6

Equalization of Time-Varying Channels

Philip Schniter, Sung-Jun Hwang, Sibasish Das, and Arun P. Kannu

The Ohio State University

6.1 Introduction

As discussed in Chapter 1, the wireless communication channel can be modeled as a time-varying (TV) linear[1] system whose output is corrupted by additive noise. To reliably recover the transmitted information from the channel output, the receiver must address the effects of both linear distortion and additive noise. While, in theory, the mitigation of linear distortion and additive noise should be done jointly, in practice the task is often partitioned into two tasks, equalization and decoding, in order to reduce implementation complexity.

Roughly speaking, *equalization* leverages knowledge of channel structure to mitigate the effects of the linear distortion while *decoding* leverages knowledge of code structure to mitigate the channel's additive noise component. The equalizer might be well informed about the channel (e.g., knowing the complete channel impulse response) or relatively uninformed (e.g., knowing only the maximum channel length). In some cases, knowledge of symbol structure (e.g., the symbol alphabet or, if applicable, the fact that the symbols have a constant modulus) is assumed to be in the domain of the equalizer, while in other cases it is assumed to be in the domain of the decoder; since the equalizer and decoder work together to infer the transmitted information from the channel output, the role of equalization versus decoding is somewhat a matter of definition. For this chapter, however, we assume that exploitation of code structure is *not* in the domain of the equalizer.

Generally speaking, the output of the equalizer is a sequence of symbol (or bit) estimates which have been, to the best of the equalizer's ability, freed of channel corruption. These estimates are then passed to the decoder for further refinement and final decision making. In so called *turbo equalization* schemes [DJB+95, KST04], the decoder passes

[1] Some channels are better modeled as nonlinear, but such channels are not the focus of this book.

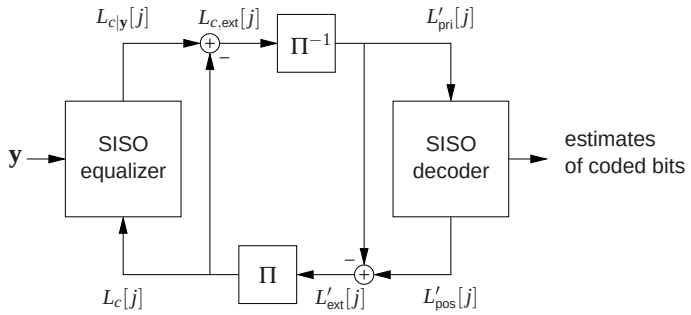

Figure 6.1 *Soft-input soft-output (SISO) equalizer connected with a SISO decoder in a turbo configuration. Note the presence of de-interleaver* Π^{-1} *fed by the extrinsic equalizer LLRs* $L_{c,\text{ext}}[j] = L_{c|\mathbf{y}}[j] - L_c[j]$*, and interleaver* Π *fed by the extrinsic decoder LLRs* $L'_{\text{ext}}[j] = L'_{\text{pos}}[j] - L'_{\text{pri}}[j]$.

refined soft bit estimates back to the equalizer for further refinement, and the equalizer passes further refined soft bit estimates to the decoder. The process is then iterated until the equalizer and decoder "agree" on the soft bit estimates. Note that the use of soft bit estimates implies that the equalizer treats the bits as (a priori) independent. Turbo equalization is illustrated in Fig. 6.1 and will be discussed in more detail later.

The inputs to an equalizer depend on its design. So-called *coherent equalizers* are assumed to know the parameters describing the *state* of the TV linear system that they are trying to mitigate, or an estimate thereof. Typical examples of channel state parameters include impulse response coefficients or inter-carrier interference coefficients. Coherent equalization requires the simultaneous operation of a *channel estimator*, whose main purpose is to provide accurate and up-to-date estimates of the TV channel state to the equalizer. Channel estimation is discussed in Chapter 4. The idea to separate channel estimation from equalization can be traced back to early work by Kailath [Kai60].

So-called *noncoherent equalizers* operate without explicit knowledge of the channel state, and therefore are not dependent on the implementation of a channel estimator. Noncoherent equalizers, however, are sometimes assumed to know the *channel statistics* (e.g., the scattering function) or an estimate thereof. In the case of a non-stationary channel, the statistics themselves would need to be tracked. In their most general form, noncoherent equalizers treat the channel parameters as "nuisance parameters" that complicate data estimation. In some cases they explicitly estimate the channel state parameters in conjunction with the data (i.e., *joint channel/symbol estimation*), while in other cases they compute data estimates without ever computing a channel estimate.

The equalization of rapidly TV communication channels is much more challenging than the equalization of their time-invariant (or slowly TV) counterparts. This can be understood intuitively as follows. From the perspective of coherent equalization, a rapidly TV channel implies that the channel state is constantly changing, which implies that the equalizer must be constantly redesigned in order to stay well matched to the channel. From the perspective of noncoherent equalization, a rapidly TV channel has more degrees

of freedom (over a given bandwidth and signaling epoch) than a slowly varying channel, and thus more nuisance parameters to contend with.

Beyond these intuitive considerations, there is another important reason why rapidly TV channels are more difficult to equalize than slowly TV ones. For time-invariant linear channels, information can be split up and transmitted in parallel on non-interfering subcarriers. In this case, equalization becomes a simple matter of adjusting the gain and phase on each received subcarrier. This is, in fact, the main idea behind multi-carrier modulation schemes like *orthogonal frequency division multiplexing* (OFDM) [Cim85]. For slowly TV channels, the same approach can be easily extended: to mimic a time-invariant channel, the OFDM symbol duration can be chosen shorter than the channel's coherence time. But, as now explained, such an approach turns out to be impractical for rapidly TV channels. To prevent interference between adjacent OFDM symbols, guard intervals are typically inserted. For time-invariant or slowly TV channels, the loss in spectral efficiency due to the inclusion of these guards can be made small, since the channel delay spread (and hence the guard interval) is much smaller than the channel coherence time (and hence the OFDM symbol length). For rapidly TV channels, the OFDM symbol length would need to be made extremely short, at which point the loss of spectral efficiency due to guard insertion would be severe. If one tried to optimize the modulation strategy, one would find that it is in fact impossible to prevent interference among the subcarriers without significant compromise in spectral efficiency [SB03]— a consequence of the Balian-Low theorem [Dau92]. To summarize: while the equalization of slowly TV channels can be trivialized via suitable choice of the transmission scheme, the equalization of rapidly TV channels cannot.

The remainder of this chapter will be organized as follows. In Section 6.2, we outline the system model assumed throughout the chapter and detail the essential features that result from rapid channel time-variation. In Section 6.3, we describe coherent approaches to equalization of rapidly TV channels and, in Section 6.4, we describe noncoherent approaches. In Section 6.5, we conclude.

6.2 System model

We now outline the system model used in the remainder of the chapter. In this chapter, we focus on systems which use a single transmitter antenna and a single receiver antenna; multi-antenna systems will be discussed in Chapter 8.

6.2.1 Basic assumptions

As discussed in Chapter 1, the time-domain received sample $r[n]$ can be written in terms of the transmitted sequence $(s[n])_{n\in\mathbb{Z}}$, the TV time-n length-M impulse response $(h[n,m])_{m=0}^{M-1}$,

and additive white Gaussian noise process $(w[n])_{n\in\mathbb{Z}}$ of variance σ_w^2 as follows:

$$r[n] \quad = \quad \sum_{m=0}^{M-1} h[n,m]s[n-m] + w[n]. \tag{6.1}$$

In this chapter, we assume that the transmitted sequence $(s[n])_{n\in\mathbb{Z}}$ is generated from the finite-alphabet symbol sequence $(a[k])_{k\in\mathbb{Z}}$ using a generic *finite-memory linear modulation* scheme, and that the demodulated sequence $(y[k])_{k\in\mathbb{Z}}$ is generated from the received sequence $(r[n])_{n\in\mathbb{Z}}$ using a corresponding *finite-memory linear demodulation* scheme. Prior to modulation, the symbol sequence $(a[k])_{k\in\mathbb{Z}}$ is mapped from a coded-bit sequence $(c[j])_{j\in\mathbb{Z}}$ which is generated from an information-bit sequence $(b[i])_{i\in\mathbb{Z}}$ via rate-R_c coding and interleaving. We denote the symbol alphabet by $\mathscr{A}$, its cardinality by $|\mathscr{A}|$, and the set of admissible symbol sequences (as allowed by coding/interleaving) by $\boldsymbol{\mathscr{A}}$.

For ease of notation, we find it convenient to assume *block transmission* with block length K, where the symbols

$$\mathbf{a} \triangleq \begin{pmatrix} a[0] & a[1] & \cdots & a[K-1] \end{pmatrix}^T \in \mathscr{A}^K$$

can be related to the demodulated channel outputs

$$\mathbf{y} \triangleq \begin{pmatrix} y[0] & y[1] & \cdots & y[K-1] \end{pmatrix}^T \in \mathbb{C}^K$$

through the matrix/vector equation

$$\mathbf{y} \quad = \quad \underbrace{\boldsymbol{\Gamma}\mathbf{H}\mathbf{G}}_{\triangleq\,\mathbf{Q}}\mathbf{a} + \mathbf{z}. \tag{6.2}$$

We note, however, that the block length K can be arbitrarily large and that the receiver might not be able to store/process the entire vector $\mathbf{y}$. In (6.2), $\boldsymbol{\Gamma}$, $\mathbf{H}$, and $\mathbf{G}$ are matrix representations of the linear demodulation operator, the linear TV channel, and the linear modulation operator, respectively, and

$$\mathbf{z} \triangleq \begin{pmatrix} z[0] & z[1] & \cdots & z[K-1] \end{pmatrix}^T \in \mathbb{C}^K$$

represents the noise after demodulation. We note that the *effective channel matrix* $\mathbf{Q} = \boldsymbol{\Gamma}\mathbf{H}\mathbf{G} \in \mathbb{C}^{K\times K}$ represents the combined effects of modulation, channel propagation, and demodulation, and will be used extensively throughout the chapter. Finally, we collect, in the vector $\mathbf{c}$, the $K\log_2|\mathscr{A}|$ coded bits that determine the K symbols in $\mathbf{a}$. Note that, with a block length of K, we have $\boldsymbol{\mathscr{A}} \subset \mathscr{A}^K$.

In writing (6.2), we have assumed that the K demodulated samples in $\mathbf{y}$ are sufficient for equalization/decoding of the K symbols in $\mathbf{a}$ (i.e., that $(y[k])_{k<0}$ and $(y[k])_{k\geq K}$ can be ignored), and that inter-block interference (IBI) is negligible. These assumptions will be satisfied for any well-designed block transmission scheme. Furthermore, we will assume that the noise $\mathbf{z}$, the symbols $\mathbf{a}$, and the effective channel $\mathbf{Q}$ are mutually independent, and that (unless otherwise noted) the symbols $\mathbf{a}$ are zero-mean (i.e., $\boldsymbol{\mu}_a = \mathbf{0}$) and white (i.e.,[2]

[2]Throughout the chapter, we use subscripted versions of $\mathbf{C}$ to denote covariance matrices.

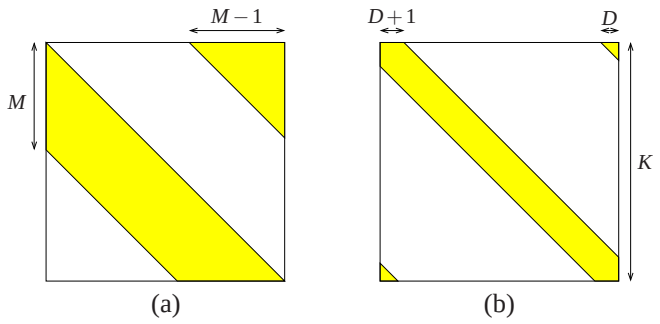

Figure 6.2 *Support region of (a) "widely quasi-banded" and (b) "narrowly quasi-banded" matrices. While M is often large (e.g., in the hundreds), D is usually very small (e.g.,* 1 *or* 2*).*

$\mathbf{C_a} = \sigma_a^2\mathbf{I}$). Finally, it should be noted that the demodulated noise $\mathbf{z}$ is *not* assumed to be white unless otherwise noted; although $(w[n])_{n\in\mathbb{Z}}$ is white, the demodulation process does not necessarily guarantee white $(z[k])_{k\in\mathbb{Z}}$.

Throughout the chapter, we assume that the equalizer knows the symbol alphabet $\mathscr{A}$ but not the code structure, i.e., $\mathscr{A}$. Thus, the topic of *joint* equalization/decoding lies outside the scope of this chapter. Turbo equalization, where separate equalization and decoding steps are iterated (as illustrated in Fig. 6.1) will, however, be discussed.

6.2.2 The structure of the effective channel matrix $\mathbf{Q}$

In block equalization, if it can be assumed that certain coefficients of $\mathbf{Q}$ will be negligible for nearly all realizations of $\mathbf{Q}$, then it is reasonable to conclude that an equalizer which ignores these coefficients will perform nearly as good as an equalizer which incorporates these coefficients. However, the equalizer which ignores these coefficients may be significantly cheaper to implement, especially if the proportion of negligible coefficients is large. This is, in fact, the guiding principle behind the design of practical equalization algorithms for rapidly TV channels.

Based on the characteristics of rapidly TV channels and commonly used modulation/demodulation schemes, we partition effective channel matrices $\mathbf{Q}$ into three classes based on the support region of non-negligible coefficients within the matrix: i) *widely quasi-banded*, ii) *narrowly quasi-banded*, and iii) *fully populated* matrices. The support regions of widely quasi-banded and narrowly quasi-banded matrices are defined in Fig. 6.2, and illustrative examples of $\mathbf{Q}$ based on a randomly generated channel impulse response and several modulation/demodulation schemes are given in Fig. 6.3 (the construction of which will be detailed below). Note that we use the term "quasi-banded" as opposed to "banded" due to the corner[3] support regions in Fig. 6.2. Banded matrices, like that illustrated in Fig. 6.5(b), will also be discussed in the sequel.

[3]Note that the one-corner support of the widely quasi-banded matrix in Fig. 6.2(a) can be transformed into the two-corner support of the narrowly quasi-banded matrix in Fig. 6.2(b) by simply rotating the columns of the former matrix right by $M/2$ places. Thus, the essential difference between these matrices is really the width of the support region (i.e., M versus $2D+1$).

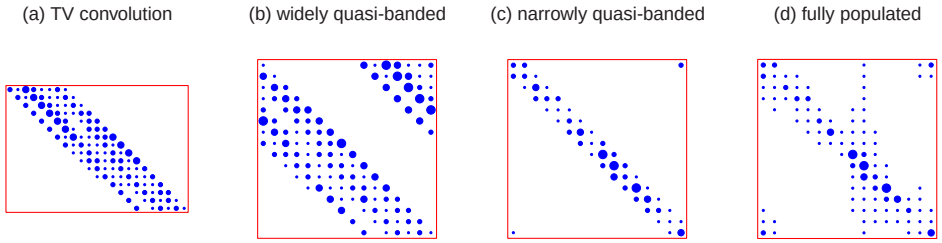

Figure 6.3 *Example of (a) a TV-channel's propagation matrix and the corresponding effective channel matrices that result from (b) CP-SCM, (c) CP-OFDM with max-SINR receiver windowing [Sch04], and (d) CP-OFDM with rectangular receiver windowing. The dot size is proportional to the coefficient magnitude.*

To understand how these patterns manifest in $\mathbf{Q} = \mathbf{\Gamma H G}$, we must consider the composite effect of linear modulation $\mathbf{G}$, propagation through the TV linear channel $\mathbf{H}$, and demodulation $\mathbf{\Gamma}$. As implied by (6.1), the channel propagation matrix $\mathbf{H}$ is a TV convolution matrix whose n^{th} row contains the impulse response coefficients $(h[n,m])_{m=0}^{M-1}$. For example, Fig. 6.3(a) shows a TV channel propagation matrix for $M = 8$ that was randomly generated according to the WSSUS Jakes [Stü01] fading assumption with $\nu_{\max} T_s = 0.03$, where $\nu_{\max}$ denotes the maximum (single-sided) Doppler spread in Hz and T_s the channel-use interval (i.e., the symbol period in a single-carrier system) in seconds. If the channel was time-invariant, the propagation matrix would have a Toeplitz structure. But here, since the channel is rapidly TV, each coefficient's magnitude varies smoothly along its diagonal of the propagation matrix. Given the construction of $\mathbf{H}$, the characteristics of $\mathbf{Q}$ will depend on the choices of $\mathbf{G}$ and $\mathbf{\Gamma}$ and their interaction with $\mathbf{H}$, as discussed next.

Single-carrier modulation/demodulation

For single-carrier modulation/demodulation schemes, $\mathbf{G}$ and $\mathbf{\Gamma}$ accomplish little more than insertion and removal of a guard interval (of length $N_g \geq M-1$). In this case, $\mathbf{Q}$ is created from the propagation matrix $\mathbf{H}$ by simply cutting the first N_g columns of $\mathbf{H}$ out and superimposing them onto the last N_g columns of $\mathbf{H}$. This operation was used, e.g., to create the widely quasi-banded matrix in Fig. 6.3(b) from the TV convolution matrix in Fig. 6.3(a). More precisely, when $\mathbf{H}$ has dimensions $K \times (K+M-1)$, cyclic-prefixed single carrier modulation (CP-SCM) [FABSE02] uses

$$\mathbf{G} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{M-1} \\ \mathbf{I}_{K-M+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-1} \end{pmatrix} \quad \text{and} \quad \mathbf{\Gamma} = \mathbf{I}_K, \tag{6.3}$$

whereas zero-padded single carrier modulation (ZP-SCM) [WMG04] uses a slightly different construction of $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{\Gamma}$ that results in an equivalent $\mathbf{Q}$ matrix. We consider the effective channel matrix generated from SCM to be *widely quasi-banded* because M, the width of the non-negligible band in $\mathbf{Q}$, is typically large: since $M \triangleq \lceil \tau_{\max}/T_s \rceil$ is the dis-

crete delay spread of the channel, it is not unusual for M to be in the hundreds (e.g., delay spread $\tau_{\max} = 20\mu$s and bandwidth $1/T_s = 10$ MHz yield $M = 200$). Though small-M applications do exist, they yield equalization problems that are not very challenging, and hence not very interesting, especially in the coherent setting. Hence, we focus on the case of large M.

Time-frequency concentrated modulation/demodulation

The effect, on the transmitted signal $\{s[n]\}$, of propagation through the linear TV channel $\{h[n,m]\}$ can be understood as *simultaneous delay and Doppler spreading*. Thus, if each symbol $a[k]$ is modulated on a time-frequency concentrated waveform $\mathbf{g}_k \triangleq \begin{pmatrix} g_k[0] & \cdots & g_k[N-1] \end{pmatrix}^T$ for suitable[4] N, so that

$$s[n] \quad = \quad \sum_{k=0}^{K-1} a[k] g_k[n] \quad \text{for} \quad n = 0, \ldots, N-M, \tag{6.4}$$

where $\mathbf{g}_k$ is sufficiently "isolated" from the other waveforms $\{\mathbf{g}_{k'}\}_{k' \neq k}$ in the time-frequency domain, then propagation through the delay/Doppler spreading channel should cause only mild interference between these $\{a[k]\}$. Extraction of the k^{th} symbol's contribution from the received signal $\{r[n]\}$ would then be accomplished via the linear demodulation operation

$$y[k] \quad = \quad \sum_{n=0}^{N-1} r[n] \gamma_k^*[n] \quad \text{for} \quad k = 0, \ldots, K-1, \tag{6.5}$$

for $\boldsymbol{\gamma}_k = \begin{pmatrix} \gamma_k[0] & \cdots & \gamma_k[N-1] \end{pmatrix}^T$ concentrated at the same time and frequency as $\mathbf{g}_k$. This is the main idea behind pulse-shaped multicarrier schemes like [LAB95, HB97, MK97, KM98, Böl02, SB03, Sch04, RBL06, DS07, MSG+07] as well as Slepian schemes like [SAMT05].

With suitably designed modulation/demodulation waveforms $\{\mathbf{g}_k\}$ and $\{\boldsymbol{\gamma}_k\}$, the combined channel matrix $\mathbf{Q}$ under (6.4)-(6.5) can be ensured to have the *narrowly quasi-banded structure* illustrated in Fig. 6.2(b). There, D can be interpreted as the (single-sided) discrete Doppler spread of the effective channel and $2D+1$ can be recognized as the width of the non-negligible interference band. Typically D is chosen as

$$D \quad = \quad \lceil \nu_{\max} T_s K + D_0 \rceil, \tag{6.6}$$

where D_0 is a small non-negative constant (e.g., $0 \leq D_0 \leq 2$ for a well-designed modulation/demodulation scheme), as discussed in the sequel. We can see that $\mathbf{Q}$ will be *narrowly quasi-banded*, so that $2D+1 \ll M$, by plugging the typical block-length choice of $K = 4M$ into (6.6) and then using the definition $M = \tau_{\max}/T_s$ to see that [HS06]

$$\begin{aligned} D \quad &\leq \quad \lceil 4\nu_{\max}\tau_{\max} \rceil + \lceil D_0 \rceil & (6.7) \\ &= \quad 1 + \lceil D_0 \rceil \quad \text{when} \quad 0 < 2\nu_{\max}\tau_{\max} \leq 0.5. & (6.8) \end{aligned}$$

[4] If N exceeds the time period between consecutive block transmissions, then inter-block interference (IBI) can result. In this case, the model (6.2) can be generalized to $\mathbf{y} = \mathbf{Q}\mathbf{a} + \mathbf{Q}_{\mathsf{pre}}\mathbf{a}_{\mathsf{pre}} + \mathbf{Q}_{\mathsf{pst}}\mathbf{a}_{\mathsf{pst}} + \mathbf{z}$, where $\mathbf{Q}_{\mathsf{pre}}\mathbf{a}_{\mathsf{pre}}$ accounts for pre-cursor IBI and $\mathbf{Q}_{\mathsf{pst}}\mathbf{a}_{\mathsf{pst}}$ accounts for post-cursor IBI. The IBI can be made negligible, however, with suitable design of modulation/demodulation pulses $\{\mathbf{g}_k\}$ and $\{\boldsymbol{\gamma}_k\}$.

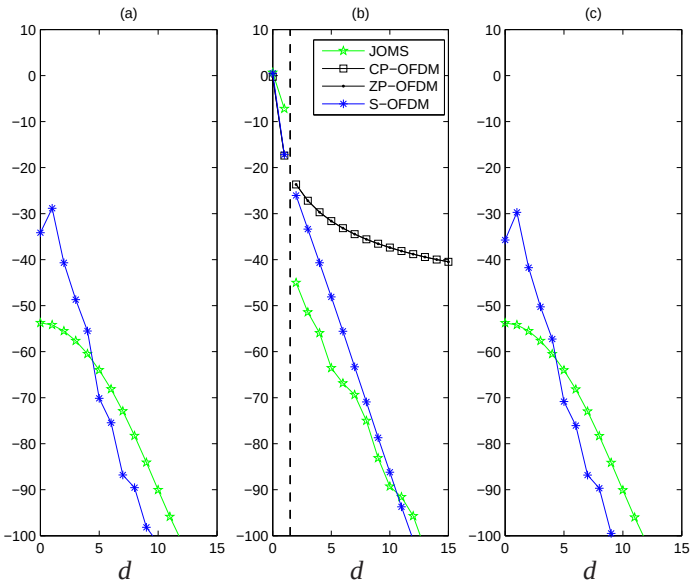

Figure 6.4 *For the IBI model* $\mathbf{y} = \mathbf{Q}\mathbf{a} + \mathbf{Q}_{\mathsf{pre}}\mathbf{a}_{\mathsf{pre}} + \mathbf{Q}_{\mathsf{pst}}\mathbf{a}_{\mathsf{pst}} + \mathbf{z}$*, with* $K = 64$*,* $N_g = 15$*,* $M = 16$*,* $\nu_{\max} T_s = 0.003$*, and WSSUS Jakes fading (see Chapter 1), subplot (b) shows the mean-square value (in dB) of a coefficient in* $\mathbf{Q}$ *versus its distance "d" from the main diagonal of* $\mathbf{Q}$*, while subplots (a) and (c) show the same for the coefficients in* $\mathbf{Q}_{\mathsf{pre}}$ *and* $\mathbf{Q}_{\mathsf{pst}}$*, respectively. The dashed vertical line indicates* $D = \lceil \nu_{\max} T_s K \rceil$*.* JOMS *refers to Das and Schniter's joint transmitter/receiver optimization max-SINR scheme [DS07] while* S-OFDM *refers to Strohmer and Beaver's orthogonal scheme [SB03].*

The quantity $2\nu_{\max}\tau_{\max}$, sometimes referred to as the "spreading index," describes the total severity of delay-Doppler spreading. The boundary between *underspread* and *overspread* channels occurs at $2\nu_{\max}\tau_{\max} = 1$, and it can be safely assumed that $2\nu_{\max}\tau_{\max} \ll 1$ for practical applications. Thus, from (6.8), we conclude that the width of the non-negligible coefficient band is $2D+1 \leq 3+2\lceil D_0 \rceil$ when suitable modulation/demodulation waveforms are used. In summary, $2D+1 \ll M$ is a reasonable claim for the values of M that are of interest in this chapter.

