{\mathrm{c},I-1}^{\{\mathscr{O}_g,\mathscr{P}\}}\right) \left(\mathbf{I}_I \otimes \mathrm{diag}\{\mathbf{p}\} \mathbf{V}_K^{\{\mathscr{P},:\}}\right), \tag{4.22}$$

where we have used that $\mathbf{C}_{\mathrm{d},i}^{\{\mathscr{P},\mathscr{P}\}} \mathbf{p} = \mathrm{diag}\{\mathbf{p}\} \sqrt{K} \mathbf{W}_K^{\{\mathscr{P},:\}} (\mathbf{c}_i^T \ \mathbf{0}_{1\times(K-M)})^T = \mathrm{diag}\{\mathbf{p}\} \mathbf{V}_K^{\{\mathscr{P},:\}} \mathbf{c}_i$, with $\mathbf{V}_K$ the first M columns of the DFT matrix $\sqrt{K}\mathbf{W}_K$.

Stacking the results obtained in (4.21) for all G pilot clusters, $\mathbf{y}^{\{\mathscr{O}\}} \triangleq (\mathbf{y}^{\{\mathscr{O}_0\}T} \ \cdots \ \mathbf{y}^{\{\mathscr{O}_{G-1}\}T})^T$, where the set $\mathscr{O}$ is given by $\mathscr{O} \triangleq \{\mathscr{O}_0, \ldots, \mathscr{O}_{G-1}\}$, we obtain

$$\mathbf{y}^{\{\mathscr{O}\}} = \mathbf{A}\mathbf{c} + \mathbf{i} + \mathbf{z}^{\{\mathscr{O}\}}, \tag{4.23}$$

where $\mathbf{i}$ and $\mathbf{z}^{\{\mathscr{O}\}}$ are similarly defined as $\mathbf{y}^{\{\mathscr{O}\}}$, and

$$\mathbf{A} \triangleq (\mathbf{A}_0^T \ \cdots \ \mathbf{A}_{G-1}^T)^T = \mathbf{A}_{\mathrm{c}}\mathbf{A}_{\mathrm{d}}, \tag{4.24}$$

with

$$\begin{aligned}\mathbf{A}_{\mathrm{c}} &\triangleq \begin{pmatrix} \mathbf{U}_{\mathrm{c},0}^{\{\mathscr{O}_0,\mathscr{P}\}} & \cdots & \mathbf{U}_{\mathrm{c},I-1}^{\{\mathscr{O}_0,\mathscr{P}\}} \\ \vdots & & \vdots \\ \mathbf{U}_{\mathrm{c},0}^{\{\mathscr{O}_{G-1},\mathscr{P}\}} & \cdots & \mathbf{U}_{\mathrm{c},I-1}^{\{\mathscr{O}_{G-1},\mathscr{P}\}} \end{pmatrix} \\ &= \mathbf{W}_K^{\{\mathscr{O},:\}} (\mathbf{U}_{\mathrm{d},0} \ \cdots \ \mathbf{U}_{\mathrm{d},I-1}) \left(\mathbf{I}_I \otimes \mathbf{W}_K^{\{\mathscr{P},:\}H} \right),\end{aligned} \tag{4.25}$$

and

$$\mathbf{A}_{\mathrm{d}} \triangleq \mathbf{I}_I \otimes \mathrm{diag}\{\mathbf{p}\}\mathbf{V}_K^{\{\mathscr{P},:\}}. \tag{4.26}$$

Remark 4.1 For the superimposed pilot scheme, the channel estimation data model can be derived in a similar way. However, since the pilot and unknown data symbols overlap in that case, i.e., $\mathbf{a} = \mathbf{p} + \mathbf{d}$, we basically have $G = 1$ and $\mathscr{P}_0 = \mathscr{D}_0 = \mathscr{O}_0 = \mathscr{P} = \mathscr{D} = \mathscr{O} = \{0, \ldots, K-1\}$. As a result, in (4.12) or (4.19) we then obtain $\mathbf{H}_{\mathrm{c}}^{\{\mathscr{O}_0,\mathscr{P}_0\}} = \mathbf{H}_{\mathrm{c}}^{\{\mathscr{O}_0,\mathscr{D}_0\}} = \mathbf{H}_{\mathrm{c}}$ or $\mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_0,\mathscr{P}\}} = \mathbf{H}_{\mathrm{d}}^{\{\mathscr{O}_0,\mathscr{D}\}} = \mathbf{H}_{\mathrm{d}}$, respectively. In spite of these differences, we can still show that the data model for the superimposed pilot scheme takes on a form that is similar to (4.16) or (4.23).

4.3.3 Channel Estimators

As is clear from the previous section, whether we work in the time or frequency domain, whether multiplexed or superimposed pilot symbols are used, the data model for channel estimation can always be expressed as a mixture of a useful part resulting from the pilot symbols, an interference part resulting from the interfering unknown data symbols, and a noise part:

$$\mathbf{y}^{\{\mathscr{O}\}} = \mathbf{A}\mathbf{c} + \mathbf{i}(\mathbf{c}) + \mathbf{z}^{\{\mathscr{O}\}}, \tag{4.27}$$

where we have made the dependence of the interference term $\mathbf{i}$ on the unknown BEM coefficients $\mathbf{c}$ explicit. Without loss of generality, we will present channel estimators based on the generic data model in (4.27). However, the inherent difference between the single-carrier channel estimators and the OFDM channel estimators will be pointed out from time to time.

We will focus on three channel estimators in this section: the linear minimum mean-squared error (LMMSE) estimator, the least squares (LS) estimator, and the best linear unbiased estimator (BLUE). All these estimators estimate the BEM coefficients $\mathbf{c}$ as

$$\hat{\mathbf{c}} = \mathbf{F}\mathbf{y}^{\{\mathscr{O}\}},$$

where $\mathbf{F}$ is the respective linear estimator. Whereas the LMMSE estimator relies on the statistics of $\mathbf{c}$, the LS estimator and the BLUE treat $\mathbf{c}$ as a deterministic variable. The unknown data symbols $\mathbf{d}$ and

the noise $\mathbf{z}$ are always assumed to be stochastic and mutually uncorrelated, i.e., $\mathsf{E}\{\mathbf{d}\mathbf{z}^H\} = \mathbf{0}$. For every estimator, we can then estimate the BEM channel vector $(\mathbf{U} \otimes \mathbf{I}_M)\mathbf{c}$ in (4.8) as $(\mathbf{U} \otimes \mathbf{I}_M)\hat{\mathbf{c}}$, and we can compute the related mean-squared error (MSE) as

$$\begin{aligned}\mathrm{MSE} &\triangleq \mathsf{E}_{\mathbf{c},\mathbf{d},\mathbf{z}^{\{\mathscr{O}\}}}\left\{\|(\mathbf{U}\otimes\mathbf{I}_M)\hat{\mathbf{c}} - (\mathbf{U}\otimes\mathbf{I}_M)\mathbf{c}\|^2\right\} \\ &= \mathrm{tr}\left\{(\mathbf{U}\otimes\mathbf{I}_M)\mathsf{E}_{\mathbf{c},\mathbf{d},\mathbf{z}^{\{\mathscr{O}\}}}\{(\hat{\mathbf{c}}-\mathbf{c})(\hat{\mathbf{c}}-\mathbf{c})^H\}(\mathbf{U}\otimes\mathbf{I}_M)^H\right\} \\ &= \mathrm{tr}\left\{\mathsf{E}_{\mathbf{c},\mathbf{d},\mathbf{z}^{\{\mathscr{O}\}}}\left\{(\hat{\mathbf{c}}-\mathbf{c})(\hat{\mathbf{c}}-\mathbf{c})^H\right\}\right\}, \end{aligned} \tag{4.28}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a matrix and the last equality is due to the fact that we have designed $\mathbf{U}$ with orthonormal columns (cf. (4.7)). Note that this MSE does not include the modeling error of the BEM.

The performance of each channel estimator is sensitive to the design parameter ℓ. Taking a larger ℓ, we basically take less interference from unknown data symbols into account, which could improve our estimate. But on the other hand, increasing ℓ reduces the number of observation symbols and makes $\mathbf{A}$ a fatter matrix, which can be detrimental to the channel estimator.

4.3.3.1 The LMMSE Estimator

The LMMSE estimator treats $\mathbf{c}$ as a stochastic variable. We further assume it is uncorrelated with the unknown data symbols and the noise, i.e., $\mathsf{E}\{\mathbf{c}\mathbf{z}^H\} = \mathbf{0}$ and $\mathsf{E}\{\mathbf{c}\mathbf{d}^H\} = \mathbf{0}$. The LMMSE estimator is the linear filter $\mathbf{F}$ that minimizes the MSE between $\mathbf{c}$ and $\hat{\mathbf{c}}$:

$$\mathbf{F}_{\mathrm{LMMSE}} \triangleq \arg\min_{\mathbf{F}} \mathsf{E}_{\mathbf{c},\mathbf{d},\mathbf{z}}\left\{\|\mathbf{F}\mathbf{y}^{\{\mathscr{O}\}} - \mathbf{c}\|^2\right\}.$$

Because the BEM coefficients $\mathbf{c}$, the unknown data symbols $\mathbf{d}$, and the noise $\mathbf{z}$ are all mutually uncorrelated, the LMMSE estimator can be derived as (Tang, Cannizzaro, Banelli, & Leus, 2007)

$$\mathbf{F}_{\mathrm{LMMSE}} = \mathbf{R}_c\mathbf{A}^H(\mathbf{A}\mathbf{R}_c\mathbf{A}^H + \mathbf{R}_i + \mathbf{R}_z^{\{\mathscr{O}\}})^{-1}, \tag{4.29}$$

where the correlation matrices $\mathbf{R}_c$, $\mathbf{R}_i$, and $\mathbf{R}_z^{\{\mathscr{O}\}}$ are given by $\mathbf{R}_c \triangleq \mathsf{E}_{\mathbf{c}}\{\mathbf{c}\mathbf{c}^H\}$, $\mathbf{R}_i \triangleq \mathsf{E}_{\mathbf{c},\mathbf{d}}\{\mathbf{i}\mathbf{i}^H\}$, and $\mathbf{R}_z^{\{\mathscr{O}\}} \triangleq \mathsf{E}_{\mathbf{z}}\{\mathbf{z}^{\{\mathscr{O}\}}\mathbf{z}^{\{\mathscr{O}\}H}\}$. The computation of these correlation matrices in the time and frequency domains is given in Tang (2007).

Note that although (4.29) appears similar to the classical LMMSE estimator (Kay, 1993), it has the extra task to process the interference term $\mathbf{i}$, which depends on $\mathbf{c}$ itself. For this purpose, the proposed LMMSE estimator treats $\mathbf{i}$ as a random vector assuming the statistics of $\mathbf{c}$ and $\mathbf{d}$. A drawback of this approach is that the statistics of $\mathbf{c}$ are difficult to determine and not always reliable in practice. For instance, the Doppler spread could only be roughly known or the assumed Doppler spectrum could deviate from the actual shape. In such cases, the proposed LMMSE estimator will function suboptimally.

The MSE obtained with the LMMSE estimator (cf. (4.28)) can be expressed as

$$\mathrm{MSE}_{\mathrm{LMMSE}} = \mathrm{tr}\left\{(\mathbf{A}^H(\mathbf{R}_i + \mathbf{R}_z^{\{\mathscr{O}\}})^{-1}\mathbf{A} + \mathbf{R}_c^{-1})^{-1}\right\}. \tag{4.30}$$

4.3.3.2 The Least-Squares Estimator

In contrast to the LMMSE estimator, the LS estimator treats $\mathbf{c}$ as a deterministic variable. It is the linear filter $\mathbf{F}$ that minimizes the squared error between $\mathbf{y}^{\{\mathscr{O}\}}$ and $\mathbf{A}\hat{\mathbf{c}}$:

$$\mathbf{F}_{\mathrm{LS}} \triangleq \arg\min_{\mathbf{F}} \|\mathbf{y}^{\{\mathscr{O}\}} - \mathbf{A}\mathbf{F}\mathbf{y}^{\{\mathscr{O}\}}\|^2.$$

The solution is well known (Kay, 1993) and given by the pseudo-inverse of $\mathbf{A}$, i.e.,

$$\mathbf{F}_{\mathrm{LS}} = \mathbf{A}^{\#}.$$

The LS estimator is the most robust estimator, in the sense that it requires no knowledge about the statistics of the BEM coefficients and the noise. This eliminates the risk of a mismatched knowledge. However, the LS estimator will perform poorly when the interference and/or the noise are prominent. More fundamentally, the conditioning of the matrix $\mathbf{A}$ plays a crucial role in the channel MSE of the LS estimator. Using (4.28), this channel MSE can be computed as

$$\begin{aligned}\mathrm{MSE}_{\mathrm{LS}} &= \mathsf{E}_{\mathbf{c},\mathbf{d},\mathbf{z}}\left\{\mathrm{tr}\left\{\mathbf{A}^{\#}(\mathbf{i}+\mathbf{z}^{\{\mathscr{O}\}})(\mathbf{i}+\mathbf{z}^{\{\mathscr{O}\}})^H\mathbf{A}^{\#H}\right\}\right\}\\ &= \mathrm{tr}\left\{\mathbf{A}^{\#}\mathsf{E}_{\mathbf{c},\mathbf{d},\mathbf{z}}\left\{(\mathbf{i}+\mathbf{z}^{\{\mathscr{O}\}})(\mathbf{i}+\mathbf{z}^{\{\mathscr{O}\}})^H\right\}\mathbf{A}^{\#H}\right\}\\ &= \mathrm{tr}\left\{\mathbf{A}^{\#}(\mathbf{R}_i+\mathbf{R}_z^{\{\mathscr{O}\}})\mathbf{A}^{\#H}\right\}. \end{aligned} \tag{4.31}$$

4.3.3.3 An Iterative BLUE

The BLUE yields a compromise between the LMMSE and LS estimators: it treats $\mathbf{c}$ as a deterministic variable, thus avoiding a possible error in calculating its statistics; at the same time, it leverages the statistics of the interference and the noise, such that they can be better suppressed than by the LS estimator. Simulation results show that the BLUE is able to yield a performance close to that of the LMMSE estimator, even if the latter uses perfect statistical knowledge.

The BLUE is the linear filter $\mathbf{F}$ that minimizes the MSE between $\mathbf{c}$ and $\hat{\mathbf{c}}$ subject to the condition that $\mathbf{c}$ is unbiased:

$$\mathbf{F}_{\mathrm{BLUE}} \triangleq \arg\min_{\mathbf{F}} \mathsf{E}_{\mathbf{d},\mathbf{z}}\left\{\|\mathbf{F}\mathbf{y}^{\{\mathscr{O}\}} - \mathbf{c}\|^2\right\}, \quad \text{s.t. } \mathsf{E}_{\mathbf{d},\mathbf{z}}\left\{\mathbf{F}\mathbf{y}^{\{\mathscr{O}\}}\right\} = \mathbf{c}.$$

The solution can be computed as in Kay (1993) by combining the interference $\mathbf{i}$ and the noise $\mathbf{z}^{\{\mathscr{O}\}}$ to a single disturbance term:

$$\mathbf{F}_{\mathrm{BLUE}} = \left(\mathbf{A}^H\tilde{\mathbf{R}}_{\delta}^{-1}(\mathbf{c})\mathbf{A}\right)^{-1}\mathbf{A}^H\tilde{\mathbf{R}}_{\delta}^{-1}(\mathbf{c}), \tag{4.32}$$

where $\tilde{\mathbf{R}}_{\delta}(\mathbf{c})$ denotes the correlation matrix of the disturbance with $\mathbf{c}$ considered as a deterministic variable:

$$\tilde{\mathbf{R}}_{\delta}(\mathbf{c}) \triangleq \tilde{\mathbf{R}}_i(\mathbf{c}) + \mathbf{R}_z^{\{\mathscr{O}\}},$$

with $\tilde{\mathbf{R}}_i(\mathbf{c}) \triangleq \mathsf{E}_{\mathbf{d}}\{\mathbf{i}\mathbf{i}^H\}$. Note here that we do not take an average over $\mathbf{c}$ as in the LMMSE case, which is also the reason why we use a different notation. The calculation of $\tilde{\mathbf{R}}_{\delta}(\mathbf{c})$ in both the time and frequency domains can be found in Tang (2007).

The problem with (4.32) is that it cannot be calculated in closed form since due to the presence of $\tilde{\mathbf{R}}_{\delta}(\mathbf{c})$, its calculation requires the knowledge of $\mathbf{c}$ itself. As a remedy, we apply a recursive approach. Suppose that at the nth iteration, an estimate of $\mathbf{c}$ has been obtained, which is denoted as $\hat{\mathbf{c}}_{\mathrm{BLUE}}^{(n)}$. Then

we utilize this temporary estimate to update the correlation matrix $\tilde{\mathbf{R}}_\delta(\mathbf{c})$, which in turn is used to produce the BLUE for the subsequent iteration, etc.:

$$\mathbf{F}_{\text{BLUE}}^{(n+1)} = \left(\mathbf{A}^H \tilde{\mathbf{R}}_\delta^{-1}(\hat{\mathbf{c}}_{\text{BLUE}}^{(n)})\mathbf{A}\right)^{-1}\mathbf{A}^H \tilde{\mathbf{R}}_\delta^{-1}(\hat{\mathbf{c}}_{\text{BLUE}}^{(n)}),$$
$$\hat{\mathbf{c}}_{\text{BLUE}}^{(n+1)} = \mathbf{F}_{\text{BLUE}}^{(n+1)}\mathbf{y}^{\{\mathscr{O}\}}.$$

Note that a similar idea is adopted in Ghogho & Swami (2004) though in a different context. To initialize the iteration, we can set $\hat{\mathbf{c}}_{\text{BLUE}}^{(0)} = \mathbf{0}$, which results in the following expression for the first iteration:

$$\mathbf{F}_{\text{BLUE}}^{(1)} = \left(\mathbf{A}^H \mathbf{R}_z^{\{\mathscr{O}\}-1}\mathbf{A}\right)^{-1}\mathbf{A}^H \mathbf{R}_z^{\{\mathscr{O}\}-1},$$
$$\hat{\mathbf{c}}_{\text{BLUE}}^{(1)} = \mathbf{F}_{\text{BLUE}}^{(1)}\mathbf{y}^{\{\mathscr{O}\}}.$$

The above expression actually corresponds to the LS estimator but weighted with the noise covariance matrix. Under the Gaussian noise assumption, this estimator is the maximum likelihood estimator (Kay, 1993) that is obtained by ignoring the interference $\mathbf{i}$.

Assuming that $\hat{\mathbf{c}}_{\text{BLUE}}^{(n)}$ will approach the theoretical BLUE $\hat{\mathbf{c}}_{\text{BLUE}}$ in (4.32), we can use (4.28) to compute an estimate of the channel MSE obtained by the iterative BLUE as

$$\begin{aligned}\text{MSE}_{\text{BLUE}} &= \mathsf{E}_{\mathbf{c},\mathbf{d},\mathbf{z}^{\{\mathscr{O}\}}}\left\{\text{tr}\left\{\mathbf{F}_{\text{BLUE}}\left(\mathbf{i}+\mathbf{z}^{\{\mathscr{O}\}}\right)\left(\mathbf{i}+\mathbf{z}^{\{\mathscr{O}\}}\right)^H \mathbf{F}_{\text{BLUE}}^H\right\}\right\}\\ &= \mathsf{E}_{\mathbf{c}}\left\{\text{tr}\left\{\mathbf{F}_{\text{BLUE}}\tilde{\mathbf{R}}_\delta(\mathbf{c})\mathbf{F}_{\text{BLUE}}^H\right\}\right\}\\ &= \mathsf{E}_{\mathbf{c}}\left\{\text{tr}\left\{(\mathbf{A}^H\tilde{\mathbf{R}}_\delta^{-1}(\mathbf{c})\mathbf{A})^{-1}\right\}\right\}. \end{aligned} \tag{4.33}$$

It must be remarked that the channel MSE of the BLUE is difficult to evaluate in closed form because the parameter $\mathbf{c}$ which is contained in $\tilde{\mathbf{R}}_\delta(\mathbf{c})^{-1}$ has to be averaged out. This forces us to resort to the Monte Carlo method to evaluate (4.33).

4.3.3.4 Optimization of ℓ

It is clear that the channel MSE of every estimator depends on ℓ (cf. (4.30), (4.31), and (4.33)). In this section, we will seek the optimal ℓ, or equivalently, the optimal number of observation samples used for channel estimation. With optimal ℓ we mean the ℓ that minimizes the channel MSE given in (4.30), (4.31), and (4.33) for the LMMSE estimator, the LS estimator, and the BLUE, respectively. A closed-form solution is, however, hard to obtain, especially for the BLUE. Alternatively, we can resort to an exhaustive search approach, which is feasible because the range of possible values for ℓ is limited. To be more specific, we have seen in Section 4.3.2 that the limits for ℓ are basically given by $\max\{0, M-\Delta-1\} \le \ell \le (P+M-1)/2$ and $\ell \le (P+D-1)/2$ for the single-carrier and OFDM system, respectively. Here, M stands for the discrete delay spread, D for the significant discrete Doppler spread, and Δ for the minimal number of samples in between two nonzero pilot symbols from different pilot clusters. In addition to these bounds, there also exist some other bounds for ℓ. First of all, we desire the matrix $\mathbf{A}$ to be of full column-rank, which is essential for the channel estimators to have a good performance in the absence of interference and noise. This means that $\ell \le (P+M)/2 - MI/(2G)$ and $\ell \le (P+D)/2 - MI/(2G)$ for the single-carrier and OFDM system, respectively. Second, we do not want the observation samples related to the nonzero pilot symbols from different pilot clusters

to overlap. With δ denoting the minimal number of samples in between two pilot clusters (note that $\delta \leq \Delta$), this means that $\ell \geq (M-\delta-1)/2$ and $\ell \geq (D-\delta-1)/2$ for the single-carrier and OFDM system, respectively.

Fortunately, even the exhaustive search may be avoided as will become evident from the simulations, where the MSE-versus-ℓ curves for each channel estimator exhibit a monotonous track. In particular, the LS estimator yields the best performance when ℓ is maximized, whereas the LMMSE estimator and the BLUE perform best when ℓ is minimized. This can be explained as follows. As we already indicated before, there are two opposite effects. On one hand, increasing ℓ reduces the amount of interference from the unknown data symbols and hence improves the channel estimate. On the other hand, increasing ℓ makes $\mathbf{A}$ a fatter matrix and deteriorates the channel estimate. For the LS estimator, the first effect plays the most important role because the LS estimator is not good at suppressing this interference due to a lack of statistical knowledge. However, this interference poses no serious problems for the LMMSE estimator and the BLUE, since both of them can take the interference into account. Hence, for these estimators the second effect is stronger and thus ℓ should be as small as possible.

4.3.4 Channel Identifiability

In this chapter, we define channel identifiability in terms of the rank condition of the pilot-related matrix $\mathbf{A}$. It is obvious that for the LS estimator and the BLUE, $\mathbf{A}$ should be of full column-rank, i.e., $\text{rank}\{\mathbf{A}\} = MI$, where MI is the total number of BEM coefficients to be estimated. If $\mathbf{A}$ is not full column-rank, the channel cannot be correctly recovered even in an interference- and noise-free situation. Although not directly visible, the full column-rank condition for $\mathbf{A}$ is also significant for the performance of the LMMSE estimator. In this section, we will describe several sufficient conditions to obtain a full column-rank $\mathbf{A}$.

4.3.4.1 *Channel Identifiability for Single-Carrier System*

To discuss channel identifiability for the single-carrier system, we will adopt the following assumption on the pilot structure:

Assumption 4.1 The pilot structure satisfies the following conditions:

(C1) The length of each pilot cluster $P+1$ satisfies $P+1 \geq 2\ell+1$.
(C2) Inside each pilot cluster, either the first or the last ℓ pilots are zero.
(C3) Every pilot cluster has at least one nonzero pilot in between the first and last ℓ pilots.
(C4) The nonzero pilot symbols from different pilot clusters should be separated by at least $M-1$ symbols.

The structure of a pilot cluster satisfying conditions (C1), (C2), and (C3) is depicted in Fig. 4.3. Condition (C4) ensures that the observation samples related to a pilot cluster are not influenced by nonzero pilot symbols from another pilot cluster. Actually, condition (C4) corresponds to the earlier condition $\Delta \geq M-\ell-1$ required for (4.12) to hold.

We can now state the following theorem:

Theorem 4.1 Under Assumption 4.1, the channel is identifiable if the number of pilot clusters G is greater than or equal to the BEM length, i.e.,

$$G \geq I.$$

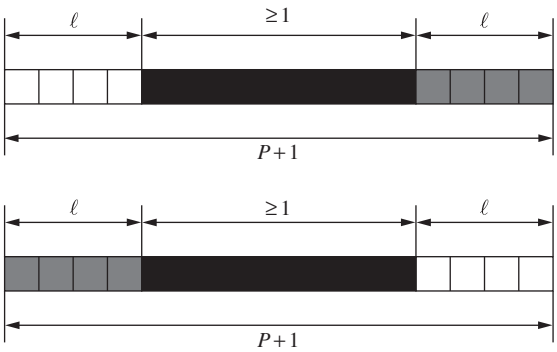

FIGURE 4.3

The two possible cases of the proposed pilot structure: the white boxes stand for the area where zero pilot symbols are located; the black boxes for the area where at least one nonzero pilot symbol is located; and the gray boxes for the area where arbitrary pilot symbols (zero or nonzero) are located.

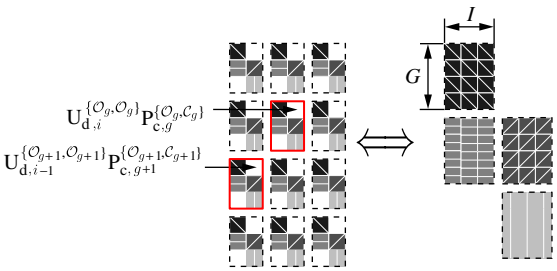

FIGURE 4.4

Changing a stack of banded matrices to a block banded matrix by row- and column-interleaving.

We only give a sketch of the proof of Theorem 4.1; a full detailed proof can be found in Tang (2007). Under (C1), (C2), (C3), and (C4), the $(P+M-2\ell)\times M$ matrix $\mathbf{P}_{\mathrm{c},g}^{\{\mathcal{O}_g,\mathcal{C}_g\}}$ used in (4.17) is a tall banded matrix with zeros above (below) the diagonal starting (ending) at the top left (bottom right) corner. From (4.17), the matrix $\mathbf{A}$ is a stack of G by I of such tall banded matrices, each one multiplied with a diagonal matrix, which does not change the tall banded structure. This is illustrated in the left part of Fig. 4.4. Let us now change the order of the rows and columns of $\mathbf{A}$ as illustrated in the right part of Fig. 4.4. Such a permutation does not affect the column-rank condition, and it enables us to obtain a new matrix that has a block-wise banded structure with blocks of size $G\times I$. It is then easy to prove that all the $G\times I$ blocks on the first (last) nonzero block diagonal of the permuted matrix have full column-rank, and this for all BEMs that have been considered in literature. This means that $\mathbf{A}$ has full column-rank.

Some remarks regarding Theorem 4.1 are now in order.

Remark 4.2 Assumption 4.1 is only sufficient for a full column-rank $\mathbf{A}$. However, it ensures that Theorem 4.1 holds for an arbitrary BEM and pilot choice. When $P+1<2\ell+1$, nonzero entries will have to emerge in the

upper and lower diagonals of $\mathbf{P}_{c,g}^{\{\mathscr{O}_g,\mathscr{C}_g\}}$, and thus also in the upper and lower block diagonals of the permuted version of $\mathbf{A}$ (see Fig. 4.4). As a result, the proof does not hold anymore and the full column-rank condition of $\mathbf{A}$ cannot be guaranteed. It will depend on the choice of the BEM and the pilot symbols.

Remark 4.3 Ma, Giannakis, & Ohno (2003) and Kannu & Schniter (2005), propose a special pilot structure where the G pilot clusters are equidistant and every cluster consists of a nonzero pilot symbol surrounded by $M-1$ zeros to the left and right, i.e., $P+1=2M-1$. We borrow the term "time-domain Kronecker delta (TDKD) structure" from Kannu & Schniter (2005) to describe this special pilot structure. The TDKD structure has been proven to be optimal in the MSE and capacity sense provided the (C)CE-BEM is used for channel modeling (Ma, Giannakis, & Ohno, 2003; Kannu & Schniter, 2005). The TDKD structure is a special case of the pilot structure defined by Assumption 4.1 with $P+1=2M-1$ and $\ell=M-1$. In that case, we know from Theorem 4.1 that the channel is identifiable if $G \geq I$, as also argued in Ma, Giannakis, & Ohno (2003) and Kannu & Schniter (2005) for the (C)CE-BEM.

4.3.4.2 Channel Identifiability for OFDM System

For OFDM systems, the full column-rank condition of $\mathbf{A}$ is much harder to analyze, due to the fact that $\mathbf{H}_d$ is basically a full matrix. To simplify the analysis, we will adopt a specific pilot scheme here.

Assumption 4.2 The pilot structure satisfies the following conditions:

(C1) The length of each pilot cluster $P+1$ satisfies $P+1 \geq 1$.
(C2) Every pilot cluster has one nonzero pilot.
(C3) The nonzero pilots are equispaced.