As an example, Fig. 6.3(c) shows $\mathbf{Q}$ constructed via cyclic-prefixed orthogonal frequency division multiplexing (CP-OFDM) [Cim85] with max-SINR receiver pulse-shaping [Sch04] using the TV convolution matrix shown in Fig. 6.3(a). Though the channel has an extremely high spreading index of $2\nu_{\max}\tau_{\max} = 0.8$, all coefficients in $\mathbf{Q}$ outside of the 3-wide band are negligible. As another example, Fig. 6.4 shows $\mathsf{E}\left\{\left|[\mathbf{Q}]_{k,k+d}\right|^2\right\}$ (in dB) versus d for several modulation/demodulation schemes and a channel with a spreading index of 0.1. ($\mathsf{E}\left\{\left|[\mathbf{Q}]_{k,k+d}\right|^2\right\}$ is invariant to k.) As can be seen in Fig. 6.4, the JOMS scheme from [DS07] suppresses coefficients outside the band of radius $D = \lceil \nu_{\max} T_s K \rceil = 1$ by at least 44 dB.

At this point we make one final observation about a narrowly quasi-banded matrix $\mathbf{Q}$. If we upper-triangularize $\mathbf{Q}$, e.g., via the QR decomposition $\mathbf{Q} = \mathbf{V}\bar{\mathbf{Q}}$ where $\mathbf{V}$ is unitary and $\bar{\mathbf{Q}}$ is upper triangular, then $\bar{\mathbf{Q}}$ will have the "V-shaped" structure shown in Fig. 6.5(c) on page xiii. Such upper-triangularization of $\mathbf{Q}$ occurs prior to decision feedback and tree-search based equalization, as discussed in Section 6.3.2.

Other modulation/demodulation schemes

When the modulation and demodulation pulses $\{\mathbf{g}_k\}_{k=0}^{K-1}$ and $\{\boldsymbol{\gamma}_k\}_{k=0}^{K-1}$ are *not* designed to curb the effects of delay/Doppler spreading, the support of non-negligible coefficients within $\mathbf{Q}$ can be widespread, to the point where $\mathbf{Q}$ must be considered as *fully populated*. Examples of such modulation/demodulation schemes include the wavelet-based schemes [Wor96, Mar00], the chirp-based schemes [Mar01, BT01, KS08], a scheme designed to maximize a lower bound on capacity [YCL07], and the diversity maximizing schemes [MG03, HS07c].

Even popular multicarrier schemes like CP-OFDM, when used with a rectangular receiver pulse, yield a near-fully populated $\mathbf{Q}$ when the channel is TV rapidly enough. Fig. 6.3 shows this by example: the effective channel matrix in Fig. 6.3(d) was constructed from the TV convolution matrix in Fig. 6.3(a) via standard CP-OFDM. Notice that the non-negligible coefficients in $\mathbf{Q}$ are not all located in the central band of the matrix. Fig. 6.4 shows a similar phenomenon: for CP-OFDM, $\mathsf{E}\{|[\mathbf{Q}]_{k,k+d}|^2\}$ decays very slowly with d, the distance from the main diagonal of $\mathbf{Q}$.

6.3 Coherent equalization

In this section, we focus on *coherent equalization*, i.e., equalization under the assumption that the channel matrix $\mathbf{H}$, and thus the effective channel matrix $\mathbf{Q}$, is known. The noncoherent case will be discussed in Section 6.4.

In Section 6.3.1, we discuss several criteria (i.e., notions of optimality) under which coherent equalizers are designed, and, in Section 6.3.2, we describe classical equalization algorithms for generic $\mathbf{Q}$. Then, in Sections 6.3.3–6.3.4, we focus on coherent equalization techniques for the specific types of $\mathbf{Q}$ anticipated for rapidly TV channels in Section 6.2.2.

6.3.1 Coherent equalization criteria

Referring to (6.2), the goal of coherent block equalization is estimation of the symbol vector $\mathbf{a}$, or the corresponding coded-bit vector $\mathbf{c}$, from the linearly distorted and noisy demodulator output vector $\mathbf{y}$, assuming knowledge of the channel matrix $\mathbf{Q}$ and the noise statistics. Note that this may or may not include a hard-decision or quantization step, as explained later. In any case, we are fundamentally interested in identifying the "opti-

mal" method to generate these symbol estimates. The answer, however, depends on how optimality is defined, i.e., which *equalization criterion* is employed.

In organizing the criteria that are most often used for equalizer design, it helps to consider how the equalizer outputs will be used by the receiver (e.g., by the decoder).

Hard symbol or bit estimates

If there is no decoder or if the decoder wants *hard estimates* of the symbols or bits, then the goal is to produce a finite-alphabet estimate $\hat{\mathbf{a}} \in \mathscr{A}^K$. (Recall that the equalizer is assumed to know the symbol alphabet $\mathscr{A}$ but not the set of coded symbol sequences $\mathscr{A}$.)

Maximum a posteriori (MAP) *sequence detection* (SD) [Poo94] minimizes the probability of sequence error. By definition, the MAPSD estimate is

$$\hat{\mathbf{a}}_{\mathsf{cMAPSD}} \triangleq \arg\max_{\mathbf{a}' \in \mathscr{A}^K} \Pr\{\mathbf{a} = \mathbf{a}' \mid \mathbf{y}, \mathbf{Q}\}. \tag{6.9}$$

In (6.9), we use the notation "cMAPSD" to emphasize that this is the *coherent* version of the MAP criterion applied to *sequence* detection. In contrast, coherent MAP *symbol* and *bit* detection takes the form

$$\hat{a}_{\mathsf{cMAP}}[k] \triangleq \arg\max_{a \in \mathscr{A}} \Pr\{a[k] = a \mid \mathbf{y}, \mathbf{Q}\} \quad \text{for} \quad k \in 0, \ldots, K-1 \tag{6.10}$$

$$\hat{c}_{\mathsf{cMAP}}[j] \triangleq \arg\max_{c \in \{0,1\}} \Pr\{c[j] = c \mid \mathbf{y}, \mathbf{Q}\} \quad \text{for} \quad j \in 0, \ldots, K\log_2|\mathscr{A}| - 1. \tag{6.11}$$

In writing (6.9)-(6.11), we have treated the channel $\mathbf{Q}$ as a random quantity.

If we assume that each of the symbol sequences in $\mathscr{A}^K$ has equal prior probability, i.e., $\Pr\{\mathbf{a} = \mathbf{a}'\} = 1/|\mathscr{A}|^K \ \forall \mathbf{a}' \in \mathscr{A}^K$, then coherent MAPSD reduces to coherent *maximum likelihood* (ML) SD [Poo94]:

$$\hat{\mathbf{a}}_{\mathsf{cMLSD}} \triangleq \arg\max_{\mathbf{a} \in \mathscr{A}^K} f(\mathbf{y} \mid \mathbf{a}, \mathbf{Q}), \tag{6.12}$$

where $f(\mathbf{y} \mid \mathbf{a}, \mathbf{Q})$ denotes the probability density function of $\mathbf{y}$ conditioned on $\mathbf{a}$ and $\mathbf{Q}$, also known as the *likelihood function*. To see this, notice from Bayes rule that

$$\Pr\{\mathbf{a} = \mathbf{a}' \mid \mathbf{y}, \mathbf{Q}\} = \frac{f(\mathbf{y} \mid \mathbf{a}', \mathbf{Q}) \Pr\{\mathbf{a} = \mathbf{a}' \mid \mathbf{Q}\}}{f(\mathbf{y} \mid \mathbf{Q})} = \frac{1}{|\mathscr{A}|^K} \frac{f(\mathbf{y} \mid \mathbf{a}', \mathbf{Q})}{f(\mathbf{y} \mid \mathbf{Q})}, \tag{6.13}$$

from which it becomes clear that maximizing $\Pr\{\mathbf{a} = \mathbf{a}' \mid \mathbf{y}, \mathbf{Q}\}$ over $\mathbf{a}'$ is equivalent to maximizing $f(\mathbf{y} \mid \mathbf{a}', \mathbf{Q})$ over $\mathbf{a}'$. Due to our assumption of zero-mean Gaussian noise with covariance $\mathbf{C_z}$, we have $f(\mathbf{y} \mid \mathbf{a}', \mathbf{Q}) = \frac{1}{\pi^K \det\{\mathbf{C_z}\}} \exp(-\|\mathbf{y} - \mathbf{Q}\mathbf{a}'\|^2_{\mathbf{C_z}^{-1}})$, so that coherent MLSD reduces to

$$\hat{\mathbf{a}}_{\mathsf{cMLSD}} = \arg\min_{\mathbf{a} \in \mathscr{A}^K} \|\mathbf{y} - \mathbf{Q}\mathbf{a}\|^2_{\mathbf{C_z}^{-1}}. \tag{6.14}$$

Above, we used the quadratic-form notation $\|\mathbf{z}\|^2_{\mathbf{A}} \triangleq \mathbf{z}^H\mathbf{A}\mathbf{z}$, where $\mathbf{A}$ is any positive semi-definite Hermitian matrix.

Complex-field symbol estimates

If the decoder prefers or tolerates complex-valued symbol estimates, rather than finite-alphabet symbol estimates, then one can consider equalization schemes that yield $\hat{\mathbf{a}} \in \mathbb{C}^K$. Note, however, that we still assume $\mathbf{a} \in \mathscr{A}^K$.

A popular criterion for this case is *minimum mean-squared error* (MMSE) [Poo94]. The coherent unconstrained MMSE sequence estimate is defined as

$$\hat{\mathbf{a}}_{\text{cMMSE}} \triangleq \arg \min_{\mathbf{a}' \in \mathbb{C}^K} \mathsf{E}\big\{ \|\mathbf{a} - \mathbf{a}'\|^2 \;\big|\; \mathbf{y}, \mathbf{Q} \big\}. \tag{6.15}$$

Since the MMSE estimate equals the conditional mean [Poo94], we have

$$\hat{\mathbf{a}}_{\text{cMMSE}} = \mathsf{E}\{\mathbf{a} \mid \mathbf{y}, \mathbf{Q}\} = \sum_{\mathbf{a} \in \mathscr{A}^K} \mathbf{a}\, p(\mathbf{a} \mid \mathbf{y}, \mathbf{Q}) \tag{6.16}$$

$$= \sum_{\mathbf{a} \in \mathscr{A}^K} \mathbf{a}\, \frac{f(\mathbf{y} \mid \mathbf{a}, \mathbf{Q}) p(\mathbf{a})}{\sum_{\mathbf{a}' \in \mathscr{A}^K} f(\mathbf{y} \mid \mathbf{a}', \mathbf{Q}) p(\mathbf{a}')}. \tag{6.17}$$

If we assume that $p(\mathbf{a})$ is uniformly distributed over $\mathscr{A}^K$, then

$$\hat{\mathbf{a}}_{\text{cMMSE}} = \frac{\sum_{\mathbf{a} \in \mathscr{A}^K} \mathbf{a}\, \exp\big(-\|\mathbf{y} - \mathbf{Q}\mathbf{a}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}\big)}{\sum_{\mathbf{a}' \in \mathscr{A}^K} \exp\big(-\|\mathbf{y} - \mathbf{Q}\mathbf{a}'\|^2_{\mathbf{C}_\mathbf{z}^{-1}}\big)}. \tag{6.18}$$

Notice from (6.18) that the finite-alphabet nature of $\mathbf{a}$ makes the conditional mean difficult to evaluate, since it requires the evaluation of $|\mathscr{A}|^K$ terms.

To reduce complexity, the MMSE criterion is often employed in conjunction with particular constraints on how the symbol estimates are generated from $\mathbf{y}$. The most common examples are MMSE *linear equalization* (6.22) and MMSE *decision feedback equalization* (6.27). Note that, if one assumes that $\mathbf{a}|\mathbf{y},\mathbf{Q}$ is Gaussian distributed, then the *un*constrained MMSE estimator (6.15) itself becomes a linear function of $\mathbf{y}$ [Poo94].

As the signal-to-noise ratio (SNR) increases, the effect of linear channel distortion overwhelms that of additive noise, motivating the so-called *zero-forcing* (ZF) criterion. Effectively, ZF equalizers "invert" the effect of the linear channel distortion while ignoring the presence of additive channel noise. The most common examples are ZF linear equalization and ZF decision feedback equalization, both described in Section 6.3.2. In the absence of additive noise, ZF equalizers are equivalent to their MMSE counterparts.

Soft bit estimates

If the decoder prefers *soft bit estimates*, then the goal is to produce reliability information on each of the coded bits in $\mathbf{c}$. Typically, bit reliabilities are expressed in the form of a *log likelihood ratio* (LLR) for each bit. The goal of coherent equalization thus becomes the computation of the coherent posterior LLRs[5]

$$L_{c|\mathbf{y},\mathbf{Q}}[j] \triangleq \ln \frac{\Pr\{c[j] = 1 \mid \mathbf{y}, \mathbf{Q}\}}{\Pr\{c[j] = 0 \mid \mathbf{y}, \mathbf{Q}\}} \quad \text{for} \quad j = 0, \ldots, K \log_2 |\mathscr{A}| - 1, \tag{6.19}$$

[5] Sometimes it is more practical to calculate posteriors using only a limited number of (say J) future observations [LVS95]. In this so-called "fixed lag" case, the conditioning in (6.19) is performed on $(y[0], \ldots, y[\lceil \frac{j}{\log_2 |\mathscr{A}|} \rceil + J])^T$ instead of $\mathbf{y}$.

given the a-priori LLRs

$$L_c[j] \quad \triangleq \quad \ln \frac{\Pr\{c[j]=1\}}{\Pr\{c[j]=0\}} \quad \text{for} \quad j=0,\ldots,K\log_2|\mathscr{A}|-1. \tag{6.20}$$

When nothing is a-priori known about the bit $c[j]$, the value $L_c[j]=0$ is used. Nonzero a-priori LLRs are used, e.g., when the equalizer is fed by the outputs of a soft decoder, as in turbo equalization (see Fig. 6.1) or when certain bits are known pilots. If $c[j]$ was a pilot (or otherwise known with complete confidence), then $L_c[j]=\pm\infty$. Recall that the use of a-priori LLRs implies that the equalizer treats the coded bits as independent.

Hard MAP bit estimates can be generated by quantizing the posterior LLRs as follows:

$$\hat{c}_{\mathsf{cMAP}}[j] \quad = \quad \frac{1}{2}\big(1+\mathrm{sign}(L_{c|\mathbf{y},\mathbf{Q}}[j])\big). \tag{6.21}$$

6.3.2 Coherent equalization tools

The coherent equalization criteria discussed in Section 6.3.1 each describe a particular goal for equalization, but not how equalization would be practically implemented. For example, the MAPSD, MLSD, and (unconstrained) MMSE estimates described in Section 6.3.1 require the evaluation of $\mathscr{O}(|\mathscr{A}|^K)$ metrics if computed via brute force, which is not practical for typical values of K. In this section, we review classical equalization implementations whose designs are guided by the various criteria in Section 6.3.1.

Trellis-based equalization

Trellis methods can be used to implement MLSD and MAP equalization when $\mathbf{Q}$ is a *banded* matrix. As illustrated in Fig. 6.5, a banded matrix differs from its quasi-banded counterpart due to the lack of corner elements. A banded matrix manifests when, e.g., the first and last few elements of $\mathbf{a}$ are known or zero-valued.[6] If $\mathbf{Q}$ is a banded matrix with a $2D+1$ wide band, then the *Viterbi algorithm* [For72] can perform MAPSD/MLSD equalization using $\mathscr{O}(KD|\mathscr{A}|^{2D+1})$ operations. Similarly, the *forward-backward* (or BCJR) algorithm [BCJR74] can be used to accomplish MAP symbol/bit equalization with a complexity of $\mathscr{O}(KD|\mathscr{A}|^{2D+1})$ operations [For73, Appendix]. Lower-complexity trellis-based approximate MAP equalizers include fixed-lag approaches [LVS95] and the soft-output Viterbi algorithm (SOVA) [HH89]. In all cases, the complexity is linear in the block length K and exponential in the effective channel length $2D+1$. Thus, these techniques will be practical if and only if $2D+1$ is very small.

Trellis methods can be modified to work on *quasi*-banded $\mathbf{Q}$ using, e.g., a "tail-biting" approach. Here, from an arbitrary location within the block, the Viterbi algorithm is initialized from each of the possible $|\mathscr{A}|^{2D+1}$ states and forced to terminate in the same state; the initialization leading to the optimum sequence metric is then chosen. This approach requires running the Viterbi algorithm $|\mathscr{A}|^{2D+1}$ times, for a total cost of $\mathscr{O}(KD|\mathscr{A}|^{4D+2})$ operations.

[6] More precisely, consider the system model (6.2). If $\mathbf{Q}$ is as illustrated in Fig. 6.5(a) and the last $M-1$ elements of $\mathbf{a}$ are zero-valued, then we can write $\mathbf{y}=\breve{\mathbf{Q}}\breve{\mathbf{a}}+\mathbf{z}$ where $\breve{\mathbf{a}}=\big(a[0] \quad \cdots \quad a[K-M]\big)^T$ and where $\breve{\mathbf{Q}}$ is a banded matrix (as illustrated in Fig. 6.5(b)) with an M-wide band. Or, if $\mathbf{Q}$ is as illustrated in Fig. 6.5(b) and the first and last D elements of $\mathbf{a}$ are zero-valued, then we can write $\mathbf{y}=\breve{\mathbf{Q}}\breve{\mathbf{a}}+\mathbf{z}$ where $\breve{\mathbf{a}}=\big(a[D] \quad \cdots \quad a[K-D-1]\big)^T$ and where $\breve{\mathbf{Q}}$ is a banded matrix (as illustrated in Fig. 6.5(b)) with a $2D+1$ wide band.

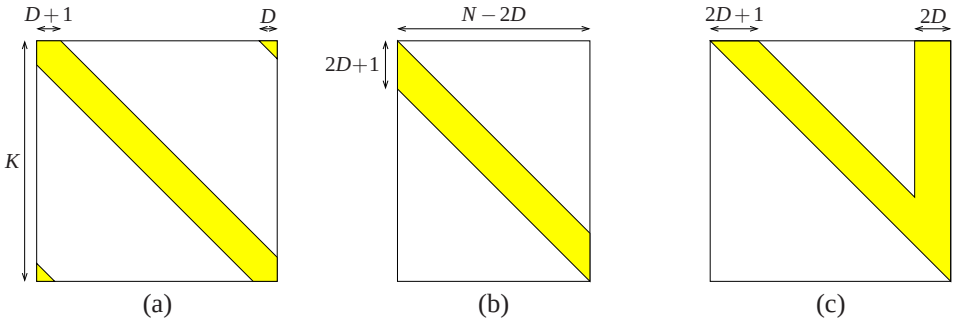

Figure 6.5 *Support region of (a) "quasi-banded," (b) "banded," and (c) "V-shaped" matrices. From quasi-banded (a), banded (b) is obtained by deleting the first and last D columns, while V-shaped (c) is obtained by upper triangularization.*

Linear equalization

In *linear equalization*, the symbol estimates are a linear function of the observation $\mathbf{y}$, i.e.,

$$\hat{\mathbf{a}}_{\mathsf{LIN}} \quad = \quad \mathbf{E}\mathbf{y}, \tag{6.22}$$

for a suitably chosen matrix $\mathbf{E} \in \mathbb{C}^{K\times K}$. In some cases, such as when it is impractical to process the entire observation $\mathbf{y}$ at once, additional constraints are placed on $\mathbf{E}$. Because linear equalization ignores the finite-alphabet property of $\mathbf{a}$, its performance is generally much worse than that of techniques which leverage the finite-alphabet property.

The coherent *linear MMSE* (LMMSE)[7] equalizer uses, for $\mathbf{E}$ in (6.22),

$$\mathbf{E}_{\mathsf{LMMSE}} \quad \triangleq \quad \arg\min_{\mathbf{E}\in\mathbb{C}^{K\times K}} \mathsf{E}\left\{\|\mathbf{a}-\hat{\mathbf{a}}_{\mathsf{LIN}}\|^2 \;\middle|\; \mathbf{Q}\right\}. \tag{6.23}$$

Given the symbol and noise statistics assumed in Section 6.2, it can be shown [Ver98] that

$$\mathbf{E}_{\mathsf{LMMSE}} \quad = \quad \mathbf{Q}^H(\mathbf{Q}\mathbf{Q}^H+\sigma_a^{-2}\mathbf{C}_{\mathbf{z}})^{-1} \tag{6.24}$$

$$\quad = \quad (\mathbf{Q}^H\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{Q}+\sigma_a^{-2}\mathbf{I}_K)^{-1}\mathbf{Q}^H\mathbf{C}_{\mathbf{z}}^{-1}. \tag{6.25}$$

The matrix inversion lemma[8] can be used to relate (6.24) and (6.25). The *linear ZF* (LZF) estimator uses (6.22) with $\mathbf{E}$ set to

$$\mathbf{E}_{\mathsf{LZF}} \quad = \quad \mathbf{Q}^{-1}, \tag{6.26}$$

assuming that $\mathbf{Q}$ is invertible. When $\mathbf{Q}$ is not invertible, the LZF equalizer is said not to exist.

Due to the matrix inversions in (6.24)–(6.26), the complexity of LMMSE and LZF equalization is $\mathscr{O}(K^3)$, which is much less than the $\mathscr{O}(|\mathscr{A}|^K)$ complexity of unconstrained MMSE estimation in (6.18). Still, $\mathscr{O}(K^3)$ may be impractical when K is large.

---

[7] Note that the LMMSE equalizer described here is a generalization of the classical tapped delay-line LMMSE equalizer [Pro01].

[8] The matrix inversion lemma can be stated as $(\mathbf{A}^{-1}+\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H)^{-1} = \mathbf{A}-\mathbf{A}\mathbf{B}(\mathbf{C}+\mathbf{B}^H\mathbf{A}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{A}$, assuming the inverses exist.

Decision feedback equalization

Decision feedback equalization (DFE) exploits the finite-alphabet symbol property while keeping complexity close to that of linear equalization. Essentially, it makes hard symbol decisions sequentially and leverages past decisions for future symbol estimates.

The DFE generates complex-valued symbol estimates as follows:[9]

$$\hat{\mathbf{a}}_{\text{DFE}} \quad = \quad \mathbf{E}\mathbf{y} - (\mathbf{U} - \mathbf{I}_K)\mathscr{D}_{\mathscr{A}}(\hat{\mathbf{a}}_{\text{DFE}}). \tag{6.27}$$

In (6.27), $\mathscr{D}_{\mathscr{A}}(\cdot) : \mathbb{C}^K \to \mathscr{A}^K$ denotes element-wise quantization w.r.t. the symbol alphabet $\mathscr{A}$, $\mathbf{U} \in \mathbb{C}^{K\times K}$ is monic upper triangular (to ensure that decision feedback is strictly causal), and $\mathbf{E} \in \mathbb{C}^{K\times K}$. Keeping the monic upper-triangular property of $\mathbf{U}$ in mind, (6.27) can be understood as follows: the estimate $\hat{a}_{\text{DFE}}[K-1]$ is linearly computed from $\mathbf{y}$ using the last row in $\mathbf{E}$; then, the estimate $\hat{a}_{\text{DFE}}[K-2]$ is linearly computed from $\mathbf{y}$ and quantized $\hat{a}_{\text{DFE}}[K-1]$ using the second-to-last rows in $\mathbf{E}$ and $\mathbf{U}$, respectively; then, the estimate $\hat{a}_{\text{DFE}}[K-3]$ is linearly computed from $\mathbf{y}$ and quantized $\{\hat{a}_{\text{DFE}}[K-2], \hat{a}_{\text{DFE}}[K-1]\}$ using the third-to-last rows in $\mathbf{E}$ and $\mathbf{U}$, respectively; and so on.