Remark 4.4 In Kannu & Schniter (2005), the counterpart of the TDKD structure in the frequency domain is proposed, which is labeled the "frequency domain Kronecker delta (FDKD) structure." In the FDKD structure, the G pilot clusters are equidistant and every cluster consists of a nonzero pilot symbol surrounded by $D-1$ zeros to the left and right, i.e., $P+1=2D-1$, where D is the significant discrete Doppler spread. The FDKD structure is optimal in terms of the MSE if the (C)CE-BEM is used for channel modeling (Kannu & Schniter, 2005). The FDKD structure fits into Assumption 4.2 with $P+1=2D-1$.

Under Assumption 4.2, the full column-rank condition for the corresponding matrix $\mathbf{A}$ will be independent of the choice of the BEM and the pilot symbols, as we will show next. According to Assumption 4.2, the gth pilot cluster $\mathbf{p}_g$ can be expressed as

$$\mathbf{p}_g = (\mathbf{0}_{1\times L}\, p_g\, \mathbf{0}_{1\times R})^T,$$

where p_g stands for the nonzero pilot symbol, and L and R are the number of zeros left and right of this nonzero pilot. The total pilot vector $\mathbf{p}$ can therefore be written as

$$\mathbf{p} = \overline{\mathbf{p}} \otimes (\mathbf{0}_{1\times L}\, 1\, \mathbf{0}_{1\times R})^T,$$

where $\overline{\mathbf{p}} \triangleq (p_0\ \ldots\ p_{G-1})^T$ contains all the G nonzero pilot symbols. The positions of these nonzero pilot symbols are collected in

$$\overline{\mathscr{P}} \triangleq \{\mu, \mu+K/G, \ldots, \mu+K(G-1)/G\}, \tag{4.34}$$

with μ denoting the position of the first nonzero pilot symbol.

With the help of the notations $\overline{\mathbf{p}}$ and $\overline{\mathscr{P}}$, and taking the zero pilot symbols into account, we can now rewrite the matrix $\mathbf{A}$ in (4.24) as

$$\mathbf{A} = \overline{\mathbf{A}}_c \overline{\mathbf{A}}_d,$$

with

$$\overline{\mathbf{A}}_{\mathrm{c}} \triangleq \mathbf{W}_K^{\{\mathscr{O},:\}}(\mathbf{U}_{\mathrm{d},0} \ \ldots \ \mathbf{U}_{\mathrm{d},I-1})\left(\mathbf{I}_I \otimes \mathbf{W}_K^{\{\overline{\mathscr{P}},:\}H}\right), \tag{4.35}$$

$$\overline{\mathbf{A}}_{\mathrm{d}} \triangleq \mathbf{I}_I \otimes \mathrm{diag}\{\overline{\mathbf{p}}\}\mathbf{V}_K^{\{\overline{\mathscr{P}},:\}}. \tag{4.36}$$

In the above equations, the columns of $\mathbf{A}_{\mathrm{c}}$ and the rows of $\mathbf{A}_{\mathrm{d}}$ corresponding to the positions of the zero pilot symbols have been discarded in $\overline{\mathbf{A}}_{\mathrm{c}}$ and $\overline{\mathbf{A}}_{\mathrm{d}}$, respectively, which obviously does not affect the result.

Let us now examine the matrix $\overline{\mathbf{A}}_{\mathrm{d}}$. First of all, $\mathbf{V}_K^{\{\overline{\mathscr{P}},:\}}$ consists of the first M columns and G equidistant rows of the DFT matrix $\sqrt{K}\mathbf{W}_K$. Hence, $\mathbf{V}_K^{\{\overline{\mathscr{P}},:\}}$ is a $G \times M$ Vandermonde matrix with rank equal to $\min\{G,M\}$. Furthermore, $\mathrm{diag}\{\overline{\mathbf{p}}\}$ is a diagonal matrix with the nonzero pilot symbols on its diagonal and as such will not change the rank of a matrix if multiplied with it. Hence, we can state the following lemma.

Lemma 4.1 Under Assumption 4.2, the $GI \times MI$ matrix $\overline{\mathbf{A}}_{\mathrm{d}}$ depends on the pilot symbols but not on the BEM and its rank is given by

$$\mathrm{rank}\{\overline{\mathbf{A}}_{\mathrm{d}}\} = \min\{G,M\}I. \tag{4.37}$$

As a result, if $G \geq M$, the rank of $\overline{\mathbf{A}}_{\mathrm{d}}$ is equal to MI.

To examine the matrix $\overline{\mathbf{A}}_{\mathrm{c}}$, let us introduce an additional assumption.

Assumption 4.3 All the received samples will be used for channel estimation, i.e., $\mathscr{O} = \{0,1,\ldots,K-1\}$.

This means that $\mathbf{W}_K^{\{\mathscr{O},:\}} = \mathbf{W}_K$, which can thus be omitted when analyzing the rank of $\overline{\mathbf{A}}_{\mathrm{c}}$. Due to (4.34), the matrix $\mathbf{W}_K^{\{\overline{\mathscr{P}},:\}H}$ can be written as

$$\mathbf{W}_K^{\{\overline{\mathscr{P}},:\}H} = \sqrt{G/K}(1 \ \mathrm{e}^{\sqrt{-1}2\pi\mu G/K} \ \mathrm{e}^{\sqrt{-1}2\pi\mu 2G/K} \ \cdots \ \mathrm{e}^{\sqrt{-1}2\pi\mu(K/G-1)G/K})^T \tag{4.38}$$

$$\otimes \mathrm{diag}\{(1 \ \mathrm{e}^{\sqrt{-1}2\pi\mu/K} \ \mathrm{e}^{\sqrt{-1}2\pi\mu 2/K} \ \cdots \ \mathrm{e}^{\sqrt{-1}2\pi\mu(G-1)/K})^T\}\mathbf{W}_G^H. \tag{4.39}$$

It is easy to show that this Kronecker structure of $\mathbf{W}_K^{\{\overline{\mathscr{P}},:\}H}$ relates the rank of $\overline{\mathbf{A}}_{\mathrm{c}}$ to the rank of

$$\begin{pmatrix} \mathbf{U}_{\mathrm{d},0,0} & \cdots & \mathbf{U}_{\mathrm{d},I-1,0} \\ \vdots & & \vdots \\ \mathbf{U}_{\mathrm{d},0,K/G-1} & \cdots & \mathbf{U}_{\mathrm{d},I-1,K/G-1} \end{pmatrix}, \tag{4.40}$$

where the $G \times G$ matrix $\mathbf{U}_{\mathrm{d},i,k}$ is the diagonal submatrix of $\mathbf{U}_{\mathrm{d},i}$ with rows and columns going from kG to $(k+1)G-1$. Permutating the rows and columns, the above matrix can be transformed into

$$\begin{pmatrix} \mathbf{U}_0 & & \\ & \ddots & \\ & & \mathbf{U}_{G-1} \end{pmatrix}, \tag{4.41}$$

where the $K/G \times I$ matrix $\mathbf{U}_g$ selects a set of K/G equidistant rows from $\mathbf{U}$ starting at g. For all BEMs that have been considered in the literature, the matrix $\mathbf{U}_g$ has a rank equal to $\min\{K/G, I\}$. So we can formulate the following lemma.

Lemma 4.2 Under Assumptions 4.2 and 4.3, the $K \times GI$ matrix $\overline{\mathbf{A}}_c$ depends on the BEM but not on the pilot symbols and its rank is given by

$$\text{rank}\{\overline{\mathbf{A}}_c\} = \min\{K/G, I\}G. \tag{4.42}$$

Hence, if $K/G \geq I$, the rank of $\overline{\mathbf{A}}_c$ is equal to GI.

For an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$, the rank inequality reads (Horn & Johnson, 1999)

$$\text{rank}\{\mathbf{A}\} + \text{rank}\{\mathbf{B}\} - k \leq \text{rank}\{\mathbf{AB}\} \leq \min\{\text{rank}\{\mathbf{A}\}, \text{rank}\{\mathbf{B}\}\}. \tag{4.43}$$

Applying this rank inequality to (4.37) and (4.42), we obtain the following theorem.

Theorem 4.2 Under Assumptions 4.2 and 4.3, the channel is identifiable if the number of pilot clusters G satisfies

$$M \leq G \leq K/I. \tag{4.44}$$

A few remarks are in order at this stage.

Remark 4.5 One can observe the link between Theorem 4.2 and the channel identifiability for a time-invariant channel. The latter can be viewed as a special case of the time-varying channel with $I = 1$. As per (4.44), in this case, we only require that the total number of pilot symbols should be larger than or equal to the channel length, which is consistent with the channel identifiability condition for time-invariant channels as given in Negi & Cioffi (1998).

Remark 4.6 Theorem 4.2 is a sufficient condition for an arbitrary choice of the nonzero pilot symbols. To understand this, it is interesting to study an extreme case where the whole block is occupied by nonzero pilot symbols. We can view such a pilot structure as a special case of Assumption 4.2 with $P = 0$ and $G = K$. In this case, the second inequality in (4.44) can only be satisfied if $I = 1$. Because $I > 1$ for time-varying channels, the column-rank of the matrix $\mathbf{A}$ will therefore not necessarily be full, and will depend on the specific values of the pilot symbols.

4.3.5 Simulation Results

In this section, we will show some simulation results illustrating the methods proposed in the previous sections. We only consider time-varying channels that follow Jakes' Doppler spectrum (Jakes, 1974). The methods given in Zheng & Xiao (2003) are used to generate these channels. Defining the normalized Doppler spread as $\xi_{\max} = \nu T_s$, where ν is the Doppler spread and T_s is the sampling period, we will concentrate on two normalized Doppler spreads: (1) $\xi_{\max} = 0.0008$ and (2) $\xi_{\max} = 0.004$.

4.3.5.1 *Results for the OFDM System*

For the OFDM system, we set the BEM length to $I = 3$. Further, we assume the channel to be an FIR filter with $M = 4$ taps, which are independent Gaussian random variables with zero mean and a uniform power delay profile, i.e., $\mathsf{E}\{|h[n,m]|^2\} = 1/M$. In summary, we can model the time-varying channel with $MI = 12$ BEM coefficients.

To examine the proposed channel estimation algorithms, we consider an OFDM system with $K = 64$ subcarriers. We adopt an FDKD structure with $G = 8$ equidistant pilot clusters, each containing

$P+1=3$ pilot symbols. One nonzero pilot symbol is positioned in the middle of the cluster with one zero pilot on both sides, i.e., the significant discrete Doppler spread D is assumed to be $D=2$. Note that roughly 37.5% of the subcarriers are devoted to training.

For the following simulations, we will use the DKL-BEM when constructing the LMMSE estimator because both of them rely on knowledge of the channel statistics. However, for both the DKL-BEM and the LMMSE estimator, we will allow for a mismatch of the statistics by assuming a fixed $\xi_{\max}=0.002$ for all the Doppler spreads under test. For the LS estimator and the BLUE, we will consider the (O)CE-BEM because both 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 Kannu & Schniter (2005). 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 frequency-domain channel matrix is viewed as strictly circularly banded.

We first study the effect of the parameter ℓ on the channel MSE of the various channel estimators. Note that we only consider the channel MSE, omitting the BEM modeling error at this point. We have mentioned in the previous section that ℓ is upper-bounded by the minimum of $(P+D-1)/2=1.5$ and $(P+D)/2-MI/(2G)=1.25$, and lower-bounded by $(D-\delta-1)/2=-2$, because $\delta=5$. For values of ℓ in this range, we evaluate the channel MSE of the LMMSE estimator (4.30) for SNR $=10$, 20, 30, and 40 dB, and depict the results in Fig. 4.5. We observe that the effect of ℓ is very limited, but one can observe a slight decrease in channel MSE if ℓ decreases. The reason why we only see a limited effect is because of the mismatch in Doppler spread. If the exact Doppler spread was known, the influence of ℓ would be more pronounced. In any case, we choose $\ell=-2$ as the optimal value, which implies that the whole OFDM symbol will be invoked for channel estimation.

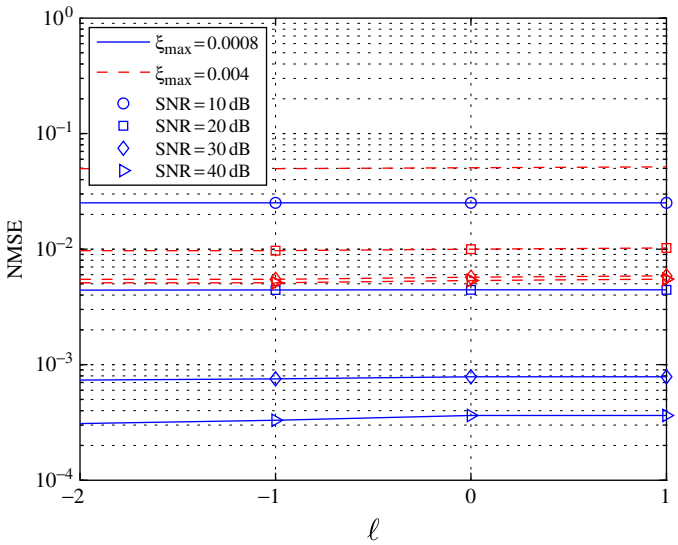

FIGURE 4.5

MSE versus ℓ for the LMMSE estimator in the OFDM case.

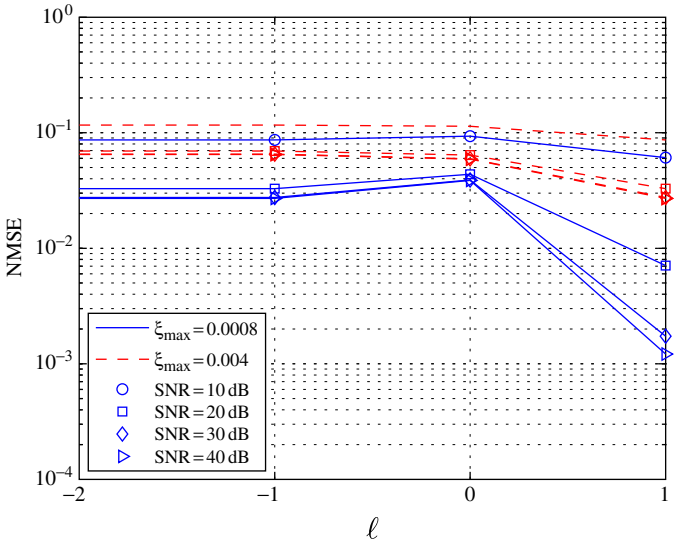

FIGURE 4.6

MSE versus ℓ for the LS estimator in the OFDM case.