The DFE matrices $\mathbf{E}$ and $\mathbf{U}$ are typically designed according to the MMSE or ZF criteria. As with linear equalization, additional constraints may be placed on $\mathbf{E}$ and/or $\mathbf{U}$. The coherent *MMSE-DFE* [CF97] uses (6.27) with $\{\mathbf{E}, \mathbf{U}\}$ set to

$$\begin{aligned} \{\mathbf{E}_{\text{MMSE-DFE}}, \mathbf{U}_{\text{MMSE-DFE}}\} \quad &= \quad \arg\min_{\mathbf{E},\mathbf{U}} \mathsf{E}\{\|\mathbf{a} - \hat{\mathbf{a}}_{\text{DFE}}\|^2 \mid \mathbf{Q}\} \\ &\qquad \text{assuming } \mathscr{D}_{\mathscr{A}}(\hat{\mathbf{a}}_{\text{DFE}}) = \mathbf{a}, \end{aligned} \tag{6.28}$$

i.e., set to minimize the MSE of $\hat{\mathbf{a}}_{\text{DFE}}$ *under the assumption of perfect decision feedback*. It can be shown that $\mathbf{U}_{\text{MMSE-DFE}}$ and $\mathbf{E}_{\text{MMSE-DFE}}$ can be computed with the aid of an LDU decomposition [ADS00]:

$$\mathbf{U}^H_{\text{MMSE-DFE}} \mathbf{\Delta}_{\text{MMSE-DFE}} \mathbf{U}_{\text{MMSE-DFE}} \quad = \quad \mathbf{Q}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{Q} + \sigma_a^{-2} \mathbf{I} \tag{6.29}$$

$$\mathbf{E}_{\text{MMSE-DFE}} \quad = \quad \mathbf{U}_{\text{MMSE-DFE}} \mathbf{E}_{\text{LMMSE}}, \tag{6.30}$$

with $\mathbf{E}_{\text{LMMSE}}$ given by (6.24)-(6.25). The *ZF-DFE* takes the form of (6.27) with $\mathbf{U}_{\text{ZF-DFE}}$ computed via the LDU decomposition $\mathbf{U}^H_{\text{ZF-DFE}} \mathbf{\Delta}_{\text{ZF-DFE}} \mathbf{U}_{\text{ZF-DFE}} = \mathbf{Q}^H \mathbf{Q}$, and with $\mathbf{E}_{\text{ZF-DFE}} = \mathbf{U}_{\text{ZF-DFE}} \mathbf{Q}^{-1}$.

In practice, the hard decisions in (6.27) are not always perfect, which leads to the phenomenon known as *error propagation* [Od85]. There, a decision error on $a[k]$ has the effect of amplifying, rather than canceling, the interference that $a[k]$ causes to the not-yet-estimated symbols $\{a[k']\}_{k'=0}^{k-1}$. While error propagation can be somewhat alleviated by detecting symbols with higher signal-to-interference noise ratio (SINR) first (e.g., by V-BLAST detection ordering [WFGV98]), error propagation is better avoided through tree search or iterative soft equalization, as discussed below.

Finally, we note that hybrid trellis/DFE techniques have been proposed with complexities and performances that lie between trellis and DFE methods. Two of the more well known techniques are *reduced state sequence estimation* [EQ88] and *delayed decision feedback estimation* [DHH89].

---

[9] The DFE described here is sometimes referred to as a "generalized" DFE to distinguish it from the classical DFE implemented using tapped delay-line forward and feedback filters [ADC95].

Equalization based on tree search

In DFE, a single hypothesis of the sequence $(a[k+1],\ldots,a[K-1])$ is used to aid the estimation of $a[k]$. Tree-search[10] methods improve on this idea by keeping and using several hypotheses of the sequence $(a[k+1],\ldots,a[K-1])$ until it is clear which is the single best hypothesis.

Tree-search algorithms can be partitioned into optimal and suboptimal approaches. *Optimal tree search* methods are capable of implementing MLSD with a complexity that is *on average* much less than that of brute-force search [Mow94]. Though this average complexity has been claimed to grow as roughly $\mathcal{O}(K^3)$ at sufficiently high SNR [HV05], a careful analysis shows that, in fact, the average complexity of optimal tree search is exponential in K [JO05]. To circumvent the potentially high complexity of exact tree search (especially at low SNR), *suboptimal tree search* may be considered, since a very small performance sacrifice can often lead to a huge reduction in complexity. In fact, a well-designed suboptimal tree search can achieve near-ML performance with near-DFE complexity [MEDC06]. When assessing a suboptimal tree search algorithm, it is most appropriate to think in terms of its *performance/complexity tradeoff*.

Before conducting a tree search, the observations $\mathbf{y}$ in (6.2) must be pre-processed to yield a *causal observation model* of the form

$$\bar{\mathbf{y}} \quad = \quad \bar{\mathbf{Q}}\bar{\mathbf{a}}+\bar{\mathbf{z}}, \tag{6.31}$$

where $\bar{\mathbf{Q}}$ is upper triangular and $\bar{\mathbf{a}}$ is some permutation of $\mathbf{a}$. Ignoring permutation for the moment (so that $\bar{\mathbf{a}}=\mathbf{a}$), the standard approach to upper-triangularization of (6.2) is *QR decomposition*: if $\mathbf{Q}=\mathbf{V}_{\mathsf{QR}}\bar{\mathbf{Q}}_{\mathsf{QR}}$, where $\mathbf{V}_{\mathsf{QR}}$ is unitary and $\bar{\mathbf{Q}}_{\mathsf{QR}}$ is upper triangular, then pre-processing according to $\bar{\mathbf{y}}=\mathbf{V}_{\mathsf{QR}}^H\mathbf{y}\triangleq\bar{\mathbf{y}}_{\mathsf{QR}}$ yields (6.31) with $\mathbf{Q}=\mathbf{Q}_{\mathsf{QR}}$ and $\bar{\mathbf{z}}=\mathbf{V}_{\mathsf{QR}}^H\mathbf{z}\triangleq\bar{\mathbf{z}}_{\mathsf{QR}}$. Notice that $\bar{\mathbf{z}}_{\mathsf{QR}}$ is statistically equivalent to $\mathbf{z}$. As we show below, QR pre-processing is closely related to the feedforward filtering operation in ZF-DFE. Since the MMSE-DFE is known to outperform the ZF-DFE in noisy environments, it has been suggested [DEC03] to replace the QR pre-processing step with its MMSE-DFE equivalent, at least for suboptimal tree search. To see this from another perspective, imagine for the moment that suboptimal tree search is conducted according to the most greedy method possible, i.e., with a single surviving hypothesis per stage. Then, if QR pre-processing is used, this suboptimal tree search is exactly the ZF-DFE, whereas, if MMSE-DFE pre-processing is used, this suboptimal tree search is exactly the MMSE-DFE.

We will now provide the technical link between the QR decomposition and the ZF-DFE, as well as the details of the MMSE-DFE pre-processor. Comparing the LDU decomposition $\mathbf{Q}^H\mathbf{Q}=\mathbf{U}_{\mathsf{ZF\text{-}DFE}}^H\mathbf{\Delta}_{\mathsf{ZF\text{-}DFE}}\mathbf{U}_{\mathsf{ZF\text{-}DFE}}$ to the QR decomposition $\mathbf{Q}=\mathbf{V}_{\mathsf{QR}}\bar{\mathbf{Q}}_{\mathsf{QR}}$, it becomes evident that $\bar{\mathbf{Q}}_{\mathsf{QR}}=\mathbf{\Delta}_{\mathsf{ZF\text{-}DFE}}^{1/2}\mathbf{U}_{\mathsf{ZF\text{-}DFE}}$, from which it follows that $\mathbf{V}_{\mathsf{QR}}^H=\bar{\mathbf{Q}}_{\mathsf{QR}}\mathbf{Q}^{-1}=\mathbf{\Delta}_{\mathsf{ZF\text{-}DFE}}^{1/2}\mathbf{U}_{\mathsf{ZF\text{-}DFE}}\mathbf{Q}^{-1}=\mathbf{\Delta}_{\mathsf{ZF\text{-}DFE}}^{1/2}\mathbf{E}_{\mathsf{ZF\text{-}DFE}}$. Thus, $\bar{\mathbf{y}}_{\mathsf{QR}}=\mathbf{\Delta}_{\mathsf{ZF\text{-}DFE}}^{1/2}\mathbf{E}_{\mathsf{ZF\text{-}DFE}}\mathbf{y}$ can be recognized as a scaled version of the ZF-DFE feedforward filter output $\mathbf{E}_{\mathsf{ZF\text{-}DFE}}\mathbf{y}$. If we repeat the same steps with MMSE-DFE quantities in place of ZF-DFE quantities, we obtain the *MMSE-*

[10] What we call "tree search" is sometimes referred to as closest lattice point search, lattice decoding, sequential decoding, or sphere decoding.

DFE pre-processed observation [DEC03]

$$\bar{\mathbf{y}}_{\text{MMSE-DFE}} \triangleq \mathbf{\Delta}_{\text{MMSE-DFE}}^{1/2}\mathbf{E}_{\text{MMSE-DFE}}\mathbf{y}, \tag{6.32}$$

and the corresponding causal model

$$\bar{\mathbf{y}}_{\text{MMSE-DFE}} = \mathbf{\Delta}_{\text{MMSE-DFE}}^{1/2}\mathbf{U}_{\text{MMSE-DFE}}\mathbf{a} + \bar{\mathbf{z}}_{\text{MMSE-DFE}}, \tag{6.33}$$

where $\bar{\mathbf{z}}_{\text{MMSE-DFE}} \triangleq \bar{\mathbf{y}}_{\text{MMSE-DFE}} - \mathbf{\Delta}_{\text{MMSE-DFE}}^{1/2}\mathbf{U}_{\text{MMSE-DFE}}\mathbf{a}$.

MMSE-DFE pre-processed tree search proceeds from the causal model (6.33), where the interference $\bar{\mathbf{z}}_{\text{MMSE-DFE}}$ is treated as (signal-independent) additive white Gaussian noise (AWGN). Although it can be shown that $\bar{\mathbf{z}}_{\text{MMSE-DFE}}$ is white (in fact, $\mathbf{C}_{\bar{\mathbf{z}}_{\text{MMSE-DFE}}} = \mathbf{I}$ for any $\mathbf{C}_{\mathbf{z}}$), it can readily be seen that $\bar{\mathbf{z}}_{\text{MMSE-DFE}}$ is signal-dependent (and hence non-Gaussian) [DEC03]. Thus, treating $\bar{\mathbf{z}}_{\text{MMSE-DFE}}$ as if it were AWGN will produce suboptimal[11] sequence estimates. However, it turns out that the increase in pre-quantization SINR (from the use of MMSE-DFE in place of ZF-DFE) more than compensates for the loss in optimality (due to non-AWGN $\bar{\mathbf{z}}_{\text{MMSE-DFE}}$). Thus, relative to QR pre-processing, MMSE-DFE pre-processing has been observed to yield significant improvements in the performance/complexity tradeoff of suboptimal tree search [MEDC06].

Other types of pre-processing include lattice reduction (e.g., the method of Lenstra, Lenstra, and Lovász [LLL82]) and column permutation (e.g., re-ordering of $\mathbf{a}$ so that stronger symbols are decided first, as in V-BLAST ordering [WFGV98]). Since these techniques would destroy the quasi-banded structure of $\mathbf{Q}$, however, we will not elaborate on them further.

Tree search algorithms (whether optimal or suboptimal) can be categorized as breadth-first, depth-first, or best-first [AM84, MEDC06]. Breadth-first search algorithms include, e.g., the M-algorithm [AM84], the T-algorithm [Sim90], statistical pruning algorithms [GH03], the Wozencraft sequential decoder [WR61], and the Pohst sphere decoder [FP85]. Depth-first search algorithms include, e.g., the Schnorr-Euchner sphere decoder and its variants [VB99, AEVZ02, DEC03]. Best-first search algorithms include, e.g., the stack and Fano algorithms [VO79, Fan63, MEDC06]. Since a thorough description and comparison of these approaches are outside the scope of this chapter, we make only a few remarks. The Fano algorithm was recently found to yield a superior complexity/performance tradeoff when $\mathbf{Q}$ was either a convolution matrix or fully populated [MEDC06]. The same result does not appear to hold when $\mathbf{Q}$ is quasi-banded, though [HS06]. The M-algorithm is popular for two reasons: simplicity and fixed complexity (i.e., complexity invariant to channel/noise realizations and SNR).

While so far we have focused on tree-search implementations of MLSD, we now describe how tree search can be used to find (approximate) posterior LLRs, and thus MAP symbol and bit estimates, using the method of Hochwald and ten Brink [Ht03]. First we define the *coherent MAP sequence metric*

$$\zeta_{\text{coh}}(\mathbf{c}) \triangleq \ln f(\mathbf{y} \mid \mathbf{c}, \mathbf{Q}) + \mathbf{l}_c^T\mathbf{c} \tag{6.34}$$

$$= -\|\mathbf{y} - \mathbf{Q}\mathbf{a}\|_{\mathbf{C}_{\mathbf{z}}^{-1}}^2 - \ln(\pi^K \det\{\mathbf{C}_{\mathbf{z}}\}) + \mathbf{l}_c^T\mathbf{c}, \tag{6.35}$$

[11] Interestingly, it has been shown that $\bar{\mathbf{z}}_{\text{MMSE-DFE}}$ can be treated as AWGN when $\mathscr{A}$ is constant modulus. In other words, $\hat{\mathbf{a}}_{\text{cMLSD}} = \arg\min_{\mathbf{a}\in\mathscr{A}^K} \|\bar{\mathbf{y}}_{\text{MMSE-DFE}} - \mathbf{\Delta}_{\text{MMSE-DFE}}^{1/2}\mathbf{U}_{\text{MMSE-DFE}}\mathbf{a}\|^2$ for constant modulus $\mathscr{A}$ [HS05].

where $\mathbf{l}_c \triangleq \begin{pmatrix} L_c[0] & \cdots & L_c[K-1] \end{pmatrix}^T$, with $L_c[k]$ being a-priori LLRs from (6.20), and where the symbols $\mathbf{a}$ are determined by the hypothesized bit vector $\mathbf{c}$. As previously remarked, the use of a-prior LLRs implies that the coded bits $\{c[j]\}$ are treated as independent. It is straighforward to show (see Appendix 6.A) that the posterior LLR defined in (6.19) can be written as

$$L_{c|\mathbf{y},\mathbf{Q}}[j] \quad = \quad \ln \frac{\sum_{\mathbf{c}:c[j]=1} e^{\zeta_{\mathsf{coh}}(\mathbf{c})}}{\sum_{\mathbf{c}:c[j]=0} e^{\zeta_{\mathsf{coh}}(\mathbf{c})}}. \tag{6.36}$$

Note that, in the summations of (6.36), all possibilities of $\mathbf{c} \in \{0,1\}^{K \log_2 |\mathscr{A}|}$ are considered, not only those in the codebook. (The same holds true in related equations throughout the chapter.) This reflects our assumption that the equalizer does *not* use knowledge of the code structure to generate posterior LLRs; code structure is exploited only by the decoder.

Computing $L_{c|\mathbf{y},\mathbf{Q}}[j]$ via (6.36) would require $2^{K \log_2 |\mathscr{A}|}$ evaluations of the MAP metric $\zeta_{\mathsf{coh}}(\mathbf{c})$, and hence would be impractical. However, as suggested in [Ht03], the "max-log" approximation $\ln \sum_{\mathbf{c}} e^{\zeta(\mathbf{c})} \approx \max_{\mathbf{c}} \zeta(\mathbf{c})$ can be applied to yield

$$L_{c|\mathbf{y},\mathbf{Q}}[j] \quad \approx \quad \max_{\mathbf{c}:c[j]=1} \zeta_{\mathsf{coh}}(\mathbf{c}) - \max_{\mathbf{c}:c[j]=0} \zeta_{\mathsf{coh}}(\mathbf{c}). \tag{6.37}$$

Suboptimal tree search can then be used to find the set of all bit vectors $\mathbf{c} \in \{0,1\}^{K \log_2 |\mathscr{A}|}$ which yield non-negligible coherent MAP metrics $\zeta_{\mathsf{coh}}(\mathbf{c})$, as detailed in [dW05]. Once the posterior LLRs have been calculated, it is possible to generate hard bit estimates via (6.21), if needed. In a turbo configuration, though, the equalizer passes the posterior LLRs to a soft-input/soft-output decoder. After decoding, the refined LLRs are passed back to the equalizer to be used as priors, i.e., $\mathbf{l}_c$. (Recall Fig. 6.1.)

Iterative soft equalization

For approximate symbol/bit MAP equalization, one can consider using *iterative soft equalization* techniques [WP99, TKS02] as an alternative to the trellis and tree-search approaches described earlier. The iterative soft equalization techniques described here use *linear estimation* strategies in conjunction with *evolving beliefs* of the interfering bits. After estimating a given bit, the equalizer updates its belief about that bit to better estimate the other bits. Once all bit beliefs (e.g., LLRs) have been updated, the process repeats. The equalizer may itself iterate several times and/or it may trade soft bit information with a soft-input/soft-output decoder in a turbo configuration.

Below we detail the main concepts behind iterative soft equalization for the simple case of BPSK.[12] This simplification allows us to make a direct mapping between each bit and a corresponding symbol, e.g., $a[j] = 2b[j] - 1$ for $j = 0, \ldots, K-1$, where $b[j] \in \{0,1\}$ and $a[j] \in \mathscr{A} = \{-1,+1\}$. In this case, the a-priori LLR from (6.20) can be rewritten as

$$L_c[j] \quad = \quad \ln \frac{\Pr\{a[j]=+1\}}{\Pr\{a[j]=-1\}}. \tag{6.38}$$

[12] The case of non-binary alphabets follows similar principles but is more tedious to describe.

Suppose that we are interested in estimating the j^{th} bit, $c[j]$, or equivalently the j^{th} symbol, $a[j]$. And say that, when doing so, we have prior information on the *other* bits, and thus the other symbols $\bar{\mathbf{a}}_j \triangleq \big(a[0] \quad \cdots \quad a[j-1] \quad 0 \quad a[j+1] \quad \cdots \quad a[K-1]\big)^T$, that comes in the form of a-priori LLRs. To facilitate the use of linear operations, the symbol estimation stage treats the elements in $\bar{\mathbf{a}}_j$ as independent Gaussian with means and variances that are calculated from the respective LLRs. In particular, the calculated mean of $a[k]$ (for $k \neq j$) is computed via $\mu_a[k] \triangleq \sum_{a\in\{-1,+1\}} a \Pr\{a[k]=a\}$ using the identity

$$\Pr\{a[k]=a\} \quad = \quad \frac{\exp\big((a-1)L_c[k]/2\big)}{1+\exp(-L_c[k])} \quad \text{for} \quad a\in\{-1,+1\}, \tag{6.39}$$

from which it can be shown that

$$\mu_a[k] \quad = \quad \frac{1-\exp(-L_c[k])}{1+\exp(-L_c[k])} \;=\; \tanh(L_c[k]/2). \tag{6.40}$$

Similarly, the calculated variance of $a[k]$ (for $k \neq j$) is computed via

$$v_a[k] \quad \triangleq \quad -\mu_a[k]^2 + \sum_{a\in\{-1,+1\}} a^2 \Pr\{a[k]=a\} \;=\; 1-\mu_a[k]^2. \tag{6.41}$$

The estimation of $a[j]$ proceeds by writing the observation as

$$\mathbf{y} \quad = \quad \mathbf{q}_j a[j] + \mathbf{Q}\bar{\mathbf{a}}_j + \mathbf{z}, \tag{6.42}$$

where $\mathbf{q}_j$ denotes the j^{th} column of $\mathbf{Q}$. For convenience, we collect the calculated means into $\bar{\boldsymbol{\mu}}_j \triangleq \big(\mu_a[0] \quad \cdots \quad \mu_a[j-1] \quad 0 \quad \mu_a[j+1] \quad \cdots \quad \mu_a[K-1]\big)^T$ and the calculated variances into $\bar{\mathbf{v}}_j \triangleq \big(v_a[0] \quad \cdots \quad v_a[j-1] \quad 0 \quad v_a[j+1] \quad \cdots \quad v_a[K-1]\big)^T$.

In the classical iterative soft equalization approach proposed by Wang and Poor [WP99], soft interference cancellation:

$$\mathbf{x}_j \quad = \quad \mathbf{y} - \mathbf{Q}\bar{\boldsymbol{\mu}}_j \tag{6.43}$$

is followed by LMMSE combining:

$$\hat{a}_{\mathsf{LMMSE}}[j] \quad = \quad \mathbf{e}_j^H \mathbf{x}_j \quad \text{with} \quad \mathbf{e}_j = \arg\min_{\mathbf{e}\in\mathbb{C}^K} \mathsf{E}\big\{\big|a[j]-\mathbf{e}^H\mathbf{x}_j\big|^2\big\}. \tag{6.44}$$

Writing the interference-canceled vector as

$$\mathbf{x}_j \quad = \quad \mathbf{q}_j a[j] + \mathbf{r}_j, \tag{6.45}$$

with residual interference vector

$$\mathbf{r}_j \quad = \quad \mathbf{Q}(\bar{\mathbf{a}}_j - \bar{\boldsymbol{\mu}}_j) + \mathbf{z}, \tag{6.46}$$

it can be seen that $\mathbf{C}_{\mathbf{r}_j} = \mathbf{Q}\,\mathrm{diag}\{\bar{\mathbf{v}}_j\}\mathbf{Q}^H + \mathbf{C}_{\mathbf{z}}$. Withholding prior belief on $a[j]$, so that $\mathsf{E}\{a[j]\}=0$ and $\mathrm{var}\{a[j]\}=1$, the LMMSE combiner in (6.44) becomes

$$\mathbf{e}_j \quad = \quad \mathbf{C}_{\mathbf{x}_j}^{-1}\mathbf{C}_{\mathbf{x}_j,a[j]} \;=\; \big(\mathbf{q}_j\mathbf{q}_j^H + \mathbf{C}_{\mathbf{r}_j}\big)^{-1}\mathbf{q}_j \;=\; \frac{1}{1+\mathbf{q}_j^H\mathbf{C}_{\mathbf{r}_j}^{-1}\mathbf{q}_j}\mathbf{C}_{\mathbf{r}_j}^{-1}\mathbf{q}_j, \tag{6.47}$$

where the matrix inversion lemma was used to obtain the right side of (6.47). Thus, $\hat{a}_{\mathsf{LMMSE}}[j]$ becomes

$$\hat{a}_{\mathsf{LMMSE}}[j] \quad = \quad \frac{\mathbf{q}_j^H \mathbf{C}_{\mathbf{r}_j}^{-1} \mathbf{x}_j}{1+\mathbf{q}_j^H \mathbf{C}_{\mathbf{r}_j}^{-1} \mathbf{q}_j}. \tag{6.48}$$

From straightforward arguments,[13] one can conclude that any scaled version of the statistic

$$g[j] \quad \triangleq \quad \mathbf{q}_j^H \mathbf{C}_{\mathbf{r}_j}^{-1} \mathbf{x}_j, \tag{6.49}$$

including the LMMSE estimate $\hat{a}_{\mathsf{LMMSE}}[j]$, is sufficient [Poo94] for ML[14] detection of $a[j]$ (and thus of $b[j]$) from $\mathbf{x}_j$. In fact, the ML symbol decision is simply the sign of

$$L_g[j] \quad \triangleq \quad \ln \frac{f(g[j] \mid a[j]=+1)}{f(g[j] \mid a[j]=-1)} \;=\; \ln \frac{f(g[j] \mid c[j]=1)}{f(g[j] \mid c[j]=0)}. \tag{6.50}$$

Expanding $g[j]$ as

$$g[j] \quad = \quad \mathbf{q}_j^H \mathbf{C}_{\mathbf{r}_j}^{-1} \mathbf{q}_j \, a[j] + \mathbf{q}_j^H \mathbf{C}_{\mathbf{r}_j}^{-1} \mathbf{r}_j, \tag{6.51}$$

it can be seen that $g[j]\big|a[j]$ is circular Gaussian with mean $a[j]\mu_{g[j]}$ and variance $\sigma^2_{g[j]}$, where $\mu_{g[j]} = \mathbf{q}_j^H \mathbf{C}_{\mathbf{r}_j}^{-1} \mathbf{q}_j = \sigma^2_{g[j]}$. Hence,

$$\begin{aligned} L_g[j] \quad &= \quad \ln \frac{\exp\left(-\left|g[j]-\mu_{g[j]}\right|^2/\sigma^2_{g[j]}\right)}{\exp\left(-\left|g[j]+\mu_{g[j]}\right|^2/\sigma^2_{g[j]}\right)} && (6.52) \\ &= \quad -\left|g[j]-\mu_{g[j]}\right|^2/\sigma^2_{g[j]} + \left|g[j]+\mu_{g[j]}\right|^2/\sigma^2_{g[j]} && (6.53) \\ &= \quad 4\,\mathrm{Re}\{g[j]\}. && (6.54) \end{aligned}$$

Finally, a posterior LLR on $a[j]$ (and hence on $c[j]$) can be generated via

$$\ln \frac{\Pr\{a[j]=+1 \mid g[j]\}}{\Pr\{a[j]=-1 \mid g[j]\}} \;=\; \ln \frac{\Pr\{c[j]=1 \mid g[j]\}}{\Pr\{c[j]=0 \mid g[j]\}} \quad = \quad L_g[j]+L_c[j], \tag{6.55}$$

where (6.50) and Bayes rule were used to obtain the right side of (6.55). The posterior LLR (6.55) can then be used in place of $L_c[j]$ in (6.40)-(6.41) to calculate the mean $\mu_a[j]$ and variance $v_a[j]$ for subsequent estimation of $\{c[k]\}_{k\neq j}$.