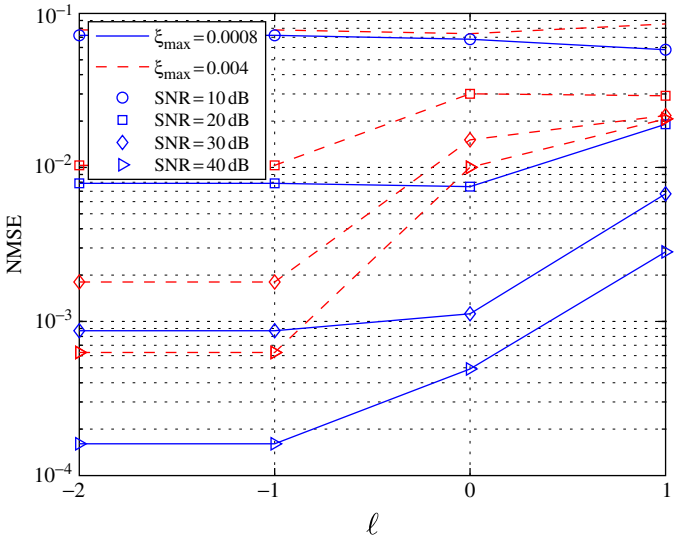

FIGURE 4.7

MSE versus ℓ for the BLUE in the OFDM case.

The results for the LS estimator are plotted in Fig. 4.6, where we observe that ℓ must be chosen as large as possible, i.e., $\ell = 1$.

For the BLUE in Fig. 4.7, a smaller ℓ always yields a lower channel MSE just as for the LMMSE estimator and we also take $\ell = -2$. Note though that for complexity reasons, one has to be careful

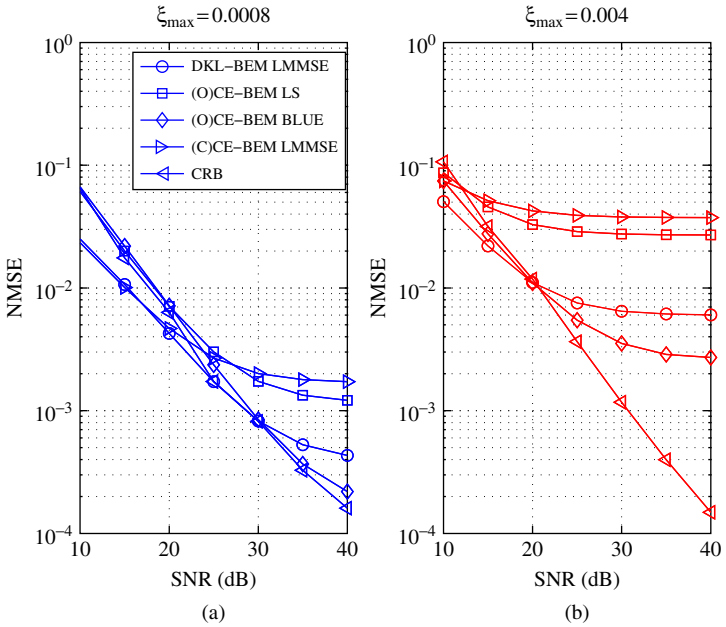

FIGURE 4.8

True channel NMSE versus SNR for different channel estimators. (a) $\xi_{max} = 0.0008$, (b) $\xi_{max} = 0.004$.

by taking ℓ too small because the BLUE has to be computed recursively and the procedure must be repeated for every OFDM symbol (the LMMSE estimator is in essence time-invariant and can thus be precomputed and stored off-line). A smaller ℓ often requires more iterations to reach convergence, and during each iteration, it requires a larger computational effort because more observation samples have to be processed.

Having set $\ell = -2$ for the LMMSE estimator, $\ell = 1$ for the LS estimator, and $\ell = -2$ for the BLUE, we now inspect their performance in terms of the true normalized mean squared error (NMSE), which can be defined as

$$\text{NMSE} \triangleq \frac{1}{K} \mathsf{E}_{\mathbf{h}} \left\{ \|\mathbf{h} - (\mathbf{U} \otimes \mathbf{I}_M)\hat{\mathbf{c}}\|^2 \right\}. \tag{4.45}$$

Note that the NMSE defined above explicitly takes the BEM modeling error into account. This is different from the channel MSE defined in (4.28), which merely indicates how close the estimated channel is to the best possible BEM fit. In Fig. 4.8, we depict the NMSE performance resulting from the LMMSE estimator based on a DKL-BEM, the LS estimator based on an (O)CE-BEM, and the BLUE based on an (O)CE-BEM. These results are compared with the LMMSE estimator that uses the (C)CE-BEM, which is proposed in Kannu & Schniter (2005). This estimator is different from the proposed LMMSE estimator because when adopting the (C)CE-BEM with $D = I = 3$, the frequency-domain channel matrix is strictly circularly banded and hence, the resulting LMMSE estimator cannot take the data-induced interference into account in the estimator design.

From Fig. 4.8, we can see that the NMSE performance of the LMMSE estimator based on the DKL-BEM is considerably worse if the channel varies faster. Recall that when constructing the DKL-BEM as well as the LMMSE estimator, we have assumed $\xi_{\max} = 0.002$ to emulate a mismatch of the channel statistics. The performance degradation suggests that underestimating the Doppler spread is more harmful than overestimating it. In contrast, the BLUE based on the (O)CE-BEM requires no knowledge of the channel statistics, and therefore exhibits a robust performance. We can see that its performance is quite close to the Cramer-Rao bound (CRB), whose derivation can be found in Tang, Cannizzaro, Banelli, & Leus (2007). In contrast, the LS estimator, although assuming no statistical knowledge at all, suffers from an inferior performance. However, it is still better than the LMMSE estimator based on the (C)CE-BEM (Kannu & Schniter, 2005) because the latter inherits a very large BEM modeling error.

4.3.5.2 Results for the Single-Carrier System

For the single-carrier system, we set the BEM length to $I = 5$. Furthermore, we assume the channel to be an FIR filter with $M = 6$ taps, which are independent Gaussian random variables with zero mean and an exponential power delay profile, i.e., $\mathsf{E}\{|h[n,m]|^2\} = c \cdot \mathrm{e}^{-m/10}$, for $m = 0,1,\ldots,M-1$, where c denotes a normalization constant. So we can model the time-varying channel with $MI = 30$ BEM coefficients.

We consider a data block of length $K = 256$, out of which 55 symbols are reserved for pilots. The pilot symbols are grouped in $G = 5$ clusters, each cluster containing $P+1 = 11$ pilots, so that the bandwidth efficiency is about 80%. In case Assumption 4.1 is applicable, we construct the pilot clusters accordingly, where in Fig. 4.3 the gray boxes are filled with zero pilots, and the black boxes are filled with nonzero pilots that are randomly picked constant-modulus symbols with the modulus chosen in such a way that the average pilot symbol power equals the data symbol power. We further take $G \geq I$ as indicated in Theorem 4.1. In case Assumption 4.1 is not applicable, we let all the pilot symbols take randomly picked constant-modulus symbols with the same modulus as the data symbols.

We first examine the effect of the parameter ℓ on the LMMSE estimator. As in the OFDM case, we use the DKL-BEM to approximate the time-varying channel. Both the DKL-BEM and the LMMSE estimator are based on the assumption of $\xi_{\max} = 0.002$ for both Doppler spreads under consideration. We know that ℓ is upper-bounded by the minimum of $(P+M-1)/2 = 7.5$ and $(P+M)/2 - MI/(2G) = 5$ and lower-bounded by the maximum of $\max\{0, M-\Delta-1\} = 0$ and $(M-\delta-1)/2 = -17.5$ because $\Delta \geq 40$ and $\delta = 40$. Hence, we look at the performance for $\ell = 0,3,5$, and depict the results in Fig. 4.9. It can be seen that a smaller ℓ leads to an improved performance although the improvement is not very significant, as for the OFDM system. The reason is again the mismatch in Doppler spread as well as the fact that the exponential power delay profile limits the spread of the pilot power over multiple observation samples.

We carry out the same simulation for the LS estimator, which is based on an (O)CE-BEM. A much more pronounced performance discrepancy can be observed from Fig. 4.10. The LS estimator only functions with the largest ℓ. In that case, the observation samples only depend on the pilot symbols, which is possible in the time domain where the channel matrix is strictly circularly banded. For other values of ℓ, data-induced interference emerges, and the LS estimator suffers from a high noise floor.

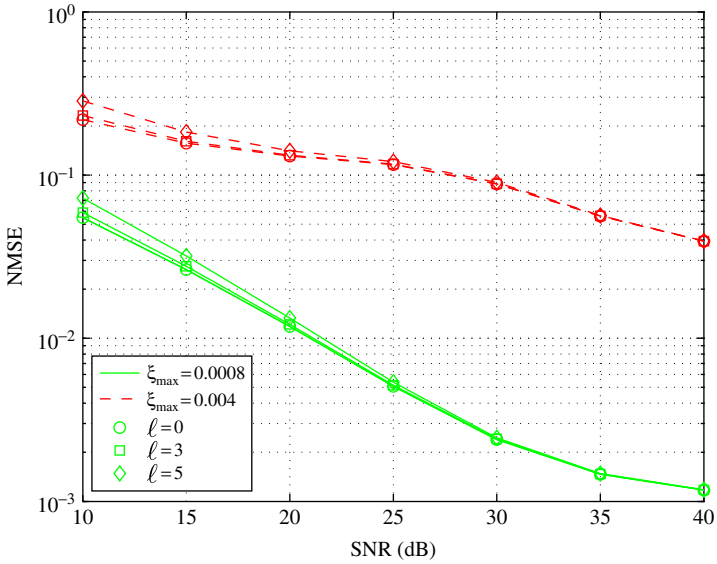

FIGURE 4.9

Performance of the LMMSE estimator in the single-carrier case.

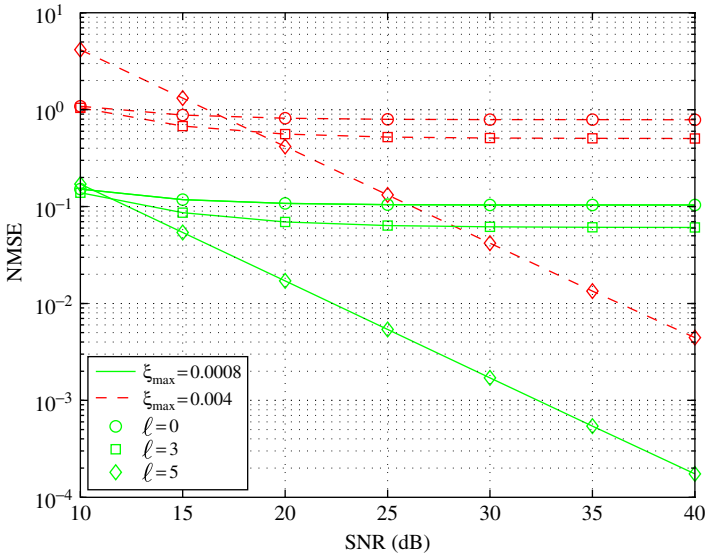

FIGURE 4.10

Performance of the LS estimator in the single-carrier case.

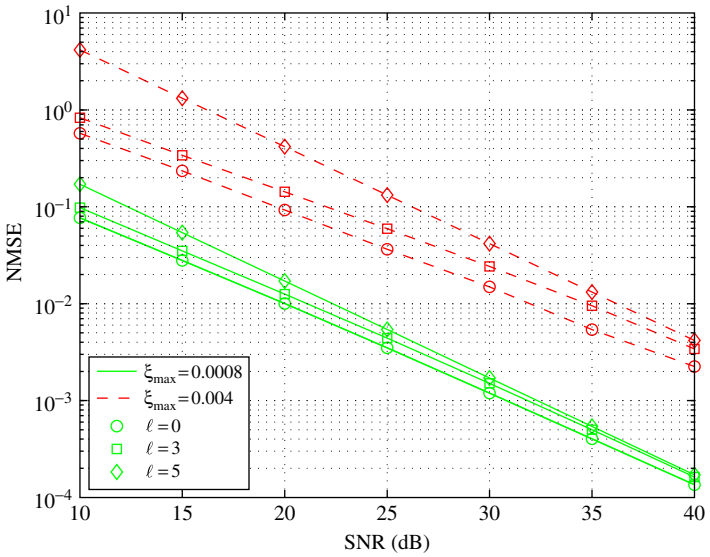

FIGURE 4.11

Performance of the BLUE in the single-carrier case.

The BLUE is also based on the (O)CE-BEM, and its behavior is analogous to that of the LMMSE estimator as can be seen in Fig. 4.11. However, the effect of ℓ can be more clearly observed than with the LMMSE estimator.

In Fig. 4.12, we compare the best performance of the different estimators, i.e., the LMMSE estimator with $\ell = 0$, the LS estimator with $\ell = 5$, and the BLUE with $\ell = 0$. The BLUE exhibits the best performance, especially if the channel varies fast. Note though that the complexity of the BLUE is the highest among the three methods.

4.4 CHANNEL ESTIMATION BASED ON MULTIPLE BLOCKS

4.4.1 Introduction

In the previous section, the channel is estimated for each block separately. To improve the performance, we will exploit the usage of multiple blocks in this section. It is, nonetheless, noteworthy that in the case of time-varying channels, the channel coherence time is rather short, which means that we cannot utilize an infinite number of blocks to enhance the estimation precision. Also note that because we are considering a time period that spans multiple blocks, the BEM could be less accurate, thereby inducing a larger modeling error.

For single-carrier systems, the channel estimation methods based on multiple blocks are basically similar to the ones based on a single block. We can simply consider a larger block size. As a result, we will not discuss this issue further.

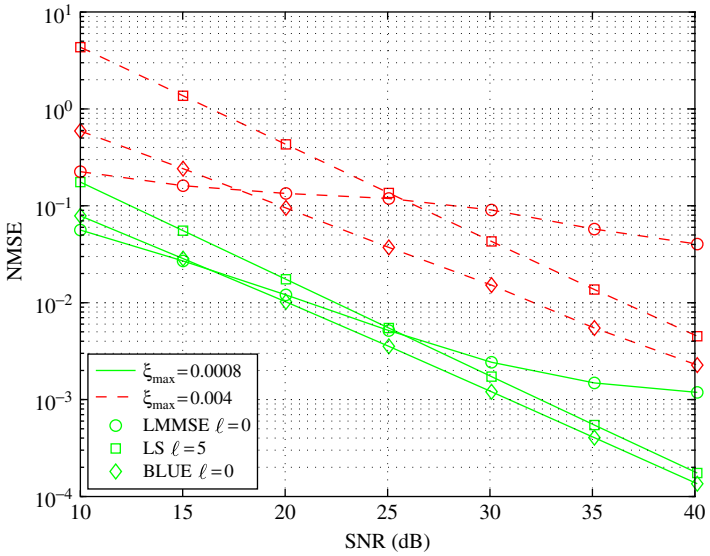

FIGURE 4.12

Comparison of the different channel estimators.

For OFDM systems, utilizing multiple blocks is, however, less straightforward. We have seen in the previous section that a pilot OFDM symbol[1] contains pilot symbols that are grouped in clusters along the frequency axis. Suppose that along the time axis, we consider J consecutive OFDM symbols, out of which V are pilot OFDM symbols. An interesting question can then be how to place the pilot symbols along both the frequency and time axes. To differentiate various pilot patterns, let us use the symbol $\mathscr{V}$ to denote the set that contains the indices of all the pilot OFDM symbols:

$$\mathscr{V} \triangleq \{j_0, \ldots, j_{V-1}\}, \tag{4.46}$$

where j_v stands for the position of the vth pilot OFDM symbol. The notation $\mathscr{P}[j_v]$ is comparable with the notation $\mathscr{P}$ introduced in the previous section, and it represents the set of pilot subcarriers within the vth pilot OFDM symbol.

Adopting the terms introduced in Coleri, Ergen, Puri, & Bahai (2002), we will basically focus on three pilot-placement scenarios.

1. The first scheme, referred to as the comb-type, is adopted in Li, Cimini, & Sollenberger (1998); Yang, Letaief, Cheng, & Cao (2001); Schafhuber, Matz, & Hlawatsch (2003); Athaudage & Jayalath (2004); Mostofi & Cox (2005); Schafhuber & Matz (2005); Tang, Leus, & Banelli (2006); Cannizzaro, Banelli, & Leus (2006). In this scheme, pilot symbols occupy only a fraction of the subcarriers, but such pilot symbols are carried by each OFDM symbol. In other words, we

[1] In this section, we refer to an OFDM symbol that carries some pilot symbols as a pilot OFDM symbol.

have $V = J$ and $|\mathscr{P}[j_v]| < K$, for $v = 0, \ldots, V-1$. This basically is the pilot scheme that we discussed in the previous section, but now extended to multiple OFDM symbols.

2. The second scheme, referred to as the block-type, is considered in Choi, Voltz, & Cassara (2001); Cui, Tellambura, & Wu (2005); Schniter (2006). In the block-type scheme, the pilot symbols occupy the entire OFDM symbol, and such pilot OFDM symbols are interleaved with data OFDM symbols. More specifically, this means that $V < J$ and $|\mathscr{P}[j_v]| = K$ for $v = 0, \ldots, V-1$.
3. The third scheme, considered in Choi & Lee (2005), is referred to as the mixed-type, which is a compromise between the comb-type and the block-type. To be more specific, the pilot symbols only occupy a fraction of the subcarriers, and such pilot OFDM symbols are interleaved with data OFDM symbols. So, we have $V < J$ and $|\mathscr{P}[j_v]| < K$ for $v = 0, \ldots, V-1$.

Some examples of these three pilot schemes are sketched in Fig. 4.13.

An intriguing question is which scheme is able to yield the most reliable channel estimate under the same bandwidth and power restrictions. Interestingly enough, conflicting results are reported; e.g., Negi & Cioffi (1998), Sanzi, Jelting, & Speidel (2003) and Choi & Lee (2005) advocate the comb-type or the mixed-type, while the block-type scheme is preferred in Cui, Tellambura, & Wu (2005) and Schniter (2006). Common to all pilot schemes is that the channel estimation method can be virtually decomposed into two steps: first, preliminary channel estimates are acquired for individual pilot OFDM symbols; second, these preliminary results are interpolated, with the aid of, e.g., channel statistics or basis functions, to obtain the final channel estimate. In terms of the intermediate channel estimation performance, the block-type scheme usually yields a better channel estimate because all the subcarriers are occupied by pilot symbols. The comb-type or mixed-type schemes work inferior because ICI leads inevitably to a noise floor.

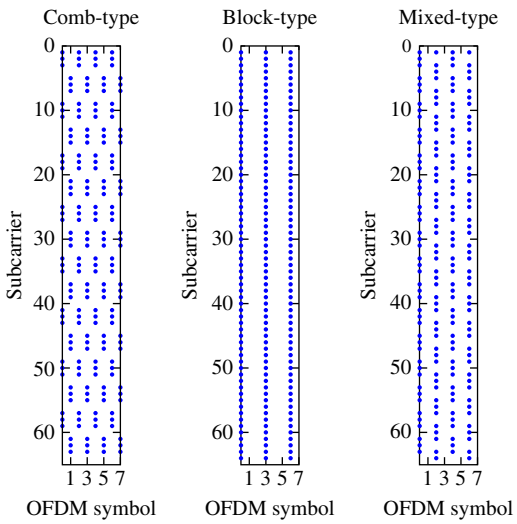

FIGURE 4.13

Illustration of different pilot schemes. The horizontal axis corresponds to the time; the vertical axis corresponds to the subcarrier positions; and the position where a pilot symbol is located is represented by a dot.

However, the block-type scheme is subject to a larger interpolation error with respect to the comb-type and mixed-type schemes because the pilot OFDM symbols are farther apart.

In the remainder of this section, we will provide a more detailed discussion of the impact of diverse pilot placement schemes on the channel estimation performance.

4.4.2 Data Model and BEM for Multiple OFDM Symbols

Compared with Section 4.3, we need to use a larger BEM to approximate the time variation of the channel over several OFDM symbols. More specifically, we will now model $J(K+M-1)$ consecutive samples of the mth channel tap, e.g., $h[0,m],\ldots,h[(J-1)(K+M-1)-1,m]$, as

$$\begin{pmatrix} h[0,m] \\ \vdots \\ h[(J-1)(K+M-1)-1,m] \end{pmatrix} \approx \underbrace{\begin{pmatrix} \mathbf{u}_0 & \cdots & \mathbf{u}_{I-1} \end{pmatrix}}_{\mathbf{U}} \begin{pmatrix} c_0[m] \\ \vdots \\ c_{I-1}[m] \end{pmatrix},$$

where $I-1$ is the BEM order, the $J(K+M-1)\times 1$ vector $\mathbf{u}_i$ is the ith BEM function, and $c_i[m]$ is the ith BEM coefficient of the mth channel tap over J OFDM symbols including the CP. Note that unlike in the single-symbol case, we also consider the part of the channel related to the CP in this model, although this is not very important.

As a result, for the jth OFDM symbol, we have

$$\begin{pmatrix} h[j(K+M-1)+M-1,m] \\ \vdots \\ h[(j+1)(K+M-1)-1,m] \end{pmatrix} \approx \underbrace{\begin{pmatrix} \mathbf{u}_0[j] & \cdots & \mathbf{u}_{I-1}[j] \end{pmatrix}}_{\mathbf{U}[j]} \begin{pmatrix} c_0[m] \\ \vdots \\ c_{I-1}[m] \end{pmatrix}, \tag{4.47}$$

where $\mathbf{U}[j]$ and $\mathbf{u}_i[j]$ are a selection of rows $j(K+M-1)+M-1$ through $(j+1)(K+M-1)-1$ from $\mathbf{U}$ and $\mathbf{u}_i$, respectively. So repeating (4.47) for all channel taps, we obtain

$$\mathbf{h}[j] \approx (\mathbf{U}[j]\otimes \mathbf{I}_M)\mathbf{c},$$

where $\mathbf{h}[j] \triangleq (\mathbf{h}_0^T[j] \cdots \mathbf{h}_{K-1}^T[j])^T$, with $\mathbf{h}_k[j] \triangleq (h[j(K+M-1)+M-1+k,0] \cdots h[j(K+M-1)+M-1+k,M-1])^T$, stacks all channel taps in the jth OFDM symbol and where $\mathbf{c} \triangleq (\mathbf{c}_0^T \cdots \mathbf{c}_{I-1}^T)^T$, with $\mathbf{c}_i \triangleq (c_i[0] \cdots c_i[M-1])^T$, stacks the BEM coefficients of all channel taps in all OFDM blocks.

By defining the BEM in this way, the resulting frequency-domain channel matrix of the jth OFDM symbol will admit an expression that is slightly different from (4.10):

$$\mathbf{H}_{\mathrm{d}}[j] = \sum_{i=0}^{I-1} \mathbf{W}_K^H \mathbf{U}_{\mathrm{d},i}[j] \mathbf{W}_K \mathbf{W}_K^H \mathbf{C}_{\mathrm{c},i} \mathbf{W}_K, \tag{4.48}$$

where $\mathbf{U}_{\mathrm{d},i}[j]$ is a diagonal matrix with $\mathbf{u}_i[j]$ on its diagonal, and $\mathbf{C}_{\mathrm{c},i}$ is a circulant matrix with $(\mathbf{c}_i^T \ \mathbf{0}_{1\times(K-M)})^T$ on its first column. It is clear that (4.48) is different from (4.10) in the sense that the BEM sequence $\mathbf{u}_i[j]$ is different for each OFDM symbol but the BEM coefficients $\mathbf{c}_i$ are not, while in (4.10), there is a common BEM for each OFDM symbol, but with different BEM coefficients.

For the same reason, we can state that for the vth pilot OFDM symbol, the observation samples collected in $\mathbf{y}^{\{\mathscr{O}[j_v]\}}[j_v]$ admit the following expression (cf. (4.23)):

$$\mathbf{y}^{\{\mathscr{O}[j_v]\}}[j_v] = \mathbf{A}[j_v]\mathbf{c} + \mathbf{i}[j_v] + \mathbf{z}^{\{\mathscr{O}[j_v]\}}[j_v],$$

with (cf. (4.24), (4.25), and (4.26))

$$\begin{aligned}\mathbf{A}[j_v] \triangleq \mathbf{W}_K^{\{\mathscr{O}[j_v],:\}}(\mathbf{U}_{\mathrm{d},0}[j_v] \cdots \mathbf{U}_{\mathrm{d},I-1}[j_v])\left(\mathbf{I}_I \otimes \mathbf{W}_K^{\{\mathscr{P}[j_v],:\}H}\right)\\ \times \left(\mathbf{I}_I \otimes \mathrm{diag}\{\mathbf{p}[j_v]\}\mathbf{V}_K{}^{\{\mathscr{P}[j_v],:\}}\right).\end{aligned}$$

Note that we have now made the OFDM symbol index j_v explicit to show which variables depend on the vth pilot OFDM symbol.

By repeating the above steps for all the pilot OFDM symbols and stacking the results, we obtain

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \mathbf{i} + \mathbf{z}. \tag{4.49}$$

It is important to remark here that in the block-type scheme, the pilot symbols and unknown data symbols are in different OFDM symbols, and thus they are not interfering. In that case, (4.49) reduces to

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \mathbf{z}.$$

Clearly, (4.49) bears an almost identical form as (4.27) for the single OFDM symbol case. Hence, all the channel estimation methods introduced in Section 4.3.3 can also be applied here with some trivial modifications. For the sake of space, we will not give all the details here, but refer the interested reader to Tang (2007).

4.4.3 Channel Identifiability Based on Multiple Blocks

In the previous section, while focusing on a single block, we said the channel is identifiable if the matrix $\mathbf{A}[j]$ has full column-rank. In practical situations, the conditions given in Theorem 4.2 may be hard to satisfy for a single OFDM symbol. We have mentioned one such example in Remark 4.6. Another quite common scenario is that the channel delay spread M could be too large to afford a sufficiently large number G of pilot clusters as per (4.44). A possible solution is to employ multiple pilot OFDM symbols. Now that the matrix $\mathbf{A}$ in (4.49) is a stack of several $\mathbf{A}[j_v]$ matrices, we will study in this section the impact of using multiple pilot OFDM symbols on the full column-rank condition for $\mathbf{A}$.

To understand this better, let us first rewrite $\mathbf{A}$ in (4.49) as a product of two matrices:

$$\mathbf{A} = \begin{pmatrix} \overline{\mathbf{A}}_{\mathrm{c}}[j_0] & & \\ & \ddots & \\ & & \overline{\mathbf{A}}_{\mathrm{c}}[j_{V-1}] \end{pmatrix} \begin{pmatrix} \overline{\mathbf{A}}_{\mathrm{d}}[j_0] \\ \vdots \\ \overline{\mathbf{A}}_{\mathrm{d}}[j_{V-1}] \end{pmatrix}, \tag{4.50}$$

where, by adopting the FDKD pilot structure of Assumption 4.2 in Section 4.3.4, we have (cf. (4.35) and (4.36))

$$\overline{\mathbf{A}}_{\mathrm{c}}[j_v] \triangleq \mathbf{W}_K^{\{\mathscr{O}[j_v],:\}}(\mathbf{U}_{\mathrm{d},0}[j_v] \cdots \mathbf{U}_{\mathrm{d},I-1}[j_v])\left(\mathbf{I}_I \otimes \mathbf{W}_K^{\{\overline{\mathscr{P}}[j_v],:\}H}\right), \tag{4.51}$$

$$\overline{\mathbf{A}}_{\mathrm{d}}[j_v] \triangleq \mathbf{I}_I \otimes \mathrm{diag}\{\overline{\mathbf{p}}[j_v]\}\mathbf{V}_K^{\{\overline{\mathscr{P}}[j_v],:\}}. \tag{4.52}$$

Using Assumption 4.3 in Section 4.3.4, we are now able to obtain two useful lemmas. The first lemma concerns the situation where every pilot OFDM symbol carries the same FDKD structure, i.e., the pilot values and positions are identical: $\overline{\mathbf{p}}[j_0] = \cdots = \overline{\mathbf{p}}[j_{V-1}]$ and $\overline{\mathscr{P}}[j_0] = \cdots = \overline{\mathscr{P}}[j_{V-1}]$. Here, $\overline{\mathbf{p}}[j_v]$ stands for the nonzero pilots, and $\overline{\mathscr{P}}[j_v]$ stands for their positions in the vth pilot OFDM symbol.

Lemma 4.3 Under Assumptions 4.2 and 4.3, if each pilot OFDM symbol carries the same pilot structure, then the $VK \times MI$ matrix $\mathbf{A}$ will have full column-rank provided that

$$M \le G \le VK/I.$$

Proof. If each pilot OFDM symbol carries the same pilot structure, i.e., $\overline{\mathbf{A}}_{\mathrm{d}}[j_0] = \cdots = \overline{\mathbf{A}}_{\mathrm{d}}[j_{V-1}] = \overline{\mathbf{A}}_{\mathrm{d}}$, (4.50) becomes

$$\mathbf{A} = \begin{pmatrix} \overline{\mathbf{A}}_{\mathrm{c}}[j_0] \\ \vdots \\ \overline{\mathbf{A}}_{\mathrm{c}}[j_{V-1}] \end{pmatrix} \overline{\mathbf{A}}_{\mathrm{d}}.$$

As per Lemma 4.1, $\overline{\mathbf{A}}_{\mathrm{d}}$ has rank $\min\{G,M\}I$. Similar to the proof of Lemma 4.2, we can show that the rank of $(\overline{\mathbf{A}}_{\mathrm{c}}^T[j_0] \cdots \overline{\mathbf{A}}_{\mathrm{c}}^T[j_{V-1}])^T$ equals $\min\{VK/G,I\}G$. Then applying the rank inequality (4.43) concludes the proof. ■

Different from Lemma 4.3, let us consider a scenario where the nonzero pilot symbols are located on different subcarriers in each pilot OFDM symbol, i.e., $\overline{\mathscr{P}}[j_0] \neq \cdots \neq \overline{\mathscr{P}}[j_{V-1}]$. This leads to the following lemma.