Remarks on complexity

The coherent equalization tools described in this section are quite general; they apply to any $\mathbf{Q}$, and thus any type of linear modulation/demodulation combined with any type of linear channel propagation (whether the channel is rapidly TV or not). In fact, when

[13] Sufficiency can be understood as follows. After constructing the interference-cancelled/whitened observation $\mathbf{C}_{\mathbf{r}_j}^{-1/2}\mathbf{x}_j = \mathbf{C}_{\mathbf{r}_j}^{-1/2}\mathbf{q}_j a[j] + \mathbf{C}_{\mathbf{r}_j}^{-1/2}\mathbf{r}_j$, the application of the matched filter $\mathbf{C}_{\mathbf{r}_j}^{-1/2}\mathbf{q}_j$, or any scaling thereof, yields a sufficient statistic for the detection of $a[j]$ [Poo94]. These two steps are combined in writing (6.49).

[14] Since we assume a uniform prior on $a[j]$, ML detection is equivalent to MAP detection.

structure in **Q** is lacking or ignored, equalization can be viewed as a form of *CDMA multiuser detection* [Ver98, Mos96] where the code matrix (in this case **Q**) changes from one bit to the next, or as a form of *MIMO decoding* [TV05] for communication over a flat-fading channel with K transmit and K receive antennas. For the case of generic **Q**, however, the cost of implementing the equalization criteria rises rapidly with K, the block size. For example, we saw that linear and DFE schemes consume $\mathcal{O}(K^3)$ operations per block, and that more sophisticated schemes can be significantly more expensive. Since, for the applications we envision, typical values of K can be in the hundreds or thousands, equalization is made practical only by leveraging the structural properties of **Q** discussed in Section 6.2.2. Using these properties, Sections 6.3.3–6.3.4 below describe equalization algorithms specifically tailored to rapidly TV channels.

6.3.3 Coherent equalization for time-frequency concentrated modulation/demodulation

Recall from Section 6.2.2 that, when sufficiently time-frequency concentrated modulation/demodulation pulses are used, the effective channel matrix **Q** falls into the "narrowly quasi-banded" class. Here, **Q** contains only negligible coefficients outside of the shaded region in Fig. 6.2(b), for some $D \ll K$. The main idea behind the equalization algorithms discussed in this section is that, by ignoring these negligible coefficients, the complexity of equalization can be significantly reduced without a significant loss in performance.

In this section, we will treat the interference caused by the negligible coefficients in **Q** as if it were part of the additive noise **z**, allowing us to regard the negligible coefficients in **Q** as if they were zero-valued. In doing so, we will assume that the interference radius D has been chosen large enough so that these additional contributions to **z** are relatively small (for the SNRs of interest). In particular, we will assume that the value of D allows us to continue treating **z** as statistically independent of **a**, as assumed in Section 6.2. With suitably designed modulation/demodulation schemes like the max-SINR schemes in [DS07], these assumptions have been shown [HS06][15] to be satisfied with $D = \lceil \nu_{\max} T_s K \rceil + 1$ at SNRs up to at least 10dB and with $D = \lceil \nu_{\max} T_s K \rceil + 2$ at SNRs up to at least 30dB. Less time-frequency concentrated schemes require larger values of D, making equalization more expensive to implement for the same level of residual interference. For example, the interference profiles in Fig. 6.4 suggest that Strohmer and Beaver's scheme [SB03] requires a radius D at least 2 higher than the max-SINR scheme of [DS07] for the same level of residual interference.

In the remainder of this section, we provide some insight into how the narrowly quasi-banded structure of **Q** can be leveraged to lower the complexity of the equalization strategies described in Section 6.3.2. In particular, we identify two principal approaches to this problem: *fast serial equalization* and *fast joint equalization*. We keep our description brief because equalization for narrowly quasi-banded **Q** is related to particular forms of equalization for OFDM, which is the topic of Chapter 7.

[15] For the specified D and SNR range, MLSD performance was found to be identical whether the out-of-band **Q** coefficients were treated as part of the channel or as part of the noise. We note that the variable "D" in [HS06] is defined to have twice the value of D in this chapter, since in [HS06] the effective channel matrix is real-valued.

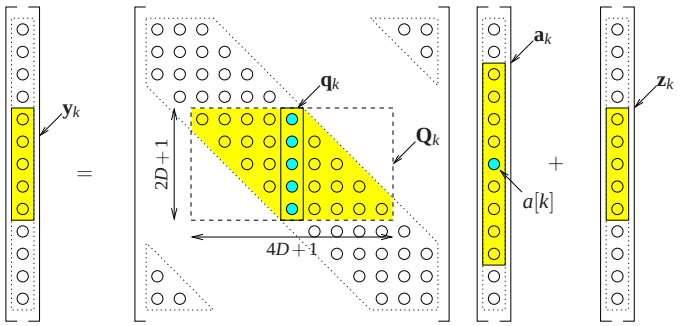

Figure 6.6 *The local observation model used for fast serial equalization.*

Fast serial equalization

Many techniques that leverage the narrowly quasi-banded structure of $\mathbf{Q}$ can be classified as *fast serial equalization* techniques. These techniques avoid the $K \times K$ matrix operations (e.g., inversion and LDU decomposition) specified in Section 6.3.2 for, e.g., linear equalization (6.24)-(6.26), decision feedback equalization (6.29)-(6.30), tree-search based equalization (6.35), and iterative soft equalization (6.49). Instead, the fast serial techniques work on the *local observation model*

$$\mathbf{y}_k \quad = \quad \mathbf{Q}_k \mathbf{a}_k + \mathbf{z}_k \tag{6.56}$$

when estimating the symbol $a[k]$ (or any coded bits represented by $a[k]$) for $k = 0, \ldots, K-1$. Here, $\mathbf{y}_k \triangleq \big(y[k-D] \quad \cdots \quad y[k+D]\big)^T$ and $\mathbf{a}_k \triangleq \big(a[k-2D] \quad \cdots \quad a[k+2D]\big)^T$ are illustrated in Fig. 6.6, along with $\mathbf{Q}_k$ and $\mathbf{z}_k$. The principal idea behind the local model is the following. *Since $a[k]$ affects only the local observations $\mathbf{y}_k \in \mathbb{C}^{2D+1}$ within $\mathbf{y} \in \mathbb{C}^K$, use only these local observations to estimate $a[k]$*. We can thus think of $\mathbf{Q}_k \in \mathbb{C}^{(2D+1)\times(4D+1)}$ as the "*local* effective channel matrix." It is usually convenient to increment the index k in steps of 1, so that estimation is performed *serially*, i.e., one symbol at a time. And sometimes it helps to start over at $k = 0$ after $k = K-1$ has been reached. Notice that, due to the corner support regions of the quasi-banded $\mathbf{Q}$ in Fig. 6.6, the local observation window shifts cyclically within $\mathbf{y}$.

To our knowledge, Jeon, Chang, and Choo [JCC99] were the first to apply this fast serial approach to the equalization of rapidly TV channels. In particular, they proposed an LMMSE approximation that required only $\mathscr{O}(KD^3)$ operations per block. Note that, when $D \ll K$, their approach is much cheaper than standard $\mathscr{O}(K^3)$ LMMSE, i.e., (6.24)-(6.25). Cai and Giannakis [CG03] proposed LMMSE and MMSE-DFE extensions of [JCC99] where the inversion of the $(2D+1)\times(2D+1)$ covariance matrix $\mathbf{C}_{\mathbf{y}_k}$ was accomplished using a rank-one update. However, their schemes require $\mathscr{O}(K^2D)$ operations per block, where the quadratic dependence on K remained as a result of not fully exploiting the quasi-banded property of $\mathbf{Q}$. Barhumi, Leus, and Moonen [BLM04, BLM06] proposed generalized $\mathscr{O}(K^2D^2)$ per-tone linear equalization schemes that allowed oversampling in the frequency domain.

Hunziker and Dahlhaus [HD03] proposed an iterative approximation to ML symbol detection in which the likelihoods of individual symbols were serially maximized assuming tentative hard decisions on the other symbols. To reduce error propagation, they initialized using an approximation of LZF that was implemented serially using Gauss-Seidel iterations. Schniter and Das [Sch04, DS07] proposed iterative soft equalization based on the local observation model (6.56), requiring only $\mathscr{O}(KD^3)$ operations per block iteration. As discussed in Section 6.3.2, a well-designed iterative soft equalizer is effective at preventing error propagation and can be used in a turbo configuration, as in [DS07]. A similar iterative soft equalization scheme was proposed later by Peng and Ryan [PR06].

Fast joint equalization

Fast joint equalization techniques have also been proposed for the coherent equalization of narrowly quasi-banded versions of $\mathbf{Q}$ that result when time-frequency concentrated modulation/demodulation is used with rapidly TV channels. As opposed to serial equalization schemes, which estimate the symbols in $\mathbf{a}$ one-at-a-time, joint equalization schemes estimate the K symbols in $\mathbf{a}$ jointly.

Early joint techniques assumed not only that $\mathbf{Q}$ is narrowly quasi-banded, but also that the off-diagonal coefficients within the support region of $\mathbf{Q}$ are themselves relatively small. For example, iterative LZF approximation techniques that require $\mathscr{O}(KD)$ operations per block iteration where proposed by Toeltsch and Molisch [TM01] and Guillaud and Slock [GS03]. Gorokhov and Linnartz [GL04] proposed $\mathscr{O}(KD)$ approximate LMMSE and DFE-like schemes using a first-order Taylor series approximation of the $K \times K$ LMMSE matrix inverse. Tomasin, Gorokhov, Yang, and Linnartz [TGYL05] extended the techniques in [GL04] to incorporate iterative hard interference cancellation. Hou and Chen [HC05] proposed an $\mathscr{O}(KD^2)$ nonlinear estimator of the form $\hat{\mathbf{a}} = \mathbf{E}_{\mathsf{FF}}\mathbf{y} - \mathbf{E}_{\mathsf{FB}}\mathscr{D}_{\mathscr{A}}(\mathbf{E}_{\mathsf{FF}}\mathbf{y})$, where $\mathbf{E}_{\mathsf{FF}}$ and $\mathbf{E}_{\mathsf{FB}}$ are both narrowly banded. Note that, in [HC05], quantization is performed on the linear estimates $\mathbf{E}_{\mathsf{FF}}\mathbf{y}$ rather than (causally) on the final estimates $\hat{\mathbf{a}}$, as in DFE (6.27).

More recently, Rugini, Banelli and Leus proposed $\mathscr{O}(KD^2)$ exact LMMSE [RBL05] and MMSE-DFE [RBL06] schemes for narrowly banded[16] $\mathbf{Q}$ based on fast LDU decomposition. Furthermore, they showed how to design a receiver window to ensure that the noise covariance $\mathbf{C}_{\mathbf{z}}$ is quasi-banded, making the observation covariance $\mathbf{C}_{\mathbf{y}} = \sigma_a^2\mathbf{Q}^H\mathbf{Q} + \mathbf{C}_{\mathbf{z}}$ banded as well. These non-approximate LMMSE and MMSE-DFE equalizers are expected to outperform their approximate counterparts. (See Chapter 7 for more details.)

Joint MLSD-based schemes exploiting the quasi-banded structure of $\mathbf{Q}$ have also been proposed. For example, Matheus and Kammeyer [MK97] applied the Viterbi algorithm to implement exact MLSD on a $(2D+1)$-banded $\mathbf{Q}$ with complexity $\mathscr{O}(KD|\mathscr{A}|^{2D+1})$. The same idea was re-invented in later works, e.g., [SAMT05]. For typical values of $|\mathscr{A}|$ and D, however, the complexity of Viterbi equalization can be orders-of-magnitude higher than that of MMSE-DFE. Thus, Hwang and Schniter [HS06] investigated tree-search approaches to approximate MLSD. Their techniques use fast MMSE-DFE pre-processing,

[16] With minor modifications, the fast LDU decomposition that Rugini, Banelli, and Leus developed for banded matrices (i.e., those matching Fig. 6.5(b)) can be extended to quasi-banded matrices (i.e., those matching Fig. 6.5(a)) [HS06].

costing $\mathcal{O}(KD^2)$ operations, followed by a tree-search that employs a fast metric update and is tuned to the V-shaped structure of the upper triangular matrix $\mathbf{\Delta}_{\text{MMSE-DFE}}^{1/2}\mathbf{U}_{\text{MMSE-DFE}}$ in the causal model (6.33). (Recall the V-shaped illustration in Fig. 6.5(c).) The resulting scheme has approximately the same complexity as the fast MMSE-DFE from [RBL06], yet results in performance that is almost indistinguishable from MLSD.

MAP schemes exploiting the quasi-banded structure of $\mathbf{Q}$ have also been proposed. For $(2D+1)$-banded $\mathbf{Q}$, e.g., Liu and Fitz [LF07] used reduced-state sequence estimation [EQ88] to compute approximate soft bit estimates while Hwang and Schniter [HS09] applied tree-search (with a fast metric update). Finally, using the technique of Tüchler, Koetter, and Singer [TKS02], it is straightforward to translate any set of LMMSE estimates into soft bit estimates. Leveraging this idea, Fang and Leus [FRL08] turned the fast joint LMMSE estimation scheme of [RBL05] into a soft bit estimation scheme.

Other approaches to equalization for time-frequency concentrated schemes

For completeness, we mention two other schemes proposed for the equalization of channels yielding a narrowly quasi-banded $\mathbf{Q}$. The paper by Choi, Voltz, and Cassara [CVC01] was among the first to consider equalization for multicarrier modulation over doubly selective (i.e., time- and frequency-selective) channels, and it proposed ZF, LMMSE, and ZF-DFE schemes for doing so. However, these schemes consumed $\mathcal{O}(K^3)$ operations per block because the narrowly quasi-banded structure of $\mathbf{Q}$ was not leveraged. For the same application, Stamoulis, Diggavi, and Al-Dhahir [SDAD02] proposed an $\mathcal{O}(K^2)$ LMMSE approximation where the matrix to be inverted during each block interval is replaced by its time average. As we have seen, however, near-optimal schemes can be designed with complexities that are *linear* in K.

6.3.4 Coherent equalization for single-carrier modulation/demodulation

Recall from Section 6.2.2 that, when single-carrier modulation/demodulation is used, the effective channel matrix $\mathbf{Q}$ falls into the "widely quasi-banded" class. Here, $\mathbf{Q}$ has only negligible coefficients outside the shaded region in Fig. 6.2(a), where M denotes the discrete channel delay spread. Since, in this case, $\mathbf{Q}$ is quasi-banded, the equalization techniques described in Section 6.3.3 can *in principle* be applied here as well (e.g., [ASLC06]). However, this approach will only be practical when M is small. Since M is often large (e.g., in the hundreds), there is good reason to study equalization schemes whose complexities are robust to large M.

Frequency-domain equalization

Frequency-domain equalization (FDE) [FABSE02] is one approach to make the equalization complexity of single-carrier schemes reasonable when M is large. To describe FDE, we focus on the case of CP-SCM modulation/demodulation, assuming adequate guard length (i.e., $N_g \geq M-1$) and white noise (i.e., $\mathbf{C_z} = \sigma_z^2\mathbf{I}_K$). The first step of FDE is transformation of the observations $\mathbf{y}$ to the frequency domain. Denoting the $K \times K$

unitary discrete Fourier transform (DFT) matrix by $\mathbf{W}$ and using under-bars to identify frequency-domain vectors (e.g., $\underline{\mathbf{y}} \triangleq \mathbf{W}\mathbf{y}$, $\underline{\mathbf{a}} \triangleq \mathbf{W}\mathbf{a}$, and $\underline{\mathbf{z}} \triangleq \mathbf{W}\mathbf{z}$), it follows from (6.2) that

$$\underline{\mathbf{y}} \;=\; \underline{\mathbf{Q}}\underline{\mathbf{a}} + \underline{\mathbf{z}} \tag{6.57}$$

where $\underline{\mathbf{Q}} = \mathbf{W}\mathbf{Q}\mathbf{W}^H$ and where $\underline{\mathbf{z}}$ also has covariance $\sigma_z^2\mathbf{I}_K$. The second step of FDE is estimation of $\underline{\mathbf{a}}$ from $\underline{\mathbf{y}}$. While the elements of $\mathbf{a}$ belong to a finite alphabet, the elements of $\underline{\mathbf{a}}$ do not, and hence the estimator must be linear. One option is LMMSE estimation, i.e., $\hat{\underline{\mathbf{a}}}_{\mathsf{LMMSE}} = (\underline{\mathbf{Q}}^H\underline{\mathbf{Q}} + \frac{\sigma_z^2}{\sigma_a^2}\mathbf{I}_K)^{-1}\underline{\mathbf{Q}}^H\underline{\mathbf{y}}$. The third and final step of FDE is transformation of $\hat{\underline{\mathbf{a}}}$ back to the time domain, yielding the symbol estimates $\hat{\mathbf{a}} \triangleq \mathbf{W}^H\hat{\underline{\mathbf{a}}}$. If the DFTs are implemented using radix-2 FFTs, they will consume only $\mathscr{O}(K\log_2 K)$ operations.

With a time-invariant channel, the use of CP-SCM makes $\mathbf{Q}$ circulant (recalling Section 6.2.2) and hence $\underline{\mathbf{Q}}$ diagonal. In this case, the LMMSE estimation step consumes only $\mathscr{O}(K)$ operations (because the matrix to invert is diagonal), and FDE consumes $\mathscr{O}(K\log_2 K)$ operations in total. Note that, for large M, FDE would be significantly cheaper than LMMSE estimation of $\mathbf{a}$ from $\mathbf{y}$ via fast LDU [RBL05] (as discussed in Section 6.3.3), which consumes $\mathscr{O}(KM^2)$ operations, where typically $M \approx K/4$.

With a TV channel, $\mathbf{Q}$ will not be circulant, and thus $\underline{\mathbf{Q}}$ will not be diagonal. In this case, the off-diagonal terms of $\underline{\mathbf{Q}}$ will be non-zero, complicating the estimation of $\underline{\mathbf{a}}$ from $\underline{\mathbf{y}}$. In fact, the interference power profile of $\underline{\mathbf{Q}}$ with CP-SCM is identical to that of $\mathbf{Q}$ with CP-OFDM, which (as shown in Fig. 6.4) decays quite slowly with distance from the diagonal. However, through the application of time-domain windowing[17] at the demodulator [SL03], it is possible to give $\underline{\mathbf{Q}}$ the narrowly quasi-banded support of Fig. 6.2(b), in which case any of the fast *linear* equalization techniques described in Section 6.3.3 can be used to estimate $\underline{\mathbf{a}}$ from $\underline{\mathbf{y}}$. For example, in [TL08], Tang and Leus proposed a method to equalize a single carrier system using the OFDM fast LMMSE technique [RBL05].

Because $\underline{\mathbf{a}}$ does not have a finite-alphabet structure, the trellis, DFE, and tree-search based techniques discussed in Section 6.3.3 are not directly applicable to the estimation of $\underline{\mathbf{a}}$. Iterative soft equalization, however, is applicable. We now summarize the approach proposed by Schniter and Liu in [SL03]. First, the fast serial iterative soft equalization technique of [Sch04] is used to compute the LMMSE interference-canceled estimate $\hat{\underline{\mathbf{a}}}$ from the frequency-domain windowed observations $\underline{\mathbf{y}}$ (given current estimates of the time-domain symbol means and variances). Next, the estimates $\hat{\underline{\mathbf{a}}}$ are transformed to the time domain via $\hat{\mathbf{a}} = \mathbf{W}^H\hat{\underline{\mathbf{a}}}$, from which posterior LLRs are calculated for each of the bits $c[j]$. The posterior LLR computation is more complicated than (6.54)-(6.55), though, due to the correlation that results from the time-frequency transformation. Finally, the posterior LLRs are used as priors in the next iteration, which begins by re-calculating the time-domain symbol means and variances. In [SL03], a fast algorithm for the entire procedure was derived that consumes only $\mathscr{O}(D^2K\log K)$ operations per block iteration. Ng and Falconer [NF04] later extended the technique of [SL03] to include widely linear estimation (though they neglected the receiver windowing step).

Though the windowed FDE method above focuses on CP-SCM, similar techniques can be applied to ZP-SCM under appropriate processing of the received guard samples.

[17] With time-domain windowing, $\underline{\mathbf{y}} = \mathbf{W}\mathbf{\Delta}\mathbf{y}$ and $\underline{\mathbf{Q}} = \mathbf{W}\mathbf{\Delta}\mathbf{Q}\mathbf{W}^H$ for suitably chosen diagonal $\mathbf{\Delta}$.

For single-carrier modulation *without* a prefix, the use of *IBI-cancellation and cyclic-prefix reconstruction* [KS98] enables the application of the CP-SCM-based windowed FDE methods discussed above, as demonstrated by Schniter and Liu in [SL04].

Other approaches to equalization for single-carrier schemes

Barhumi, Leus, and Moonen [BLM05] proposed a CP-SCM equalization technique based on LTV filters whose time-variations were constrained to obey a (possibly oversampled) complex-exponential basis expansion model of order $I-1$. Under these constraints, LZF and LMMSE equalizers, requiring $\mathcal{O}(KI^3M^3)$ operations per block, were designed. However, due to the cubic complexity in M, these schemes are much more expensive than frequency-domain equalization when M is large.

6.4 Noncoherent equalization

In Section 6.3 we discussed the coherent approach to equalization, i.e., estimation of $\mathbf{a}$ from $\mathbf{y}$ in (6.2), where the channel $\mathbf{H}$, and hence the effective channel $\mathbf{Q}$, was assumed to be known. Here we discuss *noncoherent equalization*, where the channel realization $\mathbf{H}$ is unknown but its statistics may be known.

In Section 6.4.1, we rewrite the system model in a form that is more convenient for noncoherent equalization. Then, in Sections 6.4.2 and 6.4.3, we describe criteria and algorithms for noncoherent equalization, respectively. Finally, in Sections 6.4.4–6.4.5, we describe specific strategies suitable for the noncoherent equalization of rapidly TV channels.

6.4.1 Noncoherent system model

Since the effective channel $\mathbf{Q}$ is now unknown, it helps to re-formulate the system model developed in Section 6.2 into a more convenient form. In particular, we rewrite (6.2) using an efficient parameterization for the entries of the matrix[18] $\mathbf{Q}$. To do this, we build a *basis expansion model* (BEM) for the trajectories of the effective channel coefficients that make up $\mathbf{Q}$, and then we write the observation in terms of these BEM coefficients.

From (6.2), we can see that

$$y[l] = \sum_{k=0}^{K-1} [\mathbf{Q}]_{l,k} a[k] + z[l] = \sum_{d=l}^{l-K+1} [\mathbf{Q}]_{l,l-d} a[l-d] + z[l] \tag{6.58}$$

$$= \sum_{d=0}^{K-1} q[l,d] a[l-d] + z[l] \tag{6.59}$$

[18] We choose to parameterize the entries of $\mathbf{Q}$ rather than those of $\mathbf{H}$ to avoid explicitly defining the modulation and demodulation operations.

for $q[l,d] \triangleq [\mathbf{Q}]_{l,\langle l-d\rangle_K}$ and where the index of $a[\cdot]$ in (6.59), and henceforth, is taken modulo-K. Here, $\langle j\rangle_K$ denotes "j modulo K." Notice that (6.59) expresses the relationship between $\{a[l]\}$ and $\{y[l]\}$ in exactly the same way as (6.1) expressed the relationship between $\{s[n]\}$ and $\{r[n]\}$: using a TV convolution. In fact, when the modulation and demodulation operations are trivial, as in single-carrier modulation, we have $q[l,d] = h[l,d]$. Due to the support of $\mathbf{Q}$, as described in Section 6.2.2, the summation range in (6.59) can be truncated to $d \in \{-D,\ldots,D\}$ for narrowly quasi-banded $\mathbf{Q}$ and to $d \in \{0,\ldots,M-1\}$ for widely quasi-banded $\mathbf{Q}$. In this section, we will assume the general case that $d \in \{0,\ldots,N_q-1\}$, so that the widely quasi-banded case follows directly from $N_q = M$ and the narrowly quasi-banded case follows from $N_q = 2D+1$ after cyclically left-shifting the columns of $\mathbf{Q}$ by D places. (Recall Fig. 6.2.)