Lemma 4.4 Under Assumptions 4.2 and 4.3, if the positions of the nonzero pilot symbols in each pilot OFDM symbol are different, the matrix $\mathbf{A}$ will have full column-rank provided that

$$M/V \le G \le K/I.$$

Proof. We can observe from (4.50) that the rank of the left matrix, which is a block diagonal matrix, is equal to the sum of the ranks of the diagonal blocks. As per Lemma 4.2, each of those diagonal blocks has rank $\min\{K/G,I\}G$, and the total rank of the left matrix is $\min\{K/G,I\}GV$. To study the rank of the right matrix in (4.50), let us permute its rows to obtain (cf. (4.52)):

$$\mathbf{I}_I \otimes \begin{pmatrix} \mathrm{diag}\{\overline{\mathbf{p}}[j_0]\} & & \\ & \ddots & \\ & & \mathrm{diag}\{\overline{\mathbf{p}}[j_{V-1}]\} \end{pmatrix} \begin{pmatrix} \mathbf{V}_K^{\{\overline{\mathscr{P}}[j_0],:\}} \\ \vdots \\ \mathbf{V}_K^{\{\overline{\mathscr{P}}[j_{V-1}],:\}} \end{pmatrix}. \tag{4.53}$$

Because the positions of the VG nonzero pilots are different in each pilot OFDM symbol, the $VG \times M$ matrix $\left(\mathbf{V}_K^{\{\overline{\mathscr{P}}[j_0],:\}T} \cdots \mathbf{V}_K^{\{\overline{\mathscr{P}}[j_{V-1}],:\}T}\right)^T$ consists of VG different rows of the Vandermonde matrix $\mathbf{V}_K$, and will have rank $\min\{VG, M\}$. Then again applying the rank inequality (4.43) concludes the proof. ■

Let us now combine the scenarios considered in Lemma 4.3 and Lemma 4.4 and introduce the following assumption.

Assumption 4.4 The total number of pilot OFDM symbols satisfies $V = V_a V_b$ 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 take exactly the same pilot structure, whereas among different clusters, we will give the nonzero pilot symbols different positions.

We then obtain the following theorem that guarantees the full column-rank condition for $\mathbf{A}$, or in other words channel identifiability.

Theorem 4.3 Under Assumptions 4.2, 4.3, and 4.4, the channel is identifiable over multiple OFDM symbols if the number of pilot clusters G inside each pilot OFDM symbol satisfies

$$M/V_a \leq G \leq V_b K/I. \tag{4.54}$$

The proof is essentially analogous to the proofs of Lemmas 4.3 and 4.4, and will be omitted here.

Compared with Theorem 4.2 in Section 4.3.4, we can easily observe that Theorem 4.3 relaxes the requirements on the number of pilot clusters G within a pilot OFDM symbol, which is beneficial to the bandwidth efficiency.

4.4.4 Simulation Results

In this section, we test different pilot schemes for an OFDM system, based on multiple blocks. The channel and OFDM parameters are the same as for the single OFDM symbol case.

We use the BLUE channel estimator and approximate the channel time-variation by means of the (O)CE-BEM, which spans in total $J = 8$ OFDM symbols including the CP, with each OFDM symbol containing $K = 64$ subcarriers. The order of the BEM is associated with the normalized Doppler spread $\xi_{\max}$. In the simulation, we set $I = 5$ if $\xi_{\max} \leq 0.002$; otherwise, we take $I = 9$. Note that for every pilot OFDM symbol, all observation samples are considered.

The pilot schemes to be compared are plotted in Fig. 4.13. In the comb-type scheme, all OFDM symbols are pilot OFDM symbols, thus $\mathscr{V} = \{0, 1, \ldots, 7\}$. Inside each pilot OFDM symbol, the pilots have the FDKD structure with $G = 8$ clusters of length $P + 1 = 3$. In the block-type scheme, the indices of the pilot OFDM symbols are given by $\mathscr{V} = \{0, 3, 6\}$. Inside each pilot OFDM symbol, we further set $(G, P+1) = (16, 3)$. Note that different from the traditional works, we let the block-type scheme also carry FDKD pilots to ensure the channel identifiability (see Remark 4.6). In the mixed-type scheme, we set $\mathscr{V} = \{0, 2, 4, 6\}$ and $(G, P+1) = (16, 3)$. In this way, all the pilot schemes result in an equal loss in bandwidth efficiency of about 37.5%.

As before, we use the true NMSE between the channel estimate and the actual time-varying channel as performance measure:

$$\text{NMSE}[j] \triangleq \frac{1}{K} \mathsf{E}_{\mathbf{h}}\{\|\mathbf{h}[j] - (\mathbf{U}[j] \otimes \mathbf{I}_M)\hat{\mathbf{c}}\|^2\}, \tag{4.55}$$

where $\mathbf{U}[j]$ is defined in (4.47) as the part of the BEM matrix $\mathbf{U}$ that corresponds to the jth 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-type and mixed-type schemes, and the fifth and sixth OFDM symbols in the block-type scheme. The channels in the remaining OFDM symbols can be estimated when the window slides forward.

In the left and right plot of Fig. 4.14, we depict the performance of the three pilot schemes for short channels, $M = 4$, with normalized Doppler spreads $\xi_{\max} = 0.0008$ and $\xi_{\max} = 0.004$, respectively. In addition, we list the performance based on a single OFDM symbol, which is a special case of the comb-type scheme but considering $J = V = 1$. We can observe in the left plot of Fig. 4.14 that when the channel fading is slow, the three pilot schemes yield a similar performance, which is better than

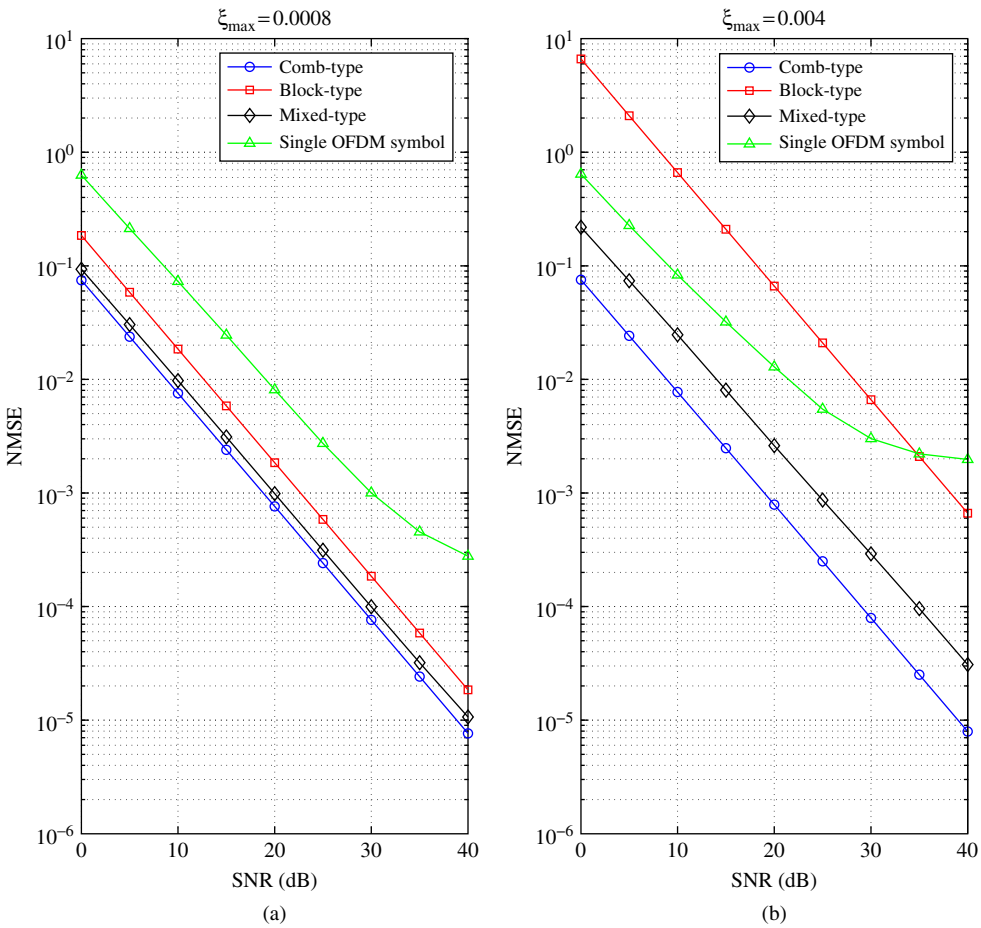

FIGURE 4.14

True channel NMSE versus SNR for $M = 4$. (a) $\xi_{\max} = 0.0008$, (b) $\xi_{\max} = 0.004$.

the performance of the single OFDM symbol case. When the channel fades faster, as illustrated in the right plot, the block-type scheme experiences more difficulty in tracking the channel compared with the other schemes. This is due to the fact that the pilot symbols are grouped in complete OFDM symbols and hence different pilot OFDM symbols are farther apart than in the other two schemes. Because the channel estimation methods implicitly carry out some interpolation over the considered time span, it is clear that the block-type scheme cannot promptly react to a time-varying channel.

The results for a much longer channel, $M = 16$, are depicted in Fig. 4.15. We observe again that when the channel varies slowly, the block-type scheme exhibits a similar performance as the other two schemes, 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 than with $M = 4$, and suffers from a high noise floor at a moderate-to-high SNR. Compared with the other channel situations, this suggests that the comb-type scheme is inferior for channels that have a large spread

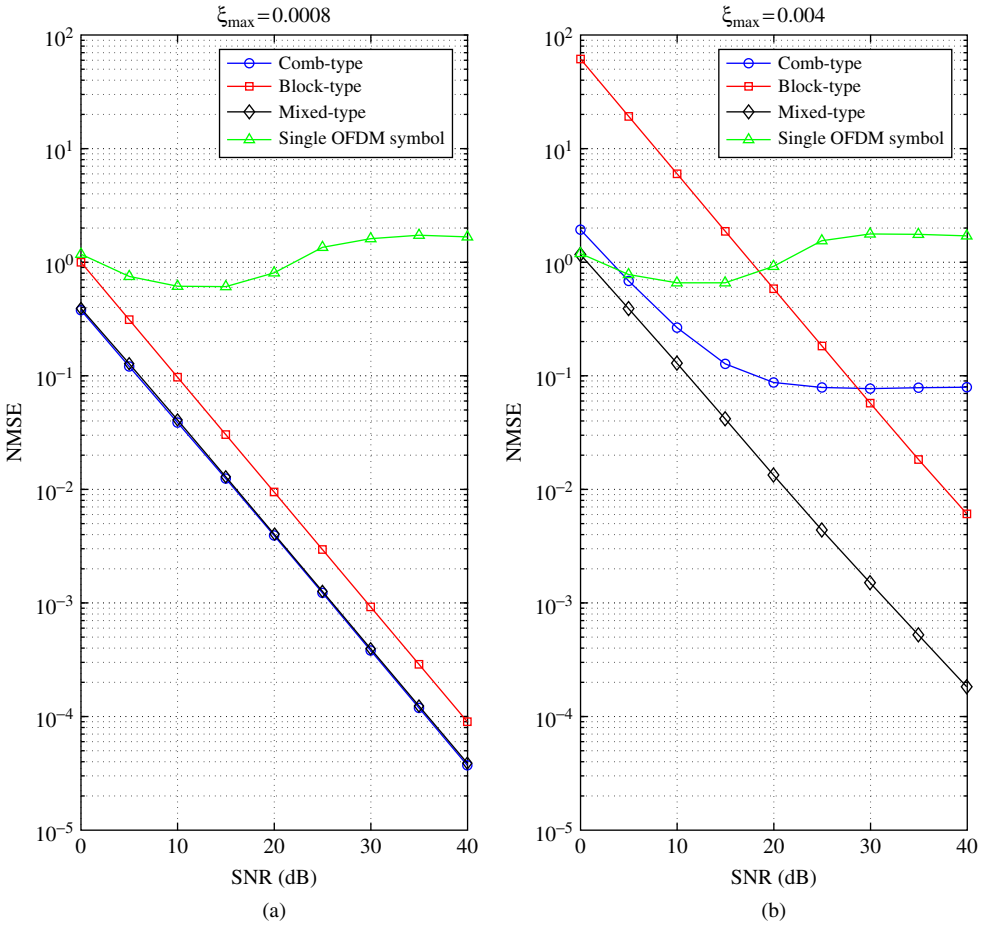

FIGURE 4.15

True channel NMSE versus SNR for $M = 16$. (a) $\xi_{max} = 0.0008$, (b) $\xi_{max} = 0.004$.

in both the Doppler and delay domains. It is noteworthy that the mixed-type scheme satisfies the conditions of both Theorem 4.2 and Theorem 4.3 for both 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.

4.5 EXTENSION TO MIMO SYSTEMS

4.5.1 Introduction

In this section, we extend the methods discussed for a single transmit and receive antenna, or the single-input single-output (SISO) case, to multiple transmit and receive antennas, or the so-called multiple-input multiple-output (MIMO) case.

Consider a system with M_{T} transmit and M_{R} receive antennas. We assume again that every transmit antenna groups the pilot symbols in G clusters of length $P+1$, where the gth pilot cluster of the mth transmit antenna $\mathbf{p}_g^{(m)}$ has indices in $\mathscr{P}_g^{(m)} = \{P_g^{(m)}, \ldots, P_g^{(m)} + P\}$. To this pilot cluster, we again assign a set of observation samples with indices collected in the set $\mathscr{O}_g^{(m)}$, similarly defined as in (4.11) (for single-carrier systems) and (4.18) (for OFDM systems), with P_g replaced by $P_g^{(m)}$. To perform channel estimation, every receive antenna then selects the observation samples with indices in $\mathscr{O} = \{\mathscr{O}_0^{(1)}, \ldots, \mathscr{O}_{G-1}^{(1)}, \ldots, \mathscr{O}_{G-1}^{(M_{\mathrm{T}})}\}$. With this in mind, it is easy to extend the data models of (4.16) and (4.23) to multiple transmit and receive antennas. Also the channel estimators are straightforward extensions of the channel estimators derived for the SISO case.

For single-carrier systems, when $\ell \geq M-1$, there is no interference from unknown data symbols, and the channel estimation problem can actually be separated into the different receive antennas. However, when $\ell < M-1$, there is interference from unknown data symbols and the performance could benefit from using the observation samples from the other receive antennas.

For OFDM systems, however, we know that there is always interference from unknown data symbols and that all receive antennas should be treated together if we want to obtain the best possible performance.

Let us now briefly summarize some channel identifiability results for MIMO systems. Proofs can be derived in a similar manner as for the SISO case.