While BEMs are usually applied to the channel impulse response trajectories $\{h[n,m]\}_{n=0}^{N-1}$ (e.g., [TG96]), here we apply a BEM to the *effective* channel impulse response trajectory, which includes the effects of modulation/demodulation. In particular, we model the d^{th} trajectory $\{q[l,d]\}_{l=0}^{K-1}$ using the BEM coefficients $\{\theta[i,d]\}_{i=0}^{I-1}$ and basis waveforms constructed from $\{\beta[l,i]\}$, as follows:

$$q[l,d] \quad = \quad \sum_{i=0}^{I-1} \beta[l,i]\theta[i,d] \quad \text{for } l = 0,\ldots,K-1. \tag{6.60}$$

If one prefers not to use a BEM, then the *trivial BEM*, specified by $I = K$ and $\beta[l,i] = \delta[l-i]$, where $\delta[\cdot]$ denotes the Kronecker delta, guarantees $\theta[l,d] = q[l,d]\ \forall l,d$. Using $\boldsymbol{\theta}_d \triangleq \begin{pmatrix}\theta[0,d] & \cdots & \theta[I-1,d]\end{pmatrix}^T \in \mathbb{C}^I$ and $\boldsymbol{\beta}_l \triangleq \begin{pmatrix}\beta[l,0] & \cdots & \beta[l,I-1]\end{pmatrix}^H \in \mathbb{C}^I$, we have $q[l,d] = \boldsymbol{\beta}_l^H \boldsymbol{\theta}_d$ and hence equation (6.59) can be rewritten in terms of BEM quantities as

$$y[l] \quad = \quad \boldsymbol{\beta}_l^H \sum_{d=0}^{N_q-1} a[l-d]\boldsymbol{\theta}_d + z[l]. \tag{6.61}$$

Collecting the demodulator outputs $\{y[l]\}_{l=0}^{k}$ in a vector, (6.61) implies

$$\underbrace{\begin{pmatrix} y[0] \\ \vdots \\ y[k] \end{pmatrix}}_{\triangleq \mathbf{y}_k} = \underbrace{\begin{pmatrix} a[0]\boldsymbol{\beta}_0^H & \cdots & a[-N_q+1]\boldsymbol{\beta}_0^H \\ \vdots & & \vdots \\ a[k]\boldsymbol{\beta}_k^H & \cdots & a[k-N_q+1]\boldsymbol{\beta}_k^H \end{pmatrix}}_{\triangleq \boldsymbol{\Lambda}_{\mathbf{a}_k}} \underbrace{\begin{pmatrix} \boldsymbol{\theta}_0 \\ \vdots \\ \boldsymbol{\theta}_{N_q-1} \end{pmatrix}}_{\triangleq \boldsymbol{\theta}} + \underbrace{\begin{pmatrix} z[0] \\ \vdots \\ z[k] \end{pmatrix}}_{\triangleq \mathbf{z}_k}, \tag{6.62}$$

summarized by

$$\mathbf{y}_k \quad = \quad \boldsymbol{\Lambda}_{\mathbf{a}_k}\boldsymbol{\theta} + \mathbf{z}_k. \tag{6.63}$$

The matrix $\boldsymbol{\Lambda}_{\mathbf{a}_k}$ is constructed from the partial symbol vector $\mathbf{a}_k \triangleq \begin{pmatrix}a[0] & \cdots & a[k]\end{pmatrix}^T$ and the BEM waveforms $\{\boldsymbol{\beta}_l\}_{l=0}^{k}$. Notice that $\mathbf{a}_{K-1} = \mathbf{a}$, $\mathbf{y}_{K-1} = \mathbf{y}$, and $\mathbf{z}_{K-1} = \mathbf{z}$ for the previously defined vectors $\mathbf{a}$, $\mathbf{y}$, and $\mathbf{z}$. Notice also that, in (6.63), the channel realization is represented by the BEM coefficient vector $\boldsymbol{\theta}$.

Throughout this section, we assume (for simplicity) that the symbol block is zero-prefixed, i.e., $a[l] = 0$ for $l \in \{-N_q + 1, \ldots, 0\}$. Note that, due to the cyclic indexing assumption on $a[\cdot]$, this implies that $a[l] = 0$ for $l \in \{K - N_q + 1, \ldots, K - 1\}$. With this assumption, (6.61) yields a causal relationship between the symbols and the demodulator outputs, i.e., $\{y[l]\}_{l \leq k}$ depends only on $\{a[l]\}_{l \leq k}$.

We also assume that the channel is *Rayleigh fading*, i.e., that the impulse response coefficients are zero-mean[19] Gaussian distributed. Due to the linearity of modulation/demodulation and basis expansion modeling, this implies that the BEM coefficients $\boldsymbol{\theta}$ will also be zero-mean Gaussian. All other model assumptions stated in Section 6.2 apply here as well.

6.4.2 Noncoherent equalization criteria

In this section, we review several well-known noncoherent equalization criteria. As in our previous discussion of coherent criteria, we partition the discussion into criteria that apply to hard symbol estimates, complex-field symbol estimates, and soft bit estimates.

Before continuing, though, we discuss the important *ambiguity* phenomenon that can arise in noncoherent equalization. For example, if $C\mathbf{a} \in \mathscr{A}^K$ for some $\mathbf{a} \in \mathscr{A}^K$ and some $C \neq 1$, then it is impossible to distinguish, from the output $\mathbf{y}$, between the hypotheses $(\mathbf{a}, \boldsymbol{\theta})$ and $(C\mathbf{a}, C^{-1}\boldsymbol{\theta})$. Notice that C accounts for both phase and/or gain ambiguity. To prevent ambiguity, one could, e.g., use an asymmetric scalar alphabet $\mathscr{A}$ or treat a single symbol (e.g., $a[0]$) as a known pilot, so that the set of candidate symbol vectors becomes asymmetric [Har00]. In stating the criteria below, we assume that the ambiguity issue has been taken care of.

Hard symbol estimates

Similar to coherent equalization, the minimal probability of sequence error is guaranteed by *noncoherent maximum a posteriori sequence detection* (MAPSD):

$$\hat{\mathbf{a}}_{\mathsf{ncMAPSD}} \triangleq \arg\max_{\mathbf{a}' \in \mathscr{A}^K} \Pr\{\mathbf{a} = \mathbf{a}' \mid \mathbf{y}\}. \tag{6.64}$$

Note that the noncoherent posterior in (6.64) is not conditioned on the effective channel matrix $\mathbf{Q}$, as was the coherent posterior in (6.9). Similarly, noncoherent MAP symbol and bit detection are defined as

$$\hat{a}_{\mathsf{ncMAP}}[k] \triangleq \arg\max_{a \in \mathscr{A}} \Pr\{a[k] = a \mid \mathbf{y}\} \text{ for } k \in 0, \ldots, K-1 \tag{6.65}$$

$$\hat{c}_{\mathsf{ncMAP}}[j] \triangleq \arg\max_{c \in \{0,1\}} \Pr\{c[j] = c \mid \mathbf{y}\} \text{ for } j \in 0, \ldots, K\log_2|\mathscr{A}| - 1. \tag{6.66}$$

If the equalizer assumes that $\mathbf{a}$ is uniformly distributed over $\mathscr{A}^K$, then noncoherent MAPSD reduces to *noncoherent maximum likelihood sequence detection* (MLSD):

$$\hat{\mathbf{a}}_{\mathsf{ncMLSD}} \triangleq \arg\max_{\mathbf{a} \in \mathscr{A}^K} f(\mathbf{y} \mid \mathbf{a}). \tag{6.67}$$

[19] Some details of the non-zero case can be found in [RC96].

(The justification is similar to (6.13).) Due to the Rayleigh fading assumption, $f(\mathbf{y} \mid \mathbf{a})$ is a Gaussian distribution with zero mean and covariance

$$\mathbf{C}_{\mathbf{y}|\mathbf{a}} \quad = \quad \mathbf{C}_{\mathbf{z}} + \mathbf{\Lambda}_{\mathbf{a}} \mathbf{C}_{\boldsymbol{\theta}} \mathbf{\Lambda}_{\mathbf{a}}^H. \tag{6.68}$$

Given this conditional distribution for $\mathbf{y}$, (6.67) reduces to

$$\hat{\mathbf{a}}_{\mathsf{ncMLSD}} \quad = \quad \arg \min_{\mathbf{a} \in \mathscr{A}^K} \left\{ \mathbf{y}^H \mathbf{C}_{\mathbf{y}|\mathbf{a}}^{-1} \mathbf{y} + \ln(\pi^K \det\{\mathbf{C}_{\mathbf{y}|\mathbf{a}}\}) \right\}. \tag{6.69}$$

There is an interesting connection between the noncoherent MAPSD/MLSD criteria and MMSE channel estimation [Kai69, HM89]. To see this, we first write the MMSE estimate of $\boldsymbol{\theta}$ from $\mathbf{y}$ under the sequence hypothesis $\mathbf{a}$ as

$$\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}} \quad \triangleq \quad \mathsf{E}\{\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{a}\} \; = \; \mathbf{C}_{\boldsymbol{\theta},\mathbf{y}|\mathbf{a}} \mathbf{C}_{\mathbf{y}|\mathbf{a}}^{-1} \mathbf{y} \tag{6.70}$$

$$= \quad \mathbf{C}_{\boldsymbol{\theta}} \mathbf{\Lambda}_{\mathbf{a}}^H \left(\mathbf{C}_{\mathbf{z}} + \mathbf{\Lambda}_{\mathbf{a}} \mathbf{C}_{\boldsymbol{\theta}} \mathbf{\Lambda}_{\mathbf{a}}^H \right)^{-1} \mathbf{y} \tag{6.71}$$

$$= \quad \left(\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \mathbf{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{\Lambda}_{\mathbf{a}} \right)^{-1} \mathbf{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{y}, \tag{6.72}$$

where the matrix inversion lemma was used to obtain (6.72). In Appendix 6.B, we use (6.72) to show that (6.69) can be rewritten as

$$\hat{\mathbf{a}}_{\mathsf{ncMLSD}} = \arg \min_{\mathbf{a} \in \mathscr{A}^K} \left\{ \left\| \mathbf{y} - \mathbf{\Lambda}_{\mathbf{a}} \hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}} \right\|_{\mathbf{C}_{\mathbf{z}}^{-1}}^2 + \| \hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}} \|_{\mathbf{C}_{\boldsymbol{\theta}}^{-1}}^2 + \ln(\pi^K \det\{\mathbf{C}_{\mathbf{y}|\mathbf{a}}\}) \right\}. \tag{6.73}$$

Equation (6.73) states that the noncoherent MLSD metric can be written as the coherent MLSD metric based on the *implicit channel estimate* $\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}}$, plus a term that penalizes the deviation in $\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}}$ from the prior statistics on $\boldsymbol{\theta}$, plus what is sometimes referred to as a "bias" term. Thus, while the noncoherent MLSD/MAPSD estimates can be found without computing a channel estimate (as in (6.69)), they can also be found via joint channel/symbol estimation (as in (6.73)).

If the channel statistics (i.e., $\mathbf{C}_{\boldsymbol{\theta}}$) are unknown, then the noncoherent ML and MAP criteria do not apply. In this case, it may be more appropriate to employ the *generalized likelihood ratio test* (GLRT) criterion [WM02]:

$$\hat{\mathbf{a}}_{\mathsf{GLRT}} \quad \triangleq \quad \operatorname*{argmax}_{\mathbf{a} \in \mathscr{A}^K} \max_{\boldsymbol{\theta} \in \mathbb{C}^{NqI}} f(\mathbf{y} \mid \mathbf{a}, \boldsymbol{\theta}). \tag{6.74}$$

Since $\ln f(\mathbf{y} \mid \mathbf{a}, \boldsymbol{\theta}) = -\|\mathbf{y} - \mathbf{\Lambda}_{\mathbf{a}} \boldsymbol{\theta}\|_{\mathbf{C}_{\mathbf{z}}^{-1}}^2 - \ln(\pi^K \det\{\mathbf{C}_{\mathbf{z}}\})$, the GLRT sequence estimate can be expressed as

$$\hat{\mathbf{a}}_{\mathsf{GLRT}} \quad = \quad \arg \min_{\mathbf{a} \in \mathscr{A}^K} \left\| \mathbf{y} - \mathbf{\Lambda}_{\mathbf{a}} \hat{\boldsymbol{\theta}}_{\mathsf{ML}|\mathbf{a}} \right\|_{\mathbf{C}_{\mathbf{z}}^{-1}}^2, \tag{6.75}$$

where $\hat{\boldsymbol{\theta}}_{\mathsf{ML}|\mathbf{a}} \triangleq \arg\max_{\boldsymbol{\theta} \in \mathbb{C}^{NqI}} f(\mathbf{y} \mid \mathbf{a}, \boldsymbol{\theta})$ denotes the $\mathbf{a}$-conditional maximum likelihood (ML) channel estimate, i.e.,

$$\hat{\boldsymbol{\theta}}_{\mathsf{ML}|\mathbf{a}} \quad = \quad \arg \min_{\boldsymbol{\theta} \in \mathbb{C}^{NqI}} \| \mathbf{y} - \mathbf{\Lambda}_{\mathbf{a}} \boldsymbol{\theta} \|_{\mathbf{C}_{\mathbf{z}}^{-1}}^2 \tag{6.76}$$

$$= \quad \left(\mathbf{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{\Lambda}_{\mathbf{a}} \right)^{-1} \mathbf{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{y}. \tag{6.77}$$

Combining (6.77) and (6.75), the GLRT sequence estimate becomes

$$\hat{\mathbf{a}}_{\mathsf{GLRT}} \;=\; \arg\max_{\mathbf{a}\in\mathscr{A}^K} \mathbf{y}^H\mathbf{C}_{\mathbf{z}}^{-1}\boldsymbol{\Lambda}_{\mathbf{a}}\big(\boldsymbol{\Lambda}_{\mathbf{a}}^H\mathbf{C}_{\mathbf{z}}^{-1}\boldsymbol{\Lambda}_{\mathbf{a}}\big)^{-1}\boldsymbol{\Lambda}_{\mathbf{a}}^H\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{y}. \tag{6.78}$$

Notice that, when the noise is white, $\hat{\boldsymbol{\theta}}_{\mathsf{ML}|\mathbf{a}}$ in (6.77) reduces to the conditional *least-squares* (LS) channel estimate: $\hat{\boldsymbol{\theta}}_{\mathsf{LS}|\mathbf{a}} = \big(\boldsymbol{\Lambda}_{\mathbf{a}}^H\boldsymbol{\Lambda}_{\mathbf{a}}\big)^{-1}\boldsymbol{\Lambda}_{\mathbf{a}}^H\mathbf{y}$.

Finally, we note that the GLRT metric in (6.78) equals the limiting case of the $\mathbf{y}^H\mathbf{C}_{\mathbf{y}|\mathbf{a}}^{-1}\mathbf{y}$ component of the noncoherent MLSD metric in (6.69) when $\mathbf{C}_{\boldsymbol{\theta}} = \sigma_\theta^2\mathbf{I}$ and $\sigma_\theta^2 \to \infty$, i.e., when the signal is white and the noise power is negligible relative to the signal power.

Complex-field symbol estimates

The *noncoherent minimum mean-squared error* (MMSE) criterion specifies the complex-valued sequence estimate

$$\hat{\mathbf{a}}_{\mathsf{ncMMSE}} \;\triangleq\; \arg\min_{\mathbf{a}'\in\mathbb{C}^K} \mathsf{E}\big\{\|\mathbf{a}-\mathbf{a}'\|^2 \;\big|\; \mathbf{y}\big\}. \tag{6.79}$$

Note that, unlike the coherent case (6.15), the expectation in (6.79) is not conditioned on the channel $\mathbf{Q}$. Writing the MMSE estimate as the conditional mean [Poo94], we find

$$\hat{\mathbf{a}}_{\mathsf{ncMMSE}} \;=\; \mathsf{E}\{\mathbf{a}\mid\mathbf{y}\} \;=\; \sum_{\mathbf{a}\in\mathscr{A}^K}\mathbf{a}\;p(\mathbf{a}\mid\mathbf{y}) \tag{6.80}$$

$$\;=\; \sum_{\mathbf{a}\in\mathscr{A}^K}\mathbf{a}\;\frac{f(\mathbf{y}\mid\mathbf{a})p(\mathbf{a})}{\sum_{\mathbf{a}'\in\mathscr{A}^K} f(\mathbf{y}\mid\mathbf{a}')p(\mathbf{a}')} \tag{6.81}$$

$$\;=\; \sum_{\mathbf{a}\in\mathscr{A}^K}\mathbf{a}\;\frac{p(\mathbf{a})\int f(\mathbf{y}\mid\mathbf{a},\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}}{\sum_{\mathbf{a}'\in\mathscr{A}^K} p(\mathbf{a}')\int f(\mathbf{y}\mid\mathbf{a}',\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{6.82}$$

If we assume that $p(\mathbf{a})$ is uniformly distributed over $\mathscr{A}^K$, then

$$\hat{\mathbf{a}}_{\mathsf{ncMMSE}} \;=\; \frac{\sum_{\mathbf{a}\in\mathscr{A}^K}\mathbf{a}\;\int\exp\big(-\|\mathbf{y}-\boldsymbol{\Lambda}_{\mathbf{a}}\boldsymbol{\theta}\|^2_{\mathbf{C}_{\mathbf{z}}^{-1}} - \|\boldsymbol{\theta}\|^2_{\mathbf{C}_{\boldsymbol{\theta}}^{-1}}\big)d\boldsymbol{\theta}}{\sum_{\mathbf{a}'\in\mathscr{A}^K}\int\exp\big(-\|\mathbf{y}-\boldsymbol{\Lambda}_{\mathbf{a}'}\boldsymbol{\theta}\|^2_{\mathbf{C}_{\mathbf{z}}^{-1}} - \|\boldsymbol{\theta}\|^2_{\mathbf{C}_{\boldsymbol{\theta}}^{-1}}\big)d\boldsymbol{\theta}}. \tag{6.83}$$

Note that the finite-alphabet nature of $\mathbf{a}$ makes the conditional mean difficult to compute, since it requires the evaluation of $|\mathscr{A}|^K$-term summations. Unlike the coherent case, imposing constraints (e.g., linear) on the noncoherent MMSE estimator does not significantly simplify its design, and so this approach is not very popular.

Soft bit estimates

In soft noncoherent equalization (as used in, e.g., turbo equalization), the equalizer computes posterior LLRs[20] for the coded bits $c[j]$

$$L_{c|\mathbf{y}}[j] \;\triangleq\; \ln\frac{\Pr\{c[j]=1\mid\mathbf{y}\}}{\Pr\{c[j]=0\mid\mathbf{y}\}} \quad\text{for}\quad j=0,\ldots,K\log_2|\mathscr{A}|-1, \tag{6.84}$$

[20] Under the "fixed lag" constraint, the conditioning in (6.84) is performed on $\Big(y[0] \;\;\cdots\;\; y[\lceil\frac{j}{\log_2|\mathscr{A}|}\rceil+J]\Big)^T$ instead of $\mathbf{y}$ [ZFG97], where the look-ahead interval $J\geq 0$ trades off between performance and complexity.

given the a-priori LLRs defined in (6.20). Note that the posterior probabilities in (6.84) are not conditioned on the channel, unlike the coherent case (6.19). If needed, the noncoherent MAP bit estimates can be generated from the noncoherent posterior LLRs via

$$\hat{c}_{\mathsf{ncMAP}}[j] \quad = \quad \frac{1}{2}\big(1+\mathrm{sign}(L_{c|\mathbf{y}}[j])\big). \tag{6.85}$$

6.4.3 Noncoherent equalization tools

The noncoherent MLSD, MAPSD, MAP, and MMSE estimates, as outlined in Section 6.4.2, require the evaluation of $\mathscr{O}(|\mathscr{A}|^K)$ metrics if computed via brute force, which is not practical for the anticipated values of K. In this section, we review algorithms that are designed for practical noncoherent equalization.

The suboptimality of trellis-based noncoherent equalization

Among the optimal coherent MAPSD, MLSD, and symbol/bit MAP algorithms in Section 6.3.2 were trellis-based methods. We now investigate whether similar approaches exist for optimal noncoherent equalization. In doing so, we use the causal[21] model summarized by (6.63). To simplify the notation, we assume BPSK, allowing us to make a direct mapping between each bit and symbol, e.g., $a[j] = 2b[j]-1$ for $j = 0,\ldots,K-1$, where $a[j] \in \mathscr{A} = \{-1,+1\}$. Finally, we allow prior beliefs on the bits (and thus symbols) in the form of a-priori LLRs $\{L_c[k]\}_{k=0}^{K-1}$, defined in (6.20) and simplified for BPSK symbols in (6.38).

Analogous to the coherent MAP sequence metric (6.34), we now define a *noncoherent MAP sequence metric*. In particular, we define a *partial* noncoherent MAP sequence metric that depends on the partial observation $\mathbf{y}_k$ from (6.63) and the partial bit vector $\mathbf{c}_k \triangleq \begin{pmatrix} c[0] & \cdots & c[k] \end{pmatrix}^T$:

$$\zeta_{\mathsf{nc}}(\mathbf{c}_k) \quad \triangleq \quad \ln f(\mathbf{y}_k \mid \mathbf{c}_k) + \mathbf{l}_k^T \mathbf{c}_k. \tag{6.86}$$

Here, $\mathbf{l}_k \triangleq \begin{pmatrix} L_c[0] & \cdots & L_c[k] \end{pmatrix}^T$ is a partial version of the a-priori LLR vector $\mathbf{l}_c$ defined just after (6.35). Notice that the complete noncoherent MAP sequence metric $\zeta_{\mathsf{nc}}(\mathbf{c})$ is obtained when $k = K-1$. If we are interested in noncoherent MLSD rather than noncoherent MAPSD, then we would use $\zeta_{\mathsf{nc}}(\mathbf{c})$ with $\mathbf{l}_c = \mathbf{0}$, and if we are interested in the GLRT criterion, then would additionally assume that $\mathbf{C}_{\boldsymbol{\theta}} = \sigma_\theta^2 \mathbf{I}$ with $\sigma_\theta^2 \to \infty$.