4.5.2 Single-Carrier System

Let us again start by adopting the following assumption on the pilot structure:

Assumption 4.5 The pilot structure satisfies the following conditions:

- **(C1)** The length $P+1$ of each pilot cluster satisfies $P+1 \geq 2\ell+1$.
- **(C2)** Inside each pilot cluster, either the first or the last ℓ pilots are zero.
- **(C3)** Every pilot cluster has at least one nonzero pilot in between the first and last ℓ pilots.
- **(C4)** The nonzero pilot symbols from different pilot clusters should be separated by at least $M-1$ symbols, viewed over all transmit antennas.

Note that these conditions are the same as in the SISO case (see Assumption 4.1 in Section 4.3.4) with the exception that condition (C4) should hold over all transmit antennas.

We can now state the following theorem:

Theorem 4.4 Under Assumption 4.5, the channel is identifiable if the number of pilot clusters per transmit antenna G is greater than or equal to the BEM length, i.e.,

$$G \geq I.$$

Note that this condition is the same as before (see Theorem 4.1), which is pleasing. However, in order to reduce the interantenna interference (IAI), it is better to overlay the nonzero pilot clusters on one transmit antenna with zero pilot clusters of the same length on the other transmit antennas. When $\ell \geq M-1$, this operation will actually separate the channel estimation problem into the different transmit antennas.

Remark 4.7 The TDKD structure has been extended to the MIMO case in Kannu & Schniter (2006) and Yang, Ma, & Giannakis (2004). It is a pilot structure where the pilot clusters on all transmit antennas overlap and where on the mth transmit antenna, the G pilot clusters are equidistant and every cluster consists of a nonzero pilot symbol surrounded by $M-1+(m-1)M$ zeros to the left and $M-1+(M_{\mathrm{T}}-m)M$ zeros to the right, i.e., $P+1=(M_{\mathrm{T}}+1)M-1$. As in the SISO case, the MIMO TDKD structure has been proven to be optimal in the MSE and capacity sense provided the (C)CE-BEM is used for channel modeling (Kannu & Schniter, 2006; Yang, Ma, & Giannakis, 2004). The MIMO TDKD structure is a special case of the pilot structure defined by Assumption 4.5 with $P+1=(M_{\mathrm{T}}+1)M-1$ and $\ell=M-1$. In that case, we know from Theorem 4.4 that the channel is identifiable if $G \geq I$, as also argued in Kannu & Schniter (2006) and Yang, Ma, & Giannakis (2004) for the (C)CE-BEM.

4.5.3 OFDM System

For OFDM systems, we again adopt a more specific pilot structure.

Assumption 4.6 The pilot structure satisfies the following conditions:

(C1) The length $P+1$ of each pilot cluster satisfies $P+1 \geq 1$.
(C2) Every pilot cluster has one nonzero pilot.
(C3) The nonzero pilots are equispaced, viewed over all transmit antennas.

As in the single-carrier case, the conditions are the same as in the SISO case (see Assumption 4.2 in Section 4.3.4) with the exception that condition (C3) should hold over all transmit antennas. Note that this means that the nonzero pilots should not necessarily be equispaced on every transmit antenna.

Remark 4.8 In Kannu & Schniter (2006), also the FDKD structure has been extended to the MIMO case. It is a pilot structure where the pilot clusters on all transmit antennas overlap and where on the mth transmit antenna, the G pilot clusters are equidistant and every cluster consists of a nonzero pilot symbol surrounded by $D-1+(m-1)D$ zeros to the left and $D-1+(M_{\mathrm{T}}-m)D$ zeros to the right, i.e., $P+1=(M_{\mathrm{T}}+1)D-1$, where D is the significant discrete Doppler spread. As in the SISO case, the MIMO FDKD structure is optimal in the MSE sense if the (C)CE-BEM is used for channel modeling (Kannu & Schniter, 2006). However, this time, we cannot say that the MIMO FDKD structure is a special case of Assumption 4.6, since condition (C3) is not satisfied. The reason why we need condition (C3) is because we want to prove identifiability for a general BEM and not only for the (C)CE-BEM. Under the (C)CE-BEM, it would be enough to replace condition (C3) by the requirement that the nonzero pilots on the same transmit antenna are equispaced and the nonzero pilots from all transmit antennas are

separated by at least $D-1$ symbols. In that case, the MIMO FDKD structure would fit into Assumption 4.6 with $P+1=(M_{\mathrm{T}}+1)D-1$.

Similarly to the SISO case (see Theorem 4.2), we can prove that for a single block, the following theorem holds:

Theorem 4.5 Under Assumptions 4.6 and 4.3, the channel is identifiable if the number of pilot clusters per transmit antenna G satisfies

$$M \leq G \leq K/(IM_T). \tag{4.56}$$

So the lower bound is the same as in the SISO case, but the upper bound is decreased with the number of transmit antennas M_T. As in the single-carrier case, we can reduce the IAI by overlaying the nonzero pilot clusters on one transmit antenna with zero pilot clusters of the same length on the other transmit antennas. In contrast to the single-carrier case though, we will generally not be able to completely separate the channel estimation problem into the different transmit antennas.

Remark 4.9 In Dai (2007), optimal pilot structures have been proposed for MIMO OFDM systems. However, it is assumed there that the discrete Doppler spread is exactly limited to D, the pilot clusters on all transmit antennas overlap and have length $P+1=D$, and one observation sample per pilot cluster is considered. The condition on the number of clusters G per transmit antenna then becomes $G \geq M_{\mathrm{T}}MI$, which is much worse than our condition in (4.56).

For multiple blocks and a mixed-mode pilot scheme, we can actually consider two cases, as illustrated in Fig. 4.16. When the pilot OFDM symbols from the different transmit antennas do not overlap (left side of Fig. 4.16), the different transmit antennas do not interfere with each other as far as channel identifiability is concerned, and we can actually use the results from the SISO case.

Theorem 4.6 Under Assumptions 4.2, 4.3, and 4.4, and assuming that the pilot OFDM symbols from the different transmit antennas do not overlap, the channel is identifiable over multiple OFDM symbols if the number of pilot clusters per transmit antenna G inside each pilot OFDM symbol satisfies

$$M/V_a \leq G \leq V_bK/I.$$

When the pilot OFDM symbols from the different transmit antennas overlap (right side of Fig. 4.16), however, we have a more complicated situation. In that case, the following theorem applies:

Theorem 4.7 Under Assumptions 4.6, 4.3, and 4.4, and assuming that the pilot OFDM symbols from the different transmit antennas overlap, the channel is identifiable over multiple OFDM symbols if the number of pilot clusters per transmit antenna G inside each pilot OFDM symbol satisfies

$$M/V_a \leq G \leq V_bK/(IM_T).$$

So as before, only the upper bound has been affected and is decreased by the factor M_T. The IAI can again be reduced by inserting additional zero pilot clusters overlaying the nonzero pilot clusters as indicated before, and this holds whether the pilot OFDM symbols from the different transmit antennas overlap or not.

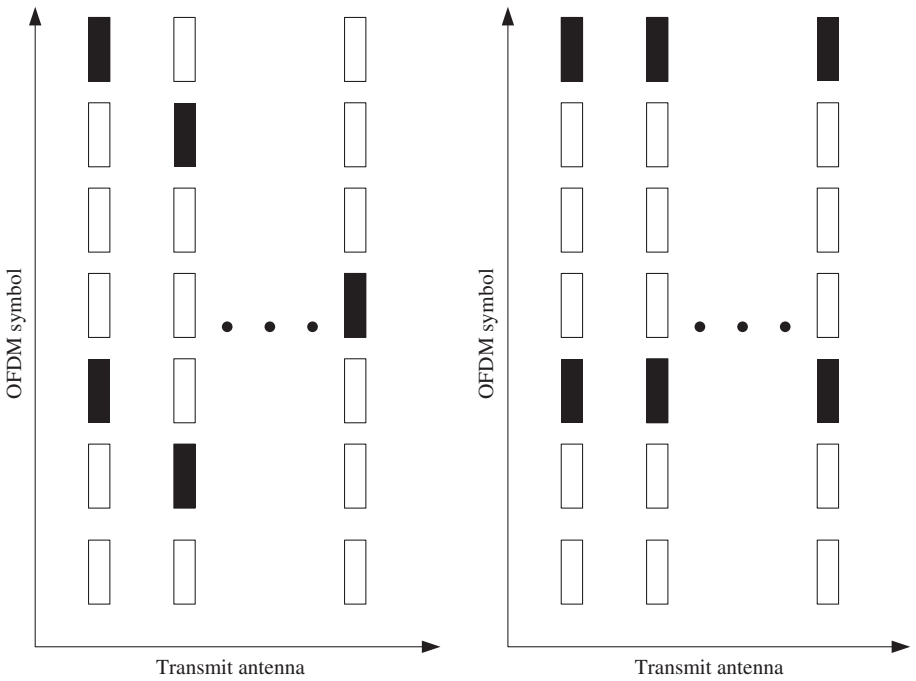

FIGURE 4.16

Two possible pilot structures for MIMO–OFDM, when multiple symbols are considered. The black boxes represent pilot OFDM symbols, whereas the white boxes represent data OFDM symbols.

4.6 ADAPTIVE CHANNEL ESTIMATION

In the earlier sections, we have shown batch approaches to estimate the time-varying channel BEM coefficients by dedicated pilot symbols in single-carrier and OFDM scenarios, with a single block or multiple blocks of observations. In this section, we will explain how all the proposed approaches can be cast in an adaptive framework, which could be useful to track not only the instantaneous channel variations but also the possible nonwide-sense stationary channel conditions and, potentially, to reduce the amount of redundant pilots. Whichever is the design of the pilots and of the observation samples, we have shown that the channel estimation relies on the observation equations (4.27). For notational simplicity, we will assume that each observation window is contiguous to the next one and coincident with a single-block duration, either in the single-carrier or the OFDM scenario. However, the time-domain formulation of the channel-estimation problem, especially in single-carrier scenarios, can naturally lead to observation windows that overlap in time. Thus, the results that follow can be easily extended to an overlapped observation approach. This overlapped observation approach may be particularly useful in very high Doppler scenarios in order to avoid abrupt changes in the evolution of the BEM coefficients, which may be hard to track otherwise.

To further simplify the problem, we also assume that the pilot design and the observation samples are chosen in such a way that the interference term in (4.27) can be neglected and, consequently, the observation equation reduces to

$$\mathbf{y}^{\{\mathscr{O}[n]\}}[n] = \mathbf{A}[n]\mathbf{c}[n] + \mathbf{z}^{\{\mathscr{O}[n]\}}[n], \tag{4.57}$$

where, as before, the index n has been used to identify the nth observation window. The observation matrix $\mathbf{A}[n]$, defined in (4.17), may depend on n or not, depending on how we select the observation index set $\mathscr{O}[n] = \{\mathscr{O}_0[n], \ldots, \mathscr{O}_G[n]\}$ and the pilot design $\mathbf{p}[n]$, which may change from one block to another. Note that $\mathbf{A}[n]$ may also depend on n via the underlaying BEM basis matrix $\mathbf{U}$, which, instead of being fixed for all n, can be smoothly linked from one block to another in order to facilitate tracking (see He & Tugnait, 2007). In the adaptive channel estimation literature, it is quite common to assume that a time-varying channel obeys an auto-regressive (AR) model evolution, which in the simplest case reduces to a Gauss–Markov process evolution. Several authors assume that also the BEM channel coefficients evolve in a similar fashion from one observation window to another, as expressed by

$$\mathbf{c}[n] = \mathbf{\Phi}[n]\mathbf{c}[n-1] + \boldsymbol{\beta}[n], \tag{4.58}$$

where $\boldsymbol{\beta}[n]$ represents the BEM coefficient innovation with respect to the $(n-1)$st observation window. Actually, (4.57) and (4.58) are the observation and the model evolution equations, respectively, that typically cast the channel-estimation problem in a vector-state vector-observation (VSVO) Kalman filter formulation. Thus, the Kalman filter estimate of the BEM coefficients can be recursively computed during each observation window according to the following set of equations:

$$\begin{aligned}
&\mathbf{M}^F[n] = \mathbf{\Phi}[n]\mathbf{M}^A[n-1]\mathbf{\Phi}^H[n] + \mathbf{Q}[n] \\
&\mathbf{K}[n] = \mathbf{M}^F[n]\mathbf{A}^H[n]\left(\mathbf{R}_{\tilde{n}} + \mathbf{A}[n]\mathbf{M}^F[n]\mathbf{A}^H[n]\right)^{-1} \\
&\hat{\mathbf{c}}[n] = \mathbf{\Phi}[n]\hat{\mathbf{c}}[n-1] + \mathbf{K}[n]\left(\mathbf{y}^{\{\mathscr{O}[n]\}}[n] - \mathbf{A}[n]\mathbf{\Phi}[n]\hat{\mathbf{c}}[n-1]\right) \\
&\mathbf{M}^A[n] = \left(\mathbf{I} - \mathbf{K}[n]\mathbf{A}[n]\right)\mathbf{M}^F[n],
\end{aligned} \tag{4.59}$$

where $\mathbf{M}^F[n]$ and $\mathbf{M}^A[n]$ are the *forward* and the *a-posteriori* error covariance matrices, respectively, $\mathbf{K}[n]$ is the *Kalman gain* matrix, and $\mathbf{Q}[n]$ is the covariance matrix of the innovation of the BEM coefficients; their initial conditions for $n = 0$ can be set according to the classical Yule–Walker equations (Kay, 1993). Moreover, together with the evolution matrix $\mathbf{\Phi}[n]$, they can be matched to the channel statistics as specifically proposed in Cannizzaro, Banelli, & Leus (2006) for a VSVO Kalman formulation. A major complexity reduction associated with the matrix inversion in the Kalman gain update equation is obtained by an equivalent vector-state scalar-observation (VSSO) formulation that decouples the estimation problem into the different BEM coefficients, as proposed in Muralidhar & Kwok (2009). Another VSVO Kalman approach, which includes also a Kalman equalizer, has been proposed in He & Tugnait (2007) for single-carrier transmissions. This approach has also been extended to SIMO (He & Tugnait, 2008) and MIMO (Kim & Tugnait, 2008), where in addition a data-aided approach has been considered, as formerly explored in Banelli, Cannizzaro, & Rugini (2007) for OFDM. In order to reduce the complexity, He & Tugnait (2007); He & Tugnait (2008); Tugnait & He (2008); Kim & Tugnait (2008, 2009a,b); Song & Tugnait (2009) ignore the coupling among the evolutions of the BEM coefficients, by approximating the evolution matrix as a diagonal one (i.e., $\mathbf{\Phi}[n] = \phi[n]\mathbf{I}$), and they

simply adapt $\phi[n]$ to the maximum Doppler frequency. Additionally, these single-carrier approaches take advantage of partially overlapped observation windows to improve tracking at the expense of some complexity increase. To further reduce the complexity, other adaptive techniques can be used, such as LMS algorithms, RLS algorithms (Cannizzaro, Banelli, & Leus, 2006; Kim & Tugnait, 2009a), or Wiener LMMSE algorithms (Lindbom, Sternad, & Ahlen, 2001), which usually trade off performance for complexity with respect to the optimum LMMSE Kalman solution.