Under the Rayleigh fading assumption, $f(\mathbf{y}_k \mid \mathbf{c}_k)$ is Gaussian with zero mean and covariance $\mathbf{C}_{\mathbf{y}_k|\mathbf{a}_k} = \mathbf{C}_{\mathbf{z}_k} + \mathbf{\Lambda}_{\mathbf{a}_k} \mathbf{C}_{\boldsymbol{\theta}} \mathbf{\Lambda}_{\mathbf{a}_k}^H$, where $\mathbf{a}_k$ denotes the BPSK symbol vector corresponding to the bit vector $\mathbf{c}_k$. Thus, we have

$$\zeta_{\mathsf{nc}}(\mathbf{c}_k) \quad = \quad -\mathbf{y}_k^H \mathbf{C}_{\mathbf{y}_k|\mathbf{a}_k}^{-1} \mathbf{y}_k - \ln(\pi^{k+1} \det\{\mathbf{C}_{\mathbf{y}_k|\mathbf{a}_k}\}) + \mathbf{l}_k^T \mathbf{c}_k, \tag{6.87}$$

[21] Note that this causal model is different from the one used for DFE in Section 6.3.2. For DFE, $\mathbf{a}$ was estimated *backwards* from the last symbol, whereas here $\mathbf{a}$ is estimated *forwards* from the first symbol.

which reduces to the noncoherent MLSD metric in (6.69) when $\mathbf{l}_k = \mathbf{0}$. Using the matrix inversion lemma, (6.87) can be rewritten as

$$\zeta_{\mathsf{nc}}(\mathbf{c}_k) \;=\; -\mathbf{y}_k^H \mathbf{C}_{\mathbf{z}_k}^{-1} \mathbf{\Lambda}_{\mathbf{a}_k} \mathbf{\Sigma}_{\mathbf{a}_k}^{-1} \mathbf{\Lambda}_{\mathbf{a}_k}^H \mathbf{C}_{\mathbf{z}_k}^{-1} \mathbf{y}_k - \ln(\pi^{k+1} \det\{\mathbf{C}_{\mathbf{y}_k|\mathbf{a}_k}\}) + \mathbf{l}_k^T \mathbf{c}_k \quad (6.88)$$

with $\mathbf{\Sigma}_{\mathbf{a}_k} \triangleq \mathbf{C}_{\boldsymbol{\theta}}^{-1} + \mathbf{\Lambda}_{\mathbf{a}_k}^H \mathbf{C}_{\mathbf{z}_k}^{-1} \mathbf{\Lambda}_{\mathbf{a}_k}$. From (6.63), it follows that the partial observation can be decomposed as

$$\mathbf{y}_k \;=\; \begin{pmatrix} \mathbf{y}_{k-1} \\ y[k] \end{pmatrix} \;=\; \begin{pmatrix} \mathbf{\Lambda}_{\mathbf{a}_{k-1}} \\ \boldsymbol{\lambda}_{\mathbf{a}_k}^H \end{pmatrix} \boldsymbol{\theta} + \begin{pmatrix} \mathbf{z}_{k-1} \\ z[k] \end{pmatrix}, \quad (6.89)$$

where $\boldsymbol{\lambda}_{\mathbf{a}_k}^H$ is the last row of $\mathbf{\Lambda}_{\mathbf{a}_k}$. When the noise is white (i.e., $\mathbf{C}_{\mathbf{z}_k} = \sigma_z^2 \mathbf{I}_{k+1}$), it can then be shown that (6.88) and (6.89) can be combined to write the noncoherent metric *recursively*:

$$\begin{aligned} \zeta_{\mathsf{nc}}(\mathbf{c}_k) \;&=\; \zeta_{\mathsf{nc}}(\mathbf{c}_{k-1}) + L_c[k]c[k] + \ln \eta_{\mathbf{a}_k} \quad (6.90) \\ &- \begin{pmatrix} \boldsymbol{\lambda}_{\mathbf{a}_k} y[k] \\ \hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}_{k-1}} \end{pmatrix}^H \begin{pmatrix} \sigma_z^{-2} \mathbf{\Sigma}_{\mathbf{a}_k}^{-1} & \mathbf{I} - \eta_{\mathbf{a}_k} \mathbf{\Sigma}_{\mathbf{a}_{k-1}}^{-1} \boldsymbol{\lambda}_{\mathbf{a}_k} \boldsymbol{\lambda}_{\mathbf{a}_k}^H \\ \mathbf{I} - \eta_{\mathbf{a}_k} \boldsymbol{\lambda}_{\mathbf{a}_k} \boldsymbol{\lambda}_{\mathbf{a}_k}^H \mathbf{\Sigma}_{\mathbf{a}_{k-1}}^{-1} & -\eta_{\mathbf{a}_k} \sigma_z^2 \boldsymbol{\lambda}_{\mathbf{a}_k} \boldsymbol{\lambda}_{\mathbf{a}_k}^H \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_{\mathbf{a}_k} y[k] \\ \hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}_{k-1}} \end{pmatrix}. \end{aligned}$$

In (6.90), $\eta_{\mathbf{a}_k} \triangleq (\sigma_z^2 + \boldsymbol{\lambda}_{\mathbf{a}_k}^H \mathbf{\Sigma}_{\mathbf{a}_{k-1}}^{-1} \boldsymbol{\lambda}_{\mathbf{a}_k})^{-1}$ and $\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}_{k-1}}$ denotes the MMSE estimate of $\boldsymbol{\theta}$ from $\mathbf{y}_{k-1}$ under the sequence hypothesis $\mathbf{a}_{k-1}$.

From (6.90) we can make a two important observations. First, we know that a trellis-based implementation exists only if the metric update depends on a *fixed* number of past symbols. While $\boldsymbol{\lambda}_{\mathbf{a}_k}$ depends only on the past N_q symbols $\{a[k], \ldots, a[k-N_q+1]\}$, the terms $\mathbf{\Sigma}_{\mathbf{a}_k}^{-1}$ and $\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}_{k-1}}$ depend, in general, on the full sequence $\mathbf{a}_k$, implying that optimal noncoherent MAPSD/MLSD/GLRT cannot be implemented by a trellis-based technique. Second, when the BEM coefficient trajectories (i.e., $\{\theta[i,d]\}_{i=0}^{I-1}$ for each d) satisfy an order-N_{AR} Gauss-Markov model and the trellis has $|\mathscr{A}|^{N_q+N_{\mathsf{AR}}}$ states, the Kalman filter can be used to recursively compute the MMSE channel estimate $\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}_{k-1}}$ and its error covariance, $\mathbf{\Sigma}_{\mathbf{a}_{k-1}}^{-1}$, conditioned on the symbols $\mathbf{a}_{k-1}$ that define each surviving path. Thus, while a trellis can facilitate the computation of the exact partial sequence metric, it cannot guarantee optimal pruning. The literature is not always clear about these points, however. For example, Chugg [Chu98] points out that some seminal and often-cited works (e.g., [MS79, DS94]) seem to claim that noncoherent MLSD can be implemented with a trellis, and shows precisely why this cannot be the case.

Not surprisingly, trellis-based implementations of the noncoherent MAP symbol and bit criteria (6.65)-(6.66) are also suboptimal. With the forward-backward algorithm,[22] there is no concept of surviving paths, and channel state information is required for each state of the trellis. When the channel is unknown, the trellis can be expanded so that a channel estimate can be calculated at each state, after which the forward-backward algorithm can again be applied, though not optimally: the performance (and complexity) depends on the amount of trellis expansion [GL97, DCH01, HP00, AC00].

[22] For fixed-lag MAP symbol estimates, the situation is a bit different since the forward-backward algorithm does not apply. There the posterior symbol probabilities can be calculated recursively (assuming Gauss-Markov BEM trajectories), but they require averaging over all possible past-symbol sequences and thus cannot be folded into a trellis [ZFG97].

Noncoherent equalization via per-survivor processing

While trellis-based noncoherent equalization is suboptimal, a trellis can be used for *approximate* noncoherent MAPSD/MLSD/GLRT and MAP symbol and bit estimation.

Of the many practical suboptimal trellis-based schemes that have been proposed, a good number can be classified as *per-survivor processing* (PSP) [RPT95]. There, the idea is to compute (at stage k of the trellis) a separate channel estimate $\hat{\boldsymbol{\theta}}_{\mathbf{a}_k}$ for each surviving path extension $\mathbf{a}_k$, and then evaluate a partial metric corresponding to the pair $(\mathbf{a}_k, \hat{\boldsymbol{\theta}}_{\mathbf{a}_k})$. Only the surviving path extensions leading to the best metrics are retained as survivors, after which the process repeats at the next stage of the trellis. As discussed above, an $|\mathscr{A}|^{N_q+N_{\mathsf{AR}}}$-state trellis facilitates recursive MMSE estimation of order-N_{AR} Gauss-Markov BEM trajectories via Kalman filtering, and thus recursive computation of MAPSD and MLSD partial metrics (recalling (6.73)). Thus, for that channel class, the Viterbi algorithm with per-survivor Kalman filtering provides near-optimal noncoherent MAPSD/MLSD. This idea seems to have been first proposed in Morley and Snyder [MS79] using continuous-time filtering. Lodge and Moher [LM90] considered discrete-time filters and realized that, if the observations are first whitened (which requires only a bank of LTI filters), then the metric calculation simplifies in a way that eliminates the need for Kalman filtering. This latter approach is known as the "innovations" approach. Similar ideas can be applied to fixed-lag MAP symbol/bit estimation processing, as proposed by Iltis, Shynk, and Giridhar [ISG94].

Because the complexity of these trellis-based PSP methods grows exponentially in $N_q + N_{\mathsf{AR}}$, however, PSP methods based on more general tree-searches may be more practical for near-optimal noncoherent detection, especially at high SNR, where sphere decoders can find the optimal solution without visiting many nodes. In fact, the proposal of noncoherent tree-search can already be found in early works, e.g., Dai and Shwedyk [DS94]. Notice that, with appropriate definition of the metric, the tree-search methods discussed in Section 6.3.2—in the context of coherent equalization—apply here too, except that the pre-processing used there now becomes unnecessary because the model (6.63) is already causal. A further advantage of tree-search is that it does not require BEM coefficients to satisfy a Gauss-Markov property, which can be useful in, e.g., multi-carrier applications. In any case, the key to a computationally efficient tree-search is minimizing both the number of nodes visited and the complexity per visited node. In regards to the latter, Hwang and Schniter [HS07b] have shown that the quantities $\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}_{k-1}}$ and $\boldsymbol{\Sigma}^{-1}_{\mathbf{a}_k}$ can be updated recursively (for the generic modulation/demodulation and BEM setup of (6.63)), yielding an $\mathscr{O}\left(N_q^2 I^2\right)$ update to the noncoherent metric (6.90).

Iterative noncoherent equalization via the EM algorithm

The *expectation-maximization* (EM) algorithm [DLR77] is a well known approach for ML estimation in the presence of "missing data." If $\mathbf{y}$ is the observation, $\mathbf{x}$ is the vector to be estimated, and $\mathbf{u}$ is the "missing data," then the EM algorithm attempts to find $\hat{\mathbf{x}}_{\mathsf{ML}} = \arg\max_{\mathbf{x}} f(\mathbf{y} \mid \mathbf{x}) = \arg\max_{\mathbf{x}} \ln f(\mathbf{y} \mid \mathbf{x})$ iteratively using the following recursion[23]

[23] If the missing data $\mathbf{u}$ was discrete, integration in (6.91) would be replaced by summation.

(where i denotes the iteration index):

$$\hat{\mathbf{x}}^{(i+1)} = \arg\max_{\mathbf{x}} \int f(\mathbf{u} \mid \mathbf{y}, \hat{\mathbf{x}}^{(i)}) \ln f(\mathbf{y}, \mathbf{u} \mid \mathbf{x}) d\mathbf{u}. \tag{6.91}$$

In Appendix 6.C, we show that (6.91) arises from the goal of maximally increasing the likelihood at each iteration. So called *generalized EM* algorithms, which increase but do not necessarily maximally increase the likelihood at each iteration, have also been proposed (e.g., [FH94]). The EM recursion is sometimes regarded as having two separate steps: an "E step" which computes the conditional expectation (i.e., the integral) in (6.91), and an "M step" which performs the maximization in (6.91). A *Bayesian EM* (EMB) algorithm, with the goal to find $\hat{\mathbf{x}}_{\mathsf{MAP}} = \max_{\mathbf{x}} f(\mathbf{x} \mid \mathbf{y})$, follows by direct extension. Using Bayes rule and disregarding irrelevant terms, we can write $\hat{\mathbf{x}}_{\mathsf{MAP}} = \arg\max_{\mathbf{x}} \big(\ln f(\mathbf{y} \mid \mathbf{x}) + \ln f(\mathbf{x}) \big)$ and attempt to find $\hat{\mathbf{x}}_{\mathsf{MAP}}$ using the recursion

$$\hat{\mathbf{x}}^{(i+1)} = \arg\max_{\mathbf{x}} \left\{ \int f(\mathbf{u} \mid \mathbf{y}, \hat{\mathbf{x}}^{(i)}) \ln f(\mathbf{y}, \mathbf{u} \mid \mathbf{x}) d\mathbf{u} + \ln f(\mathbf{x}) \right\}. \tag{6.92}$$

One can immediately think of two ways that the EM(B) algorithms could be applied to noncoherent equalization: i) the coded bits could be estimated while treating the channel as missing (i.e., "info EM(B)") [GH97], or ii) the channel could be estimated, while treating the data as missing, and later used for coherent sequence detection (i.e., "channel EM(B)") [KV94, AHFF97, CT01, CV01, YR03, NP03, NL05]. EM(B) algorithms that treat both channel and data values as parameters to be estimated have also been proposed, e.g., [ZJP99].

We now describe the *channel EM(B)* algorithm for noncoherent equalization. (See Appendix 6.D for a discussion of the less practical *info EM(B)* algorithm.) For this, we use the model $\mathbf{y} = \mathbf{\Lambda_a} \boldsymbol{\theta} + \mathbf{z}$ from (6.63), where we once again find it convenient to assume BPSK in order to ensure a one-to-one correspondence between symbols $\mathbf{a}$ and bits $\mathbf{c}$. For noncoherent equalization, $\boldsymbol{\theta}$ is the vector to estimate and $\mathbf{c}$ is the missing data, so that (from (6.92)) channel EMB[24] performs the recursion

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \arg\max_{\hat{\boldsymbol{\theta}}} \left\{ \sum_{\mathbf{c} \in \{0,1\}^K} p(\mathbf{c} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(i)}) \ln f(\mathbf{y}, \mathbf{c} \mid \hat{\boldsymbol{\theta}}) + \ln f(\hat{\boldsymbol{\theta}}) \right\}. \tag{6.93}$$

Using the property $\ln f(\mathbf{y}, \mathbf{c} \mid \boldsymbol{\theta}) = \ln f(\mathbf{y} \mid \mathbf{c}, \boldsymbol{\theta}) + \ln f(\mathbf{c})$ in conjunction with the Rayleigh fading and Gaussian noise assumptions, (6.93) reduces to

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \arg\min_{\hat{\boldsymbol{\theta}}} \left\{ \sum_{\mathbf{c} \in \{0,1\}^K} p(\mathbf{c} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(i)}) \|\mathbf{y} - \mathbf{\Lambda_a} \hat{\boldsymbol{\theta}}\|^2_{\mathbf{C_z}^{-1}} + \hat{\boldsymbol{\theta}}^H \mathbf{C}_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\theta}} \right\} \tag{6.94}$$

$$= \left(\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \sum_{\mathbf{c} \in \{0,1\}^K} p(\mathbf{c} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(i)}) \mathbf{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{\Lambda}_{\mathbf{a}} \right)^{-1} \sum_{\mathbf{c} \in \{0,1\}^K} p(\mathbf{c} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(i)}) \mathbf{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{y}. \tag{6.95}$$

Above, $\mathbf{a}$ denotes the symbol vector corresponding to the bit vector $\mathbf{c}$. From (6.95), channel EM(B) can be interpreted as performing *iterative soft decision-directed channel estimation*, using soft decisions computed from the previous channel estimate. In fact, with

[24] Channel EM would yield (6.93) without the $\ln f(\hat{\boldsymbol{\theta}})$ term and hence (6.94)-(6.97) without the $\mathbf{C}_{\boldsymbol{\theta}}^{-1}$ term. This relationship is reminiscent of that between GLRT and ncMLSD.

constant modulus (CM) $\mathscr{A}$ and white noise, the summed terms in (6.95) can be rewritten using the $\tanh(\cdot)$ operator, reminiscent of (6.40). Soft decision-directed channel estimation can be contrasted with per-survivor channel estimation, as used in noncoherent MAPSD/MLSD and GLRT, in that soft decision-directed channel estimation generates a *single* channel estimate after "averaging" the soft bit estimates, whereas per-survivor channel estimation generates *multiple* channel estimates, one for each hypothesized bit sequence. Iterative soft decision-directed channel estimation has also been considered outside of the EM(B) context in, e.g., [BC98, OT04, SSS04, FRL08, LF08].

The posterior bit probabilities $\{p(\mathbf{c} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(i)})\}_{\mathbf{c}\in\{0,1\}^K}$ required for (6.95) can be obtained in various ways. For the case of BEM trajectories that satisfy a Gauss-Markov model, the forward-backward algorithm can be employed [KV94]. A different approach was proposed in [HS09] that allows the use of a-priori LLRs and more general BEM statistics. We now briefly describe this approach. Writing

$$p(\mathbf{c} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(i)}) \quad = \quad \frac{f(\mathbf{y} \mid \mathbf{c}, \hat{\boldsymbol{\theta}}^{(i)}) p(\mathbf{c})}{\sum_{\mathbf{c}'\in\{0,1\}^K} f(\mathbf{y} \mid \mathbf{c}', \hat{\boldsymbol{\theta}}^{(i)}) p(\mathbf{c}')} \quad = \quad \frac{e^{\zeta_{\text{coh}}(\mathbf{c};\hat{\boldsymbol{\theta}}^{(i)})}}{\sum_{\mathbf{c}'\in\{0,1\}^K} e^{\zeta_{\text{coh}}(\mathbf{c}';\hat{\boldsymbol{\theta}}^{(i)})}}, \quad (6.96)$$

where we have used the coherent MAP sequence metric $\zeta_{\text{coh}}(\mathbf{c}; \hat{\boldsymbol{\theta}}^{(i)}) = \ln f(\mathbf{y} \mid \mathbf{c}, \hat{\boldsymbol{\theta}}^{(i)}) + \mathbf{l}_c^T \mathbf{c}$ from (6.34) with the $\hat{\boldsymbol{\theta}}^{(i)}$-dependence explicitly noted, the EM recursion (6.95) can be restated as

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \left(\sum_{\mathbf{c}\in\{0,1\}^K} e^{\zeta_{\text{coh}}(\mathbf{c};\hat{\boldsymbol{\theta}}^{(i)})} (\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}}) \right)^{-1} \sum_{\mathbf{c}\in\{0,1\}^K} e^{\zeta_{\text{coh}}(\mathbf{c};\hat{\boldsymbol{\theta}}^{(i)})} \boldsymbol{\Lambda}_{\mathbf{a}}^H \mathbf{C}_{\mathbf{z}}^{-1} \mathbf{y}. \quad (6.97)$$

Equation (6.97) suggests iterating a soft decision-directed channel estimator, with input $\{\zeta_{\text{coh}}(\mathbf{c}; \hat{\boldsymbol{\theta}}^{(i)})\}_{\mathbf{c}\in\{0,1\}^K}$ and output $\hat{\boldsymbol{\theta}}^{(i+1)}$, and a soft-input/soft-output coherent equalizer, with input $\hat{\boldsymbol{\theta}}^{(i)}$ and output $\{\zeta_{\text{coh}}(\mathbf{c}; \hat{\boldsymbol{\theta}}^{(i)})\}_{\mathbf{c}\in\{0,1\}^K}$. Together, the pair forms a soft-input/soft-output *non*coherent equalizer, which could be iterated with a soft-input/soft-output decoder for turbo reception.

When the channel is frequency-nonselective, the noise is white, and the BEM is trivial, $\ln f(\mathbf{y} \mid \mathbf{c}, \hat{\boldsymbol{\theta}}^{(i)}) = C - \frac{1}{\sigma_w^2} \sum_{k=0}^{K-1} \left| y[k] - c[k] \hat{\theta}[k,0] \right|^2$, so that the 2^K term summations in (6.97) decouple into K binary summations [CV01], greatly simplifying the evaluation of (6.97). In the general case, the 2^K-term summations do not decouple, but not all 2^K metrics $\{\zeta_{\text{coh}}(\mathbf{c}; \hat{\boldsymbol{\theta}}^{(i)})\}_{\mathbf{c}\in\{0,1\}^K}$ need to be calculated, because very few of them yield non-negligible $e^{\zeta_{\text{coh}}(\mathbf{c};\hat{\boldsymbol{\theta}}^{(i)})}$. The dominant posterior probabilities can be found without too much effort using, e.g., M-algorithm tree-search [HS09], as discussed for the coherent case in Section 6.3.2.

Other noncoherent equalization schemes

Other approaches to noncoherent equalization exist as well. For example, Anastasopoulos et al. [ACC+07, MAKA07] applied message passing algorithms to MLSD and MAP symbol detection over time-selective flat-fading channels. They have shown that, under

certain conditions, MLSD complexity scales as $\mathscr{O}(K^{2\text{rank}\{\mathbf{C}_{\boldsymbol{\theta}}\}})$. Due to space limitations, these techniques will not be discussed here. One can also imagine ad hoc combination of decoupled (coherent) equalization and channel estimation. Examples will be provided in the sequel.

6.4.4 Noncoherent equalization for single-carrier modulation/demodulation

When single-carrier modulation/demodulation is used, the effective channel coincides with the propagation channel (i.e., $q[l,d] = h[l,d]\ \forall l,d$ and $N_q = M$), whose trajectories $\{h[n,m]\}_{n=0}^{N-1}$ (for each $m \in \{0,\ldots,M-1\}$) are well described by a Gauss-Markov model of suitable order N_{AR}. Thus, when the trivial BEM is employed to write (6.2), so that $\theta[i,d] = q[i,d]\ \forall i,d$, the BEM trajectories themselves are well described by a Gauss-Markov model of order N_{AR}. In this case, and assuming the noise $\mathbf{z}$ is white, an $|\mathscr{A}|^{M+N_{\mathsf{AR}}}$-state trellis can be used to implement near-optimal noncoherent MAPSD/MLSD, as well as near-optimal MAP symbol and bit detection, as described in Section 6.4.3. In fact, these ideas dominated much of the early literature on noncoherent equalization of rapidly TV channels.

Near-optimal trellis-PSP equalization for single-carrier schemes

As mentioned earlier, Lodge and Moher [LM90] were one of the first papers to propose a near-optimal trellis-based implementation of noncoherent MLSD. In particular, they proposed to use the Viterbi algorithm with per-branch linear prediction for MLSD of CM signals with ARMA time-selective channels. Soon after, Iltis [Ilt92] proposed to use the Viterbi algorithm in conjunction with an extended Kalman filter for per-survivor joint estimation of symbol timing offset and an AR doubly selective channel. Dai and Shwedyk [DS94] proposed similar near-optimal trellis-based implementations of noncoherent MLSD for general signal alphabets and ARMA doubly selective channels, using per-branch Kalman filtering. Yu and Pasupathy [YP95] then extended [LM90] to general signal alphabets and ARMA doubly selective channels. The latter technique was extended further to carrier-frequency-offset Rician channels by Hart and Taylor in [HT98].

In related work, Gertsman and Lodge [GL97] showed that the forward-backward algorithm, with per-branch linear prediction, can be used for near-MAP symbol detection under AR time-selective channels and CM alphabets. Independently, these ideas were generalized this approach to doubly selective channels and general signal alphabets by Hart and Pasupathy [HP00] and Davis, Collings, and Hoeher [DCH01]. For fixed-lag MAP symbol estimation, Zhang, Fitz, and Gelfand [ZFG97] proposed to use per-survivor Kalman filtering, echoing earlier work by Iltis et al. [ISG94]. Anastasopoulos and Chugg [AC00] then presented two general families of trellis algorithms, one based on parameter-first combining and the other on sequence-first combining, that yield both forward-backward and fixed-lag algorithms.

Reduced-complexity trellis-PSP equalization for single-carrier schemes

Due to the complexity of near-optimal PSP methods, which are typically based on per-sequence Kalman filtering, simpler PSP techniques have also been proposed based on simpler forms of adaptive filtering, such as the RLS[25] and LMS algorithms. For example, Kubo, Murakami, and Fujino [KMF94] proposed to use the Viterbi algorithm in conjunction with the LMS algorithm [Hay01] for per-survivor channel estimation, whereas Raheli, Polydoros, and Tzou [RPT95] proposed to use the Viterbi algorithm in conjunction with the RLS algorithm [Hay01]. Other LMS and RLS approaches were discussed in [AC00].

While the previously described PSP algorithms assumed a trivial BEM, PSP for more general BEMs has also been considered. For example, trellis-based PSP algorithms for joint estimation of symbols and polynomial BEM [BH99a] coefficients were proposed, for time-selective channels, by Borah and Hart [BH99c] and Leon and Taylor [LT03]. DFE and trellis methods for both polynomial and Karhunen-Loève BEMs [BH99b] were studied by Borah and Hart in [BH99b] for doubly selective channels. Trellis-based PSP using a complex-exponential BEM [TG96] was discussed by El-Mahdy [EM04].

Near-optimal tree-PSP equalization for single-carrier schemes

Since the trellis-based approaches to noncoherent equalization typically use an $(M+N_{\mathsf{AR}})$-state trellis, with $\mathscr{O}(K|\mathscr{A}|^{M+N_{\mathsf{AR}}})$ complexity, they are impractical for all but very short delay spreads. Tree-search based PSP is one way to circumvent this complexity. In one of the earliest proposals, Dai and Shwedyk [DS94] suggested to use a Fano-like tree-search with per-survivor Kalman estimation to non-coherently equalize a doubly selective ARMA channel (assuming a trivial BEM). The method in Zhang, Fitz, and Gelfand [ZFG97] can be considered as using the T-algorithm to obtain symbol-MAP fixed-lag metrics for the same channel. For a doubly selective channel modeled by a generic BEM, Hwang and Schniter proposed PSP-based noncoherent M-algorithm tree-searches that accomplish approximate MLSD, in [HS07a], and approximate MAP bit detection, in [HS07b]. The latter, with complexity $\mathscr{O}(KM^2I^2)$, was combined with soft decoding in a turbo receiver.

Iterative noncoherent equalization for single-carrier schemes

For EM-based iterative noncoherent equalization of single-carrier systems, the channel EM(B) algorithm described in Section 6.4.3 is the most popular approach; the info EM(B) algorithm, proposed by Georghiades and Han [GH97] for AR time-selective fading and CM signaling and described in Appendix 6.D, was found, in the more recent studies [CPB03, YR03], to have convergence problems.

One of the first applications of the channel EMB algorithm to noncoherent equalization of frequency-selective channels was given by Kaleh and Vallet [KV94]. Anton-Haro, Fonollosa and Fonollosa [AHFF97] proposed a channel EM algorithm for doubly selective channels that used a polynomial BEM, assuming CM signaling. More recently, Yan

[25] Recall that the ML channel estimate (6.77) reduces to an LS estimate in the case of white noise, which can be computed recursively using RLS.

and Rao [YR03] proposed a channel EMB method for AR-1 time-selective channels and CM signaling that employs a Kalman filter. Nissilä and Pasupathy [NP03] generalized these ideas to AR doubly selective channels and arbitrary constellations through the use of Kalman smoothing. All of these approaches used the forward-backward algorithm to evaluate the posterior bit probabilities in (6.95). Because the number of trellis states in the forward-backward algorithm is $|\mathscr{A}|^{M+N_{\mathsf{AR}}}$, however, these approaches are practical for only very short delay spread M.

To circumvent the complexity of trellis processing, Hwang and Schniter [HS09] proposed to use suboptimal tree-search to compute the dominant posterior bit probabilities, via (6.96)-(6.97), leading to a complexity of only $\mathscr{O}(KM^2I^2)$. This approach can be considered an extension of the technique originally proposed by Chiavaccini and Vitetta [CV01] for a time-selective channel and trivial BEM, to doubly selective channels modeled by generic BEMs.

6.4.5 Noncoherent equalization for time-frequency concentrated modulation/demodulation

We saw, in the previous section, that the use of single-carrier modulation/demodulation facilitated Gauss-Markov modeling of the effective channel trajectory $\{q[l,d]\}_{l=0}^{K-1}$. Time-frequency concentrated modulation/demodulation schemes[26] generally do not facilitate the use of a Gauss-Markov model with order $N_{\mathsf{AR}} \ll K$, since $q[l,d]$ can change very quickly in l. For example, with multicarrier schemes, $\{q[l,0]\}_{l=0}^{K-1}$ represents the channel frequency response, which may exhibit deep and sudden nulls. Thus, the trellis-based approaches to noncoherent equalization (whether optimal, PSP approximate, or EM-based) do not apply here. For this reason, the literature on noncoherent equalization for time-frequency concentrated modulation/demodulation schemes is somewhat sparse.

For time-frequency concentrated modulation/demodulation, the dimensionality of the effective channel response $\{q[l,d]\}$ is more efficiently reduced by a BEM, e.g., the complex-exponential BEM (as proposed in the classical OFDM work [ESJJ+98]). Cui and Tellambura [CT07] applied the complex exponential (CE)-BEM, made several approximations to the noncoherent MLSD metric in (6.69) to reduce it to a simple quadratic form $\mathbf{a}^H\mathbf{R}\mathbf{a}$ (with $\mathbf{a}$-independent $\mathbf{R}$), and then used tree-search to find the optimal $\mathbf{a} \in \mathscr{A}^K$, all under the assumption that $\mathscr{A}$ was constant modulus. Hwang and Schniter took a more direct approach, leveraging the CE-BEM to design PSP-based and channel EMB-based noncoherent MAP bit equalization algorithms in [HS08] and [HS09], respectively, whose complexities scale as $\mathscr{O}\big(K(2D+1)^2I^2\big)$. The key to these low complexities is the use of a fast metric update. Here, the BEM dimension I refers to the number of *active* channel taps; it is typical that $I \ll M$ when the delay power profile is *sparse*. These latter algorithms achieve near-MAP performance with a complexity that is quite reasonable, even for large simultaneous channel delay and Doppler spreads.

[26] In this section, we will include OFDM in the "time-frequency concentrated" class under the assumption that the Doppler spread is mild enough to guarantee a short inter-carrier interference spread.

6.5 Conclusion

In this chapter, we have given a broad overview of coherent and noncoherent equalization for rapidly TV channels, focusing on the case of significant delay spread. To better understand the problem, we described the combined effect of modulation, channel propagation, and demodulation using an effective channel matrix **Q**, and then examined the key features of **Q**. We found that the support of the significant coefficients within **Q** can be described as widely quasi-banded when single-carrier modulation/demodulation is used, and narrowly quasi-banded when time-frequency concentrated modulation/demodulation is used. This structure of **Q** was later used to explain the design of low-complexity equalization algorithms.

We then discussed coherent equalization, where **Q** is assumed to be known. Various equalization criteria were described, including those based on ML, MAP, MMSE, and the computation of posterior LLRs. Equalization tools were described next, including trellis-based, linear, decision feedback, tree-search based, and iterative methods. We then described how these criteria and tools have been applied to the design of coherent equalizers for time-frequency concentrated modulation/demodulation over rapidly TV channels, highlighting fast serial and fast joint equalization schemes. For equalization of single-carrier modulation/demodulation over rapidly TV channels, we focused on frequency-domain equalization approaches that yield high performance with low complexity.

Finally, we discussed noncoherent equalization, where **Q** is assumed unknown (though sometimes its statistics are known). For this, the system model was reformulated to accommodate an efficient BEM-based parameterization of the effective channel **Q**. Various equalization criteria were described, including those based on ML, MAP, GLRT, MMSE, and posterior LLRs. Equalization tools were described next, including those based on trellis, tree search, per-survivor processing, and the EM algorithm. We then described how these criteria and tools have been applied to the design of noncoherent equalizers for single-carrier modulation/demodulation over rapidly TV channels. While the traditional approach was to leverage a Gauss-Markov fading model for the channel trajectory, general BEM approaches have been developed more recently. Finally, we described noncoherent equalization for time-frequency concentrated modulation/demodulation over rapidly TV channels, a problem which has received attention only recently.

Appendix

6.A Derivation of posterior LLR expression (6.36)

Applying Bayes rule to the numerator of the posterior LLR in (6.19), we find

$$\Pr\{c[j]=1 \mid \mathbf{y},\mathbf{Q}\} \;=\; \sum_{\mathbf{c}:c[j]=1} p(\mathbf{c} \mid \mathbf{y},\mathbf{Q}) \;=\; \sum_{\mathbf{c}:c[j]=1} \frac{f(\mathbf{y} \mid \mathbf{c},\mathbf{Q})p(\mathbf{c} \mid \mathbf{Q})}{f(\mathbf{y} \mid \mathbf{Q})}. \quad (6.98)$$

As in the text, all possibilities of coded bit vectors $\mathbf{c} \in \{0,1\}^{K\log_2|\mathscr{A}|}$ are considered in the summations, not only those in the codebook. Doing the same to the denominator of (6.19) and then taking the log of their ratio yields

$$L_{c|\mathbf{y},\mathbf{Q}}[j] \;=\; \ln \frac{\sum_{\mathbf{c}:c[j]=1} f(\mathbf{y} \mid \mathbf{c},\mathbf{Q})p(\mathbf{c} \mid \mathbf{Q})}{\sum_{\mathbf{c}:c[j]=0} f(\mathbf{y} \mid \mathbf{c},\mathbf{Q})p(\mathbf{c} \mid \mathbf{Q})}. \quad (6.99)$$

Assuming independent coded bits, so that $p(\mathbf{c} \mid \mathbf{Q}) = p(\mathbf{c}) = \prod_{j'=0}^{K\log_2|\mathscr{A}|-1} p(c[j'])$, and using the identity

$$p(c[j']) \;=\; \frac{\exp\big((c[j']-1)L_c[j']\big)}{1+\exp(-L_c[j'])} \quad \text{for} \quad c[j'] \in \{0,1\}, \quad (6.100)$$

we can rewrite (6.99) as

$$L_{c|\mathbf{y},\mathbf{Q}}[j] \;=\; \ln \frac{\sum_{\mathbf{c}:c[j]=1} f(\mathbf{y} \mid \mathbf{c},\mathbf{Q})\exp(\mathbf{l}_c^T\mathbf{c})}{\sum_{\mathbf{c}:c[j]=0} f(\mathbf{y} \mid \mathbf{c},\mathbf{Q})\exp(\mathbf{l}_c^T\mathbf{c})} \quad (6.101)$$

for $\mathbf{l}_c \triangleq \begin{pmatrix} L_c[0] & \cdots & L_c[K-1] \end{pmatrix}^T$. Finally, writing the LLR expression (6.101) in terms of the MAP metric (6.34) yields (6.36).

6.B Derivation of the noncoherent MLSD expression (6.73)

Using (6.68), we can write the first term in (6.69) as

$$\mathbf{y}^H\mathbf{C}_{\mathbf{y}|\mathbf{a}}^{-1}\mathbf{y} \;=\; \mathbf{y}^H\big(\mathbf{C}_\mathbf{z}+\mathbf{\Lambda}_\mathbf{a}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{\Lambda}_\mathbf{a}^H\big)^{-1}\mathbf{y}. \quad (6.102)$$

Applying the matrix inversion lemma, and introducing a pair of terms that sum to zero,

$$\begin{aligned}\mathbf{y}^H\mathbf{C}_{\mathbf{y}|\mathbf{a}}^{-1}\mathbf{y} \;=\;& \mathbf{y}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{y} - \mathbf{y}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{\Lambda}_\mathbf{a}\big(\mathbf{C}_{\boldsymbol{\theta}}^{-1}+\mathbf{\Lambda}_\mathbf{a}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{\Lambda}_\mathbf{a}\big)^{-1}\mathbf{\Lambda}_\mathbf{a}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{y} \\ & - \mathbf{y}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{\Lambda}_\mathbf{a}\big(\mathbf{C}_{\boldsymbol{\theta}}^{-1}+\mathbf{\Lambda}_\mathbf{a}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{\Lambda}_\mathbf{a}\big)^{-1}\mathbf{\Lambda}_\mathbf{a}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{y} \\ & + \mathbf{y}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{\Lambda}_\mathbf{a}\big(\mathbf{C}_{\boldsymbol{\theta}}^{-1}+\mathbf{\Lambda}_\mathbf{a}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{\Lambda}_\mathbf{a}\big)^{-1}\mathbf{\Lambda}_\mathbf{a}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{y}. \end{aligned} \quad (6.103)$$

Plugging in the expression for $\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}}$ in (6.72),

$$\begin{aligned}
\mathbf{y}^H\mathbf{C}_{\mathbf{y}|\mathbf{a}}^{-1}\mathbf{y} &= \mathbf{y}^H\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{y} - \hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}}^H\boldsymbol{\Lambda}_{\mathbf{a}}^H\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{y} - \mathbf{y}^H\mathbf{C}_{\mathbf{z}}^{-1}\boldsymbol{\Lambda}_{\mathbf{a}}\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}} \\
&\quad + \hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}}^H\big(\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{\Lambda}_{\mathbf{a}}^H\mathbf{C}_{\mathbf{z}}^{-1}\boldsymbol{\Lambda}_{\mathbf{a}}\big)\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}} \qquad (6.104)\\
&= \big\|\mathbf{y} - \boldsymbol{\Lambda}_{\mathbf{a}}\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}}\big\|_{\mathbf{C}_{\mathbf{z}}^{-1}}^2 + \|\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}|\mathbf{a}}\|_{\mathbf{C}_{\boldsymbol{\theta}}^{-1}}^2. \qquad (6.105)
\end{aligned}$$

6.C Explanation of EM recursion (6.91)

Given an estimate $\hat{\mathbf{x}}^{(i)}$ at the i^{th} iteration, the EM algorithm [DLR77] attempts to find $\mathbf{x}$ which maximizes the *increase* in log-likelihood, i.e.,

$$\begin{aligned}
&\ln f(\mathbf{y} \mid \mathbf{x}) - \ln f(\mathbf{y} \mid \hat{\mathbf{x}}^{(i)}) \\
&= \ln \int f(\mathbf{y},\mathbf{u} \mid \mathbf{x}) d\mathbf{u} - \ln f(\mathbf{y} \mid \hat{\mathbf{x}}^{(i)}) \qquad (6.106)\\
&= \ln \int f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)}) \frac{f(\mathbf{y},\mathbf{u} \mid \mathbf{x})}{f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)})} d\mathbf{u} - \ln f(\mathbf{y} \mid \hat{\mathbf{x}}^{(i)}) \qquad (6.107)\\
&\geq \int f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)}) \ln \frac{f(\mathbf{y},\mathbf{u} \mid \mathbf{x})}{f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)})} d\mathbf{u} - \ln f(\mathbf{y} \mid \hat{\mathbf{x}}^{(i)}) \;\triangleq\; \Delta(\mathbf{x} \mid \hat{\mathbf{x}}^{(i)}), \qquad (6.108)
\end{aligned}$$

where Jensen's inequality was used in (6.108). From (6.108), it can be seen that $\hat{\mathbf{x}}^{(i+1)} \triangleq \arg\max_{\mathbf{x}} \Delta(\mathbf{x} \mid \hat{\mathbf{x}}^{(i)})$ can be written as (6.91) after dropping non-essential terms. Since

$$\begin{aligned}
\Delta(\hat{\mathbf{x}}^{(i)} \mid \hat{\mathbf{x}}^{(i)}) &= \int f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)}) \ln \frac{f(\mathbf{y},\mathbf{u} \mid \hat{\mathbf{x}}^{(i)})}{f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)})} d\mathbf{u} - \ln f(\mathbf{y} \mid \hat{\mathbf{x}}^{(i)}) \qquad (6.109)\\
&= \int f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)}) \ln \frac{f(\mathbf{y},\mathbf{u} \mid \hat{\mathbf{x}}^{(i)})}{f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)})} d\mathbf{u} \\
&\quad + \ln \frac{1}{f(\mathbf{y} \mid \hat{\mathbf{x}}^{(i)})} \int f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)}) d\mathbf{u} \qquad (6.110)\\
&= \int f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)}) \ln \underbrace{\frac{f(\mathbf{y},\mathbf{u} \mid \hat{\mathbf{x}}^{(i)})}{f(\mathbf{u} \mid \mathbf{y},\hat{\mathbf{x}}^{(i)}) f(\mathbf{y} \mid \hat{\mathbf{x}}^{(i)})}}_{=1} d\mathbf{u} \qquad (6.111)\\
&= 0, \qquad (6.112)
\end{aligned}$$

if follows that the increase in log-likelihood associated with the EM estimate $\hat{\mathbf{x}}^{(i+1)}$ equals $\Delta(\hat{\mathbf{x}}^{(i+1)} \mid \hat{\mathbf{x}}^{(i)}) = \max_{\mathbf{x}} \Delta(\mathbf{x} \mid \hat{\mathbf{x}}^{(i)}) \geq \Delta(\hat{\mathbf{x}}^{(i)} \mid \hat{\mathbf{x}}^{(i)}) = 0$, and hence the EM recursion never decreases the log likelihood. Thus, when the likelihood $f(\mathbf{y} \mid \mathbf{x})$ is unimodal in $\mathbf{x}$, the EM recursions will converge to $\hat{\mathbf{x}}_{\mathsf{ML}}$.

6.D Info EM(B) algorithms for noncoherent equalization

In info EMB, the bits $\mathbf{c}$ are estimated while treating the channel parameters $\boldsymbol{\theta}$ as missing data, so that (6.92) becomes (with $\mathbf{x}=\mathbf{c}$ and $\mathbf{u}=\boldsymbol{\theta}$)

$$\hat{\mathbf{c}}^{(i+1)} = \arg\max_{\mathbf{c}\in\{0,1\}^K}\left\{\int f(\boldsymbol{\theta}\mid\mathbf{y},\hat{\mathbf{c}}^{(i)})\ln f(\mathbf{y},\boldsymbol{\theta}\mid\mathbf{c})d\boldsymbol{\theta}+\ln p(\mathbf{c})\right\}. \quad (6.113)$$

The identity $\ln f(\mathbf{y},\boldsymbol{\theta}\mid\mathbf{c})=\ln f(\mathbf{y}\mid\boldsymbol{\theta},\mathbf{c})+\ln f(\boldsymbol{\theta})$, in conjunction with the Gaussian noise assumption, yields

$$\hat{\mathbf{c}}^{(i+1)} = \arg\min_{\mathbf{c}\in\{0,1\}^K}\left\{\int f(\boldsymbol{\theta}\mid\mathbf{y},\hat{\mathbf{c}}^{(i)})\,\|\mathbf{y}-\boldsymbol{\Lambda}_{\mathbf{a}'}\boldsymbol{\theta}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}d\boldsymbol{\theta}-\ln p(\mathbf{c})\right\}, \quad (6.114)$$

where $\mathbf{a}$ is a one-to-one function of $\mathbf{c}$. Then using Bayes rule for

$$f(\boldsymbol{\theta}\mid\mathbf{y},\hat{\mathbf{c}}^{(i)})=\frac{f(\mathbf{y}\mid\boldsymbol{\theta},\hat{\mathbf{c}}^{(i)})f(\boldsymbol{\theta})}{\int f(\mathbf{y}\mid\boldsymbol{\theta}',\hat{\mathbf{c}}^{(i)})f(\boldsymbol{\theta}')d\boldsymbol{\theta}'},$$

in conjunction with the Rayleigh fading assumption, (6.113) reduces to

$$\hat{\mathbf{c}}^{(i+1)} = \arg\min_{\mathbf{c}\in\{0,1\}^K}\left\{\frac{\int\|\mathbf{y}-\boldsymbol{\Lambda}_\mathbf{a}\boldsymbol{\theta}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}f(\mathbf{y}\mid\boldsymbol{\theta},\hat{\mathbf{c}}^{(i)})f(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int f(\mathbf{y}\mid\boldsymbol{\theta}',\hat{\mathbf{c}}^{(i)})f(\boldsymbol{\theta}')d\boldsymbol{\theta}'}-\ln p(\mathbf{c})\right\} \quad (6.115)$$

$$= \arg\min_{\mathbf{c}\in\{0,1\}^K}\left\{\frac{\int\exp\big(-\|\mathbf{y}-\boldsymbol{\Lambda}_{\hat{\mathbf{a}}^{(i)}}\boldsymbol{\theta}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}-\|\boldsymbol{\theta}\|^2_{\mathbf{C}_{\boldsymbol{\theta}}^{-1}}\big)\|\mathbf{y}-\boldsymbol{\Lambda}_\mathbf{a}\boldsymbol{\theta}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}d\boldsymbol{\theta}}{\int\exp\big(-\|\mathbf{y}-\boldsymbol{\Lambda}_{\hat{\mathbf{a}}^{(i)}}\boldsymbol{\theta}'\|^2_{\mathbf{C}_\mathbf{z}^{-1}}-\|\boldsymbol{\theta}'\|^2_{\mathbf{C}_{\boldsymbol{\theta}}^{-1}}\big)d\boldsymbol{\theta}'}\right.$$

$$\left.-\ln p(\mathbf{c})\right\}. \quad (6.116)$$

The optimization problem (6.116) is, in general, difficult to solve. In the simplified case of frequency-nonselective fading, white noise, CM alphabet, and the trivial BEM, though, $\|\mathbf{y}-\boldsymbol{\Lambda}_\mathbf{a}\boldsymbol{\theta}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}=C_1+C_2\prod_{k=0}^{K-1}\mathrm{Re}\{a[k]\hat{\theta}[k,0]\}$ for C_1 and C_2 that do not depend on $\mathbf{a}$, making the optimization problem (6.116) tractable [GH97]. Even then, the hard-decision nature of info EMB makes it subject to error propagation. Thus, it is not surprising that channel EMB has been shown to outperform info EMB for this simplified setup [YR03].

For info EM, or without an informative prior distribution for $\mathbf{c}$, the $\ln p(\mathbf{c})$ term can be neglected, so that

$$\hat{\mathbf{c}}^{(i+1)} = \arg\min_{\mathbf{c}\in\{0,1\}^K}\int\exp\big(-\|\mathbf{y}-\boldsymbol{\Lambda}_{\hat{\mathbf{a}}^{(i)}}\boldsymbol{\theta}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}-\|\boldsymbol{\theta}\|^2_{\mathbf{C}_{\boldsymbol{\theta}}^{-1}}\big)\|\mathbf{y}-\boldsymbol{\Lambda}_\mathbf{a}\boldsymbol{\theta}\|^2_{\mathbf{C}_\mathbf{z}^{-1}}d\boldsymbol{\theta}. \quad (6.117)$$

Note, however, that the minimization is not significantly simplified relative to (6.116).

Table 6.1 *Variables introduced for the equalization chapter*

no.	*symbol*	*description*	*avoid*
1.1	$\mathbf{H}$	channel convolution matrix	
1.2	$\mathbf{G}$	modulation matrix	
1.3	$\mathbf{\Gamma}$	demodulation matrix	
1.4	$\mathbf{Q} = \mathbf{\Gamma H G}$	effective channel matrix	
1.5	$\mathbf{U}$	upper triangular matrix	
1.6	$\mathscr{A} \subset \mathscr{A}^K$	codebook	
1.7	$L_c[i] = \frac{\Pr\{c[i]=1\}}{\Pr\{c[i]=0\}}$	a-priori log likelihood ratio (LLR),	
1.8	$L_{c\|\mathbf{y},\mathbf{Q}}[i] = \frac{\Pr\{c[i]=1 \mid \mathbf{y},\mathbf{Q}\}}{\Pr\{c[i]=0 \mid \mathbf{y},\mathbf{Q}\}}$	coherent posterior log likelihood ratio (LLR),	
1.9	$L_{c\|\mathbf{y}}[i] = \frac{\Pr\{c[i]=1 \mid \mathbf{y}\}}{\Pr\{c[i]=0 \mid \mathbf{y}\}}$	noncoherent posterior log likelihood ratio (LLR),	
1.10	$\|\mathbf{x}\|_{\mathbf{C}} = \sqrt{\mathbf{x}^H \mathbf{C} \mathbf{x}}$	weighted norm of $\mathbf{x}$ w.r.t Hermitian PSD matrix $\mathbf{C}$	
1.11	$\mathscr{D}_{\mathscr{A}}(\cdot)$	quantization w.r.t. alphabet $\mathscr{A}$	
1.12	N_g	OFDM guard length	
1.13	N_{AR}	Gauss-Markov model order	
1.14	$\mathrm{diag}\{.\}$	create diagonal matrix from vector	

References

[AC00] A. Anastasopoulos and K. M. Chugg. Iterative detection for channels with memory. *IEEE Trans. Commun.*, 48(10):1638–1649, Oct. 2000.

[ACC+07] A. Anastasopoulos, K. M. Chugg, G. Colavolpe, G. Ferrari, and R. Raheli. Iterative detection for channels with memory. *Proc. IEEE*, 95(6):1272–1294, June 2007.

[ADC95] N. Al-Dhahir and J. M. Cioffi. MMSE decision feedback equalizers: Finite-length results. *IEEE Trans. Inform. Theory*, 41(4):961–976, July 1995.

[ADS00] N. Al-Dhahir and A. H. Sayed. The finite-length multi-input multi-output MMSE-DFE. *IEEE Trans. Signal Process.*, 48(10):2921–2936, Oct. 2000.

[AEVZ02] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger. Closest point search in lattices. *IEEE Trans. Inform. Theory*, 48(8):2201–2214, Aug. 2002.

[AHFF97] C. Antón-Haro, J. A. R. Fonollosa, and J. R. Fonollosa. Blind channel estimation and data detection using hidden Markov models. *IEEE Trans. Signal Process.*, 45(1):241–247, Jan. 1997.

[AM84] J. B. Anderson and S. Mohan. Sequential decoding algorithms: A survey and cost analysis. *IEEE Trans. Commun.*, 32:169–172, 1984.

[ASLC06] S. Ahmed, M. Sellathurai, S. Lambotharan, and J. A. Chambers. Low-complexity iterative method of equalization for single carrier with cyclic prefix in doubly selective channels. *IEEE Signal Process. Lett.*, 13(1):5–8, Jan. 2006.

[BC98] E. Baccarelli and R. Cusani. Combined channel estimation and data detection using soft statistics for frequency-selective fast-fading digital links. *IEEE Trans. Commun.*, 46(4):424–427, Apr. 1998.

[BCJR74] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. Inform. Theory*, 20:284–287, Mar. 1974.