4.7 CONCLUSIONS

In this chapter, we have discussed pilot-aided channel estimation for time-varying channels. We have focused on a block approach where the channel is modeled as a BEM within the block. Channel estimation has been studied for single-carrier as well as OFDM systems. The channel is always estimated in the time domain, but the pilot symbols are inserted in the time domain for single-carrier systems and in the frequency domain for OFDM systems. Different channel estimators have been studied and the number of observation samples taken into account for channel estimation has been optimized. We have first looked at the case where a single block of data is considered, and we have subsequently extended our discussion to the case where multiple blocks of data are used. The latter leads to less restrictive channel identifiability conditions, and can allow for a higher bandwidth efficiency. We have also considered extensions to multiple antenna systems and briefly discussed adaptive channel estimation.

References

Athaudage, C. R. N., & Jayalath, A. D. S. (2004). Enhanced MMSE channel estimation using timing error statistics for wireless OFDM systems. *IEEE Transactions on Broadcasting*, *50*(4), 369–376.

Banelli, P., Cannizzaro, R. C., & Rugini, L. (2007). Data-aided Kalman tracking for channel estimation in Doppler-affected OFDM systems. In *Proceedings of IEEE ICASSP-07* (pp. 133–136). Honolulu, HI.

Borah, D. K., & Hart, B. D. (1999a). A robust receiver structure for time-varying, frequency-flat Rayleigh fading channels. *IEEE Transactions on Communications*, *47*(3), 862–873.

Borah, D. K., & Hart, B. D. (1999b). Frequency-selective fading channel estimation with a polynomial time-varying channel model. *IEEE Transactions on Communications*, *47*(6), 862–873.

Cai, X., & Giannakis, G. B. (2003). Bounding performance and suppressing intercarrier interference in wireless mobile OFDM. *IEEE Transactions on Communications*, *51*(12), 2047–2056.

Cannizzaro, R. C., Banelli, P., & Leus, G. (2006, July). Adaptive channel estimation for OFDM systems with Doppler spread. In *Proceedings of IEEE SPAWC-06*, Cannes, France.

Choi, J.-W., & Lee, Y.-H. (2005). Optimum pilot pattern for channel estimation in OFDM systems. *IEEE Transactions on Wireless Communications*, *4*(5), 2083–2088.

Choi, Y.-S., Voltz, P. J., & Cassara, F. A. (2001). On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels. *IEEE Transactions on Communications*, *49*(8), 1375–1387.

Cirpan, H. A., & Tsatsanis, M. K. (1999). Maximum likelihood blind channel estimation in the presence of Doppler shifts. *IEEE Transactions on Signal Processing*, *47*(6), 1559–1569.

Coleri, S., Ergen, M., Puri, A., & Bahai, A. (2002). Channel estimation techniques based on pilot arrangement in OFDM systems. *IEEE Transactions on Broadcasting*, *48*(3), 223–229.

Cui, T., Tellambura, C., & Wu, Y. (2005, May). Low-complexity pilot-aided channel estimation for OFDM systems over doubly-selective channels. In *Proceedings of IEEE ICC-05*, Vol. 3 (pp. 1980–1984). Seoul, Korea.

Dai, X. (2007). Optimal training design for linearly time-varying MIMO/OFDM channels modelled by a complex exponential basis expansion. *IET Communications*, *1*(5), 945–953.

Falconer, D., Ariyavisitakul, S. L., Benyamin-Seeyar, A., & Eidson, B. (2002). Frequency domain equalization for single-carrier broadband wireless systems. *IEEE Communications Magazine*, *40*(9), 58–66.

Ghogho, M., & Swami, A. (2004, July). Improved channel estimation using superimposed training. In *Proceedings of IEEE SPAWC-04* (pp. 110–114). Lisbon, Portugal.

Ghogho, M., & Swami, A. (2006, September). Estimation of doubly-selective channels in block transmissions using data-dependent superimposed training. In *Proceedings of EUSIPCO-06*, Antalya, Turkey.

Gorokhov, A., & Linnartz, J.-P. (2004). Robust OFDM receivers for dispersive time-varying channels: equalization and channel acquisition. *IEEE Transactions on Communications*, *52*(4), 572–583.

Guillaud, M., & Slock, D. T. M. (2003, April). Channel modeling and associated intercarrier interference equalization for OFDM systems with high Doppler spread. In *Proceedings of IEEE ICASSP-03*, Vol. IV (pp. 237–240). Hong Kong.

Haykin, S. (1996). *Adaptive filter theory*. Englewood Cliffs: Prentice-Hall.

He, S., & Tugnait, J. (2007, April). Doubly-selective multiuser channel estimation using superimposed training and discrete prolate spheroidal basis expansion models. In *Proceedings of IEEE ICASSP-07*, Honolulu, HI.

He, S., & Tugnait, J. K. (2007). Doubly-selective channel estimation using exponential basis models and subblock tracking. In *Proceedings of IEEE GC-07* (pp. 2847–2851). Washington, DC.

He, S., & Tugnait, J. K. (2008). Decision-directed tracking of doubly-selective channels using exponential basis models. In *Proceedings of IEEE ICC-08* (pp. 5098–5102). Beijing, China.

Horn, R. A., & Johnson, C. R. (1999). *Matrix analysis*. Cambridge University Press, Cambridge, UK.

Jakes, W. C. (1974). *Microwave mobile channels*. New York: Wiley.

Kannu, A. P., & Schniter, P. (2005, March). MSE-optimal training for linear time-varying channels. In *Proceedings of IEEE ICASSP-05*, Vol. III (pp. 789–792). Philadelphia, PA.

Kannu, A. P., & Schniter, P. (2006, March). Minimum mean-squared error pilot-aided transmission for MIMO doubly selective channels. In *Proceedings of CISS-06* (pp. 134–139). Princeton, NJ.

Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River, NJ: Prentice Hall.

Kim, H., & Tugnait, J. K. (2008). Doubly-selective MIMO channels estimation using exponential basis models and subblock tracking. In *Proceedings of IEEE SPAWC-08* (pp. 1258–1261). Recife, Brasil.

Kim, H., & Tugnait, J. K. (2009a). Recursive least-squares decision-directed tracking of doubly-selective channels using exponential basis models. In *Proceedings of IEEE ICASSP-09* (pp. 2689–2693). Taipei, Taiwan.

Kim, H., & Tugnait, J. K. (2009b). Turbo equalization of doubly-selective MIMO fading channels using exponential basis models. In *Proceedings of IEEE SPAWC-09* (pp. 21–24). Perugia, Italy.

Leus, G. (2004, September). On the estimation of rapidly time-varying channels. In *Proceedings of EUSIPCO-04* (pp. 2227–2230). Vienna, Austria.

Leus, G., & Moonen, M. (2003, June). Deterministic subspace based blind channel estimation for doubly-selective channels. In *Proceedings of IEEE SPAWC-2003* (pp. 210–214). Rome, Italy.

Leus, G., & van der Veen, A.-J. (2005). Channel estimation, in *Smart antennas – state of the art*. Ed. T. Kaiser, A. Bourdoux, H. Boche, J. R. Fonollosa, J. B. Andersen and W. Utschick, Hindawi, New York, USA.

Lindbom, L., Sternad, M., & Ahlen, A. (2001). Tracking of time-varying mobile radio channels–part I: The Wiener LMS algorithm. *IEEE Transactions on Communications*, *49*(12), 2207–2217.

Li, Y., Cimini, L. J., & Sollenberger, N. R. (1998). Robust channel estimation for OFDM systems with rapid dispersive fading channels. *IEEE Transactions on Communications*, *46*(7), 1146–1162.

Ma, X., & Giannakis, G. B. (2003). Maximum-diversity transmissions over doubly selective wireless channels. *IEEE Transactions on Information Theory*, *49*(7), 1832–1840.

Ma, X., Giannakis, G., & Ohno, S. (2003). Optimal training for block transmissions over doubly-selective fading channels. *IEEE Transactions on Signal Processing*, *51*(5), 1351–1366.

Mostofi, Y., & Cox, D. C. (2005). ICI mitigation for pilot-aided OFDM mobile systems. *IEEE Transactions on Wireless Communications*, *4*(2), 765–774.

Muralidhar, K., & Kwok, H. L. (2009). A low-complexity Kalman approach for channel estimation in doubly-selective OFDM systems. *IEEE Signal Processing Letters*, *16*(7), 632–635.

Negi, R., & Cioffi, J. (1998). Pilot tone selection for channel estimation in a mobile OFDM system. *IEEE Transactions on Consumer Electronics*, *44*(3), 1122–1128.

Nicoli, M., Simeone, O., & Spagnolini, U. (2003). Multislot estimation of frequency-selective fast-varying channels. *IEEE Transactions on Communications*, *51*(8), 1337–1347.

Rousseaux, O., Leus, G., Stoica, P., & Moonen, M. (2004). Gaussian maximum likelihood channel estimation with short training. *IEEE Transactions on Wireless Communications*, *4*(6), 2945–2955.

Rousseaux, O., Leus, G., & Moonen, M. (2006). Estimation and equalization of doubly-selective channels using known symbol padding. *IEEE Transactions on Signal Processing*, *54*(3), 979–990.

Sanzi, F., Jelting, S., & Speidel J. (2003). A comparative study of iterative channel estimation for mobile OFDM systems. *IEEE Transactions on Wireless Communications*, *2*(5), 849–859.

Scaglione, A., Giannakis, G. B., & Barbarossa, S. (1999a). Redundant filterbank precoders and equalizers – part I: Unification and optimal designs. *IEEE Transactions on Signal Processing*, *47*(7), 1988–2006.

Scaglione, A., Giannakis, G. B., & Barbarossa, S. (1999b). Redundant filterbank precoders and equalizers – part II: Blind channel estimation, synchronization and direct equalization. *IEEE Transactions on Signal Processing*, *47*(7), 2007–2022.

Schafhuber, D., & Matz, G. (2005). MMSE and adaptive prediction of time-varying channels for OFDM systems. *IEEE Transactions on Wireless Communications*, *4*(2), 593–602.

Schafhuber, D., Matz, G., & Hlawatsch, F. (2003, November). Kalman tracking of time-varying channels in wireless MIMO-OFDM systems. In *Proceedings of Asilomar-03* (pp. 1261–1265). Pacific Grove, CA.

Schniter, P. (2006, March). On doubly dispersive channel estimation for pilot-aided pulse-shaped multicarrier modulation. In *Proceedings of CISS-06* (pp. 1296–1301). Princeton, NJ.

Schniter, P., & Liu, H. (2003, November). Iterative equalization for single-carrier cyclic-prefix in doubly dispersive channels. In *Proceedings of Asilomar-03* (pp. 502–506). Pacific Grove, CA.

Song, L., & Tugnait, J. K. (2009). Doubly-selective fading channels equalization: A comparison of the Kalman filter approach with the exponential basis expansion model-based equalizers. *IEEE Transactions on Wireless Communications*, *1*(8), 60–64.

Stamoulis, A., Diggavi, S. N., & Al-Dhahir, N. (2002). Intercarrier interference in MIMO OFDM. *IEEE Transactions on Signal Processing*, *50*(10), 2451–2464.

Tang, Z. (2007). *OFDM transmission over rapidly changing channels*. Ph.D Disseration, Delft University of Technology, The Netherlands.

Tang, Z., & Leus, G. (2008). A novel receiver architecture for single-carrier transmission over time-varying channels. *IEEE Journal on Selected Areas in Communications*, *26*(2), 366–377.

Tang, Z., Leus, G., & Banelli, P. (2006, June). Pilot-assisted time-varying OFDM channel estimation based on multiple OFDM symbols. In *Proceedings of IEEE SPAWC-06*, Cannes, France.

Tang, Z., Cannizzaro, R. C., Banelli, P., & Leus, G. (2007). Pilot-assisted time-varying channel estimation for OFDM systems. *IEEE Transactions on Signal Processing*, *55*(5), 2226–2238.

Teo, K. D., & Ohno, S. (2005). Optimal MMSE finite parameter model for doubly-selective channels. In *Proceedings of IEEE GLOBECOM-05*, Vol. 6 (pp. 3503–3507). St. Louis, MO.

Thomas, T. A., & Vook, F. W. (2000, March). Multi-user frequency-domain channel identification, interference suppression, and equalization for time-varying broadband wireless communications. In *Proceedings of IEEE SAM-00* (pp. 444–448). Cambridge, MA.

Tomasin, S., Gorokhov, A., Yang, H., & Linnartz, J.-P. (2005). Iterative interference cancellation and channel estimation for mobile OFDM. *IEEE Transactions on Wireless Communications*, *4*(1), 238–245.

Tong, L., Sadler, B. M., & Dong, M. (2004). Pilot-assisted wireless transmissions: General model, design criteria, and signal processing. *IEEE Signal Processing Magazine*, *21*(6), 12–25.

Tsatsanis, M. K., & Giannakis, G. B. (1996). Modeling and equalization of rapidly fading channels. *International Journal of Adaptive Control and Signal Processing*, *10*(2/3), 159–176.

Tugnait, J. K., & He, S. (2008). Recursive least-squares doubly-selective channel estimation using exponential basis models and subblock-wise tracking. In *Proceedings of IEEE ICASSP-08* (pp. 2861–2864). Las Vegas, NV.

Visintin, M. (1996). Karhunen-Loève expansion of a fast Rayleigh fading process. *IET Electronics Letters*, *32*(18), 1712–1713.

Yang, L., Ma, X., & Giannakis, G. B. (2004, May). Optimal training for MIMO fading channels with time- and frequency-selectivity. In *Proceedings of IEEE ICASSP-04*, Vol. III (pp. 821–824). Montreal, Canada.

Yang, B., Letaief, K. B., Cheng, R. S., & Cao, Z. (2001). Channel estimation for OFDM transmission in multi-path fading channels based on parametric channel modeling. *IEEE Transactions on Communications*, *49*(3), 467–479.

Yip, K.-W., & Ng, T.-S. (1997). Karhunen-Loève expansion of the WSSUS channel output and its application to efficient simulation. *IEEE Journal on Selected Areas in Communications*, *15*(4), 640–646.

Zakharov, Y. V., Tozer, T. C., & Adlard, J. F. (2004). Polynomial spline-approximation of Clarke's model. *IEEE Transactions on Signal Processing*, *52*(5), 1198–1208.

Zemen, T., & Mecklenbräuker, C. F. (2005). Time-variant channel estimation using discrete prolate spheroidal sequences. *IEEE Transactions on Signal Processing*, *53*(9), 3597–3607.

Zheng, Y. R., & Xiao, C. (2003). Simulation models with correct statistical properties for Rayleigh fading channels. *IEEE Transactions on Communications*, *51*(6), 920–928.