[BH99a] D. K. Borah and B. D. Hart. Frequency-selective fading channel estimation with a polynomial time-varying channel model. *IEEE Trans. Commun.*, 47(6):862–873, June 1999.

[BH99b] D. K. Borah and B. D. Hart. Receiver structures for time-varying frequency-selective fading channels. *IEEE J. Sel. Areas Commun.*, 17(11):1863–1875, Nov. 1999.

[BH99c] D. K. Borah and B. D. Hart. A robust receiver structure for time-varying, frequency-flat, Rayleigh fading channels. *IEEE Trans. Commun.*, 47(3):360–364, Mar. 1999.

[BLM04] I. Barhumi, G. Leus, and M. Moonen. Time-domain and frequency-domain per-tone equalization for OFDM in doubly selective channels. *Signal Process.*, 84:2055–2066, Nov. 2004.

[BLM05] I. Barhumi, G. Leus, and M. Moonen. Time-varying FIR equalization for doubly selective channels. *IEEE Trans. Wireless Commun.*, 4(1):202–214, Jan. 2005.

[BLM06] I. Barhumi, G. Leus, and M. Moonen. Equalization for OFDM over doubly selective channels. *IEEE Trans. Signal Process.*, 54(4):1445–1458, Nov. 2006.

[Böl02] Helmut Bölcskei. Orthogonal frequency division multiplexing based on offset QAM. In H. G. Feichtinger and T. Strohmer, editors, *Advances in Gabor Analysis*, pages 321–352. Birkhäuser, Boston, 2002.

[BT01] S. Barbarossa and R. Torti. Chirped-OFDM for transmissions over time-varying channels with linear delay/Doppler spreading. In *Proc. IEEE Int. Conf. Acoust. Speech & Signal Process.*, volume 4, pages 2377–2380, Salt Lake City, UT, May 2001.

[CF97] J. M. Cioffi and G. D. Forney. Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise. In A. Paulraj, V. Roychowdhury, and C. Schaper, editors, *Communication, Computation, Control and Signal Processing*, chapter 4, pages 79–127. Kluwer, Boston, MA, 1997.

[CG03] X. Cai and G. B. Giannakis. Bounding performance and suppressing intercarrier interference in wireless mobile OFDM. *IEEE Trans. Commun.*, 51(12):2047–2056, Dec. 2003.

[Chu98] K. Chugg. The condition for the applicability of the Viterbi algorithm with implications for fading channel MLSD. *IEEE Trans. Commun.*, 46(9):1112–1116, Sep. 1998.

[Cim85] L. J. Cimini, Jr. Analysis and simulation of a digital mobile radio channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.*, 33:665–765, July 1985.

[CPB03] H. Chen, R. Perry, and K. Buckley. On MLSE algorithms for unknown fast time-varying channels. *IEEE Trans. Commun.*, 51(5):730–734, May 2003.

[CT01] H. A. Cirpan and M. K. Tsatsanis. Maximum-likelihood estimation of fir channels excited by convolutionally encoded inputs. *IEEE Trans. Commun.*, 49:1125–1128, July 2001.

[CT07] T. Cui and C. Tellambura. Blind receiver design for OFDM systems over doubly selective channels. *IEEE Trans. Commun.*, 55(5):906–917, May 2007.

[CV01] E. Chiavaccini and G. M. Vitetta. MAP symbol estimation on frequency-flat Rayleigh fading channels via a Bayesian EM algorithm. *IEEE Trans. Commun.*, 49(11):1869–1872, Nov. 2001.

[CVC01] Y.-S. Choi, P. J. Voltz, and F. A. Cassara. On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels. *IEEE Trans. Commun.*, 49(8):1375–1387, Aug. 2001.

[Dau92] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.

[DCH01] L. M. Davis, I. B. Collings, and P. Hoeher. Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels. *IEEE Trans. Commun.*, 49(12):2106–2114, Dec. 2001.

[DEC03] M. O. Damen, H. El Gamal, and G. Caire. On maximum-likelihood detection and the search for the closest lattice point. *IEEE Trans. Inform. Theory*, 49(10):2389–2402, Oct. 2003.

[DHH89] A. Duel-Hallen and C. Heegard. Delayed decision feedback sequence equalization. *IEEE Trans. Commun.*, pages 428–436, May 1989.

[DJB+95] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux. Iterative correction of intersymbol interference: Turbo equalization. *European Trans. Telecommun.*, 6:507–511, Sept.-Oct. 1995.

[DLR77] A. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc.*, 39:1–17, 1977.

[DS94] Q. Dai and E. Shwedyk. Detection of bandlimited signals over frequency selective Rayleigh fading channels. *IEEE Trans. Commun.*, 42(2/3/4):941–950, Feb./Mar./Apr. 1994.

[DS07] S. Das and P. Schniter. Max-SINR ISI/ICI-shaping multi-carrier communication over the doubly dispersive channel. *IEEE Trans. Signal Process.*, 55(12):5782–5795, Dec. 2007.

[dW05] Y. L. C. de Jong and T. J. Willink. Iterative tree search detection for MIMO wireless systems. *IEEE Trans. Commun.*, 53:930–935, June 2005.

[EM04] A. E.-S. El-Mahdy. Adaptive channel estimation and equalization for rapidly mobile communication channels. *IEEE Trans. Commun.*, 52(7):1126–1135, July 2004.

[EQ88] M. V. Eyuboglu and S. U. Qureshi. Reduced-state sequence estimation with set partitioning and decision feedback. *IEEE Trans. Commun.*, 36(1):12–20, Jan. 1988.

[ESJJ+98] O. Edfors, M. Sandell, J.-J.van de Beek, S. K. Wilson, and P. O. Börjesson. OFDM channel estimation by singular value decomposition. *IEEE Trans. Commun.*, 46:931–939, July 1998.

[FABSE02] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson. Frequency domain equalization for single-carrier broadband wireless systems. *IEEE Commun. Mag.*, 40(4):58–66, Apr. 2002.

[Fan63] R. M. Fano. A heuristic discussion of probabilistic decoding. *IEEE Trans. Inform. Theory*, 9(2):64–74, Apr. 1963.

[FH94] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Trans. Signal Process.*, 42:2664–2677, Oct. 1994.

[For72] G. Forney. Maximum-likelihood sequence estimation of digital sequences in the presence of inter symbol interference. *IEEE Trans. Inform. Theory*, 18(3):363–378, May 1972.

[For73] G. D. Forney, Jr. The Viterbi algorithm. *IEEE Trans. Inform. Theory*, 61:262–278, Mar. 1973.

[FP85] U. Fincke and M. Pohst. Improved methods for calculating vectors of short length in a lattice. *Math. Comput.*, 44:463–471, Apr. 1985.

[FRL08] K. Fang, L. Rugini, and G. Leus. Iterative channel estimation and turbo equalization for time-varying OFDM systems. In *Proc. IEEE Int. Conf. Acoust. Speech & Signal Process.*, 2008.

[GH97] C. N. Georghiades and J. C. Han. Sequence estimation in the presence of random parameters via the EM algorithm. *IEEE Trans. Commun.*, 45:300–308, Mar. 1997.

[GH03] R. Gowaikar and B. Hassibi. Efficient statistical pruning algorithms for maximum likelihood decoding. In *Proc. IEEE Int. Conf. Acoust. Speech & Signal Process.*, volume 5, pages 49–52, Hong Kong, Apr. 2003.

[GL97] M. J. Gertsman and J. H. Lodge. Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels. *IEEE Trans. Commun.*, 45(7):788–799, Jul. 1997.

[GL04] A. Gorokhov and J.-P. Linnartz. Robust OFDM receivers for dispersive time-varying channels: Equalization and channel acquisition. *IEEE Trans. Commun.*, 52(4):572–583, Apr. 2004.

[GS03] M. Guillaud and D. T. M. Slock. Channel modeling and associated inter-carrier interference equalization for OFDM systems with high Doppler spread. In *Proc. IEEE Int. Conf. Acoust. Speech & Signal Process.*, volume 4, pages 237–240, Hong Kong, Apr. 2003.

[Har00] B. D. Hart. Maximum likelihood sequence detection using a pilot tone. *IEEE Trans. Veh. Tech.*, 49(2):550–560, Mar. 2000.

[Hay01] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Englewood Cliffs, NJ, 4th edition, 2001.

[HB97] R. Haas and J.-C. Belfiore. A time-frequency well-localized pulse for multiple carrier transmission. *Wireless Pers. Commun.*, 5:1–18, July 1997.

[HC05] W.-S. Hou and B.-S. Chen. ICI cancellation for OFDM communication systems in time-varying multipath fading channels. *IEEE Trans. Wireless Commun.*, 4(5):2100–2110, Sep. 2005.

[HD03] T. Hunziker and D. Dahlhaus. Iterative detection for multicarrier transmission employing time-frequency concentrated pulses. *IEEE Trans. Commun.*, 51(4):641–651, Apr. 2003.

[HH89] J. Hagenauer and P. Hoeher. A Viterbi algorithm with soft-decision outputs and its applications. In *Proc. IEEE Global Telecommun. Conf.*, pages 1680–1686, Dallas, TX, Nov. 1989.

[HM89] R. Haeb and H. Meyr. A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels. *IEEE Trans. Commun.*, 37(7):748–754, July 1989.

[HP00] B. D. Hart and S. Pasupathy. Innovations-based MAP detection for time-varying frequency-selective channels. *IEEE Trans. Veh. Tech.*, 48(9):1507–1519, Sept. 2000.

[HS05] S.-J. Hwang and P. Schniter. On the optimality of MMSE-GDFE preprocessed sphere decoding. In *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Oct. 2005.

[HS06] S.-J. Hwang and P. Schniter. Efficient sequence detection of multi-carrier transmissions over doubly dispersive channels. *EURASIP J. Appl. Signal Process.*, Article ID 93638, 17 pages, 2006.

[HS07a] S.-J. Hwang and P. Schniter. Efficient communication over highly spread underwater acoustic channels. In *Proc. ACM Int. Workshop Underwater Netw. (WUWNet)*, Montreal, Quebec, Sept. 2007.

[HS07b] S.-J. Hwang and P. Schniter. Fast noncoherent decoding of block transmissions over doubly dispersive channels. In *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 2007.

[HS07c] S.-J. Hwang and P. Schniter. Maximum-diversity affine precoding for the noncoherent doubly dispersive channel. In *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, pages 1–5, Helsinki, Finland, June 2007.

[HS08] S.-J. Hwang and P. Schniter. Efficient multicarrier communication for highly spread underwater acoustic channels. *IEEE J. Sel. Areas Commun.*, 26(9):1674–1683, Dec. 2008.

[HS09] S.-J. Hwang and P. Schniter. EM-based soft noncoherent equalization of doubly selective channels using tree search and basis expansion. In *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Perugia, Italy, June 2009.

[HT98] B. D. Hart and D. P. Taylor. Maximum-likelihood synchronization, equalization, and sequence estimation for unknown time-varying frequency-selective Rician channels. *IEEE Trans. Commun.*, 46(2):211–221, Feb. 1998.

[Ht03] B. M. Hochwald and S. ten Brink. Achieving near-capacity on a multiple-antenna channel. *IEEE Trans. Commun.*, 51:389–399, Mar. 2003.

[HV05] B. Hassibi and H. Vikalo. On the sphere-decoding algorithm I. Expected complexity. *IEEE Trans. Signal Process.*, 53(8):2806–2818, Aug. 2005.

[Ilt92] R. A. Iltis. A Bayesian maximum-likelihood sequence estimation algorithm for a priori unknown channels and symbol timing. *IEEE J. Sel. Areas Commun.*, 10(3):579–588, Apr. 1992.

[ISG94] R. A. Iltis, J. J. Shynk, and K. Giridhar. Bayesian algorithms for blind equalization using parallel adaptive filtering. *IEEE Trans. Commun.*, 42(2/3/4):1017–1032, Feb./Mar./Apr. 1994.

[JCC99] W. G. Jeon, K. H. Chang, and Y. S. Cho. An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels. *IEEE Trans. Commun.*, 47(1):27–32, Jan. 1999.

[JO05] J. Jalden and B. Ottersten. On the complexity of sphere decoding in digital communications. *IEEE Trans. Signal Process.*, 53(4):1474–1484, 2005.

[Kai60] T. Kailath. Correlation detection of signals perturbed by a random channel. *IRE Trans. Inform. Theory*, 6:361–366, June 1960.

[Kai69] T. Kailath. A general likelihood formula for random signals in Gaussian noise. *IEEE Trans. Inform. Theory*, 15(3):350–361, May 1969.

[KM98] W. Kozek and A. F. Molisch. Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels. *IEEE J. Sel. Areas Commun.*, 16(8):1579–1589, Oct. 1998.

[KMF94] H. Kubo, K. Murakami, and T. Fujino. An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels. *IEEE Trans. Commun.*, 42(2/3/4):1872–1880, Feb./Mar./Apr. 1994.

[KS98] D. Kim and G. Stüber. Residual ISI cancellation for OFDM with application to HDTV broadcasting. *IEEE J. Sel. Areas Commun.*, 16(8):1590–1599, 1998.

[KS08] A. P. Kannu and P. Schniter. Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels. *IEEE Trans. Signal Process.*, 56(3):1148–1160, Mar. 2008.

[KST04] R. Koetter, A. C. Singer, and M. Tüchler. Turbo equalization. *IEEE Signal Process. Mag.*, 21(1):67–80, Jan. 2004.

[KV94] G. K. Kaleh and R. Vallet. Joint parameter estimation and symbol detection for linear or nonlinear unknown channels. *IEEE Trans. Commun.*, 42:2406–2413, July 1994.

[LAB95] B. Le Floch, M. Alard, and C. Berrou. Coded orthogonal frequency division multiplex. *Proc. IEEE*, 83(6):982–996, June 1995.

[LF07] D. Liu and M. P. Fitz. Reduced state iterative MAP equalization and decoding in wireless mobile coded OFDM. In *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Sep. 2007.

[LF08] D. Liu and M. P. Fitz. Joint turbo channel estimation and data recovery in fast fading mobile coded OFDM. In *Proc. IEEE Int. Symposium Pers. Indoor Mobile Radio Commun.*, Sep. 2008.

[LLL82] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, Dec. 1982.

[LM90] J. H. Lodge and M. L. Moher. Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels. *IEEE Trans. Commun.*, 38(6):787–794, June 1990.

[LT03] W. S. Leon and D. P. Taylor. Generalized polynomial-based receiver for the flat fading channel. *IEEE Trans. Commun.*, 51(6):896–899, June 2003.

[LVS95] Y. Li, B. Vucetic, and Y. Sato. Optimum soft-output detection for channels with intersymbol interference. *IEEE Trans. Inform. Theory*, 41:704–713, May 1995.

[MAKA07] I. Motedayen-Aval, A. Krishnamoorthy, and A. Anastasopoulos. Optimal joint detection/estimation in fading channels with polynomial complexity. *IEEE Trans. Inform. Theory*, 53(1):209–223, Jan. 2007.

[Mar00] M. Martone. Wavelet-based separating kernels for array processing of cellular DS/CDMA signals in fast fading. *IEEE Trans. Commun.*, 48(6):979–995, June 2000.

[Mar01] M. Martone. A multicarrier system based on the fractional Fourier transform for time-frequency selective channels. *IEEE Trans. Commun.*, 49(6):1011–1020, June 2001.

[MEDC06] A. Murugan, H. El Gamal, M. O. Damen, and G. Caire. A unified framework for tree search decoding: Rediscovering the sequential decoder. *IEEE Trans. Inform. Theory*, 52(3):933–53, Mar. 2006.

[MG03] X. Ma and G. B. Giannakis. Maximum-diversity transmissions over doubly selective wireless channels. *IEEE Trans. Inform. Theory*, 49(7):1832–1840, July 2003.

[MK97] K. Matheus and K.-D. Kammeyer. Optimal design of a multicarrier system with soft impulse shaping including equalization in time or frequency direction. In *Proc. IEEE Global Telecommun. Conf.*, volume 1, pages 310–314, Phoenix, AZ, Nov. 1997.

[Mos96] S. Moshavi. Multi-user detection for DS-CDMA systems. *IEEE Commun. Mag.*, 34(10):124–136, Oct. 1996.

[Mow94] W. H. Mow. Maximum likelihood sequence estimation from the lattice viewpoint. *IEEE Trans. Inform. Theory*, 40(5):1591–1600, 1994.

[MS79] R. E. Morley, Jr. and D. L . Snyder. Maximum likelihood sequence estimation for randomly dispersive channels. *IEEE Trans. Commun.*, 27(6):833–839, June 1979.

[MSG+07] G. Matz, D. Schafhuber, K. Grochenig, M. Hartmann, and F. Hlawatsch. Analysis, optimization, and implementation of low-interference wireless multicarrier systems. *IEEE Trans. Wireless Commun.*, 6(5):1921–1931, May 2007.

[NF04] B. K. Ng and D. Falconer. A novel frequency domain equalization method for single-carrier wireless transmissions over doubly-selective fading channels. In *Proc. IEEE Global Telecommun. Conf.*, volume 1, pages 237–241, Dallas, TX, Dec. 2004.

[NL05] H. Nguyen and B. C. Levy. Blind equalization of dispersive fast fading Ricean channels via the EMV algorithm. *IEEE Trans. Veh. Tech.*, 54(5):1793–1801, Sep. 2005.

[NP03] M. Nissilä and S. Pasupathy. Adaptive Bayesian and EM-based detectors for frequency-selective fading channels. *IEEE Trans. Commun.*, 51(8):1325–1336, Aug. 2003.

[Od85] J. J. O'Reilly and A. M. de Oliviera Duarte. Error propagation in decision feedback receivers. *IEE Proc. F: Image Radar & Signal Process.*, 132(7):561–566, Dec. 1985.

[OT04] R. Otnes and M. Tüchler. Iterative channel estimation for turbo equalization of time-varying frequency-selective channels. *IEEE Trans. Wireless Commun.*, 3(6):1918–1923, Nov. 2004.

[Poo94] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer, New York, 2nd edition, 1994.

[PR06] F. Peng and W. E. Ryan. A low-complexity soft demapper for OFDM fading channels with ICI. In *Proc. IEEE Wireless Commun. & Netw. Conf.*, pages 1549–1554, Las Vegas, NV, Apr. 2006.

[Pro01] J. G. Proakis. *Digital Communications*. McGraw-Hill, New York, 4th edition, 2001.

[RBL05] L. Rugini, P. Banelli, and G. Leus. Simple equalization of time-varying channels for OFDM. *IEEE Commun. Lett.*, 9(7):619–621, July 2005.

[RBL06] L. Rugini, P. Banelli, and G. Leus. Low-complexity banded equalizers for OFDM systems in Doppler spread channels. *EURASIP J. Appl. Signal Process.*, Article ID 67404, 13 pages, 2006.

[RC96] D. J. Reader and W. G. Cowley. Blind maximum likelihood sequence detection over fast fading channels. In *Proc. European Signal Process. Conf.*, Trieste, Italy, Sep. 1996.

[RPT95] R. Raheli, A. Polydoros, and C. K. Tzou. Per-survivor processing: A general approach to MLSE in uncertain environments. *IEEE Trans. Commun.*, 43(2/3/4):354–364, Feb./Mar./Apr. 1995.

[SAMT05] K. Sigloch, M. R. Andrews, P. P. Mitra, and D. J. Thomson. Communicating over nonstationary nonflat wireless channels. *IEEE Trans. Signal Process.*, 53(6):2216–2227, June 2005.

[SB03] T. Strohmer and S. Beaver. Optimal OFDM design for time-frequency dispersive channels. *IEEE Trans. Commun.*, 51(7):1111–1122, July 2003.

[Sch04] P. Schniter. Low-complexity equalization of OFDM in doubly selective channels. *IEEE Trans. Signal Process.*, 52(4):1002–1011, Apr. 2004.

[SDAD02] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir. Intercarrier interference in MIMO OFDM. *IEEE Trans. Signal Process.*, 50(10):2451–2464, Oct. 2002.

[Sim90] S. J. Simmons. Breadth-first trellis decoding with adaptive effort. *IEEE Trans. Commun.*, 38(1):3–12, Jan. 1990.

[SL03] P. Schniter and H. Liu. Iterative equalization for single carrier cyclic-prefix in doubly-dispersive channels. In *Proc. Asilomar Conf. Signals Syst. Comput.*, volume 1, pages 502–506, Pacific Grove, CA, Oct. 2003.

[SL04] P. Schniter and H. Liu. Iterative frequency-domain equalization for single-carrier systems in doubly-dispersive channels. In *Proc. Asilomar Conf. Signals Syst. Comput.*, volume 1, pages 667–671, Pacific Grove, CA, Nov. 2004.

[SSS04] S. Song, A. C. Singer, and K.-M. Sung. Soft input channel estimation for turbo equalization. *IEEE Trans. Signal Process.*, 52(10):2885–2894, Oct. 2004.

[Stü01] G. L. Stüber. *Principles of Mobile Communication*. Springer, New York, 2nd edition, 2001.

[TG96] M. K. Tsatsanis and G. B. Giannakis. Modeling and equalization of rapidly fading channels. *Int. J. Adaptive Control & Signal Process.*, 10(2/3):159–176, Mar. 1996.

[TGYL05] S. Tomasin, A. Gorokhov, H. Yang, and J.-P. Linnartz. Iterative interference cancellation and channel estimation for mobile OFDM. *IEEE Trans. Wireless Commun.*, 4(1):238–245, Jan. 2005.

[TKS02] M. Tüchler, R. Koetter, and A. C. Singer. Turbo equalization: Principles and new results. *IEEE Trans. Commun.*, 50(5):754–767, May 2002.

[TL08] Z. Tang and G. Leus. A novel receiver architecture for single-carrier transmission over time-varying channels. *IEEE J. Sel. Areas Commun.*, 26(2):366–377, Feb. 2008.

[TM01] M. Toeltsch and A.F. Molisch. Equalization of OFDM-systems by interference cancellation techniques. In *Proc. IEEE Int. Conf. Commun.*, volume 6, pages 1950–1954, Helsinki, Finland, June 2001.

[TV05] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, New York, 2005.

[VB99] E. Viterbo and J. Boutros. A universal lattice code decoder for fading channels. *IEEE Trans. Inform. Theory*, 45(5):1639–1642, July 1999.

[Ver98] S. Verdú. *Multiuser Detection*. Cambridge, New York, 1998.

[VO79] A. J. Viterbi and J. Omura. *Principles of Digital Communication and Decoding*. McGraw-Hill, New York, 1979.

[WFGV98] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela. V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel. In *Proc. URSI ISSSE-98*, pages 295–300, Pisa, Italy, Sept. 1998.

[WM02] D. Warrier and U. Madhow. Spectrally efficient noncoherent communication. *IEEE Trans. Inform. Theory*, 48(3):651–668, Mar. 2002.

[WMG04] Z. Wang, X. Ma, and G. B. Giannakis. OFDM or single-carrier block transmissions? *IEEE Trans. Commun.*, 52(3):380–394, Mar. 2004.

[Wor96] G. W. Wornell. Emerging applications of multirate signal processing and wavelets in digital communications. *Proc. IEEE*, 84(4):586–603, Apr. 1996.

[WP99] X. Wang and H. V. Poor. Iterative (turbo) soft interference cancellation and decoding for coded CDMA. *IEEE Trans. Commun.*, 47:1046–1061, July 1999.

[WR61] J. M. Wozencraft and B. Reiffen. *Sequential Decoding*. MIT Press and Wiley, New York, 1961.

[YCL07] J. Y. Yun, S.-Y. Chung, and Y. H. Lee. Design of ICI canceling codes for OFDM systems based on capacity maximization. *IEEE Signal Process. Lett.*, 14(3):169–172, Mar. 2007.

[YP95] X. Yu and S. Pasupathy. Innovations-based MLSE for Rayleigh fading channels. *IEEE Trans. Commun.*, 43(2/3/4):1534–1544, Feb./Mar./April 1995.

[YR03] M. Yan and B. D. Rao. Soft decision-directed MAP estimate of fast Rayleigh flat fading channels. *IEEE Trans. Commun.*, 51(12):1965–1969, Dec. 2003.

[ZFG97] Y. Zhang, M. P. Fitz, and S. B. Gelfand. Soft output demodulation on frequency-selective Rayleigh fading channels using AR channel models. In *Proc. IEEE Global Telecommun. Conf.*, Oct. 1997.

[ZJP99] H. Zamiri-Jafarian and S. Pasupathy. Adaptive MLSDE using the EM algorithm. *IEEE Trans. Commun.*, 47:1181–1193, Aug. 1999.