## 3 Equalization

3.1 Intersymbol Interference and Receivers for Successive Message Transmission ........................................... 164
   3.1.1 Transmission of Successive Messages ........................................................................................................ 164
   3.1.2 Bandlimited Channels .................................................................................................................................. 166
       3.1.2.1 Noise Equivalent Pulse Response ...................................................................................................... 170
   3.1.3 The ISI-Channel Model .................................................................................................................................. 171
3.2 Basics of the Receiver-generated Equivalent AWGN ......................................................................................... 178
   3.2.1 Receiver Signal-to-Noise Ratio ..................................................................................................................... 178
   3.2.2 Receiver Biases .............................................................................................................................................. 179
   3.2.3 The Matched-Filter Bound .......................................................................................................................... 181
3.3 Nyquist Criterion .................................................................................................................................................. 184
   3.3.1 Vestigial Symmetry ........................................................................................................................................ 185
   3.3.2 Raised Cosine Pulses ..................................................................................................................................... 188
       3.3.2.1 Raised Cosine Derivation ..................................................................................................................... 189
   3.3.3 Square-Root Splitting of the Nyquist Pulse ................................................................................................. 191
       3.3.3.1 Square-Root Raised Cosine Derivation .............................................................................................. 191
3.4 Linear Zero-Forcing Equalization ......................................................................................................................... 194
   3.4.1 Performance Analysis of the ZFE ................................................................................................................ 194
   3.4.2 Noise Enhancement ....................................................................................................................................... 196
3.5 Minimum Mean-Square Error Linear Equalization ............................................................................................. 206
   3.5.1 Optimization of the Linear Equalizer ........................................................................................................ 206
   3.5.2 Performance of the MMSE-LE .................................................................................................................... 207
   3.5.3 Examples Revisited ...................................................................................................................................... 210
   3.5.4 Fractionally Spaced Equalization .............................................................................................................. 214
       3.5.4.1 Infinite-Length FSE Settings ................................................................................................................ 215
   3.5.4.1 Passband Equalization .......................................................................................................................... 217
3.6 Decision Feedback Equalization .......................................................................................................................... 219
   3.6.1 Minimum-Mean-Square-Error Decision Feedback Equalizer (MMSE-DFE) ............................................. 219
   3.6.2 Performance Analysis of the MMSE-DFE .................................................................................................. 221
   3.6.3 Zero-Forcing DFE ........................................................................................................................................ 223
   3.6.4 Examples Revisited ...................................................................................................................................... 224
3.7 Finite Length Equalizers ..................................................................................................................................... 229
   3.7.1 FIR MMSE-LE ............................................................................................................................................. 229
   3.7.2 FIR ZFE ......................................................................................................................................................... 233
   3.7.3 example ....................................................................................................................................................... 233
   3.7.4 FIR MMSE-DFE ........................................................................................................................................... 236
       3.7.4.1 FIR ZF-DFE ....................................................................................................................................... 241
   3.7.5 An Alternative Approach to the DFE ......................................................................................................... 241
       3.7.5.1 Optimizing the feed-forward and feedback filters .............................................................................. 242
   3.7.6 The Stanford DFE Program ......................................................................................................................... 245
   3.7.7 Error Propagation in the DFE ...................................................................................................................... 247
3.7.8 Look-Ahead .................................................. 250
3.8 Precoding .......................................................... 251
  3.8.1 The Tomlinson Precoder .................................. 251
    3.8.1.1 Laroia or Flexible Precoder .......................... 254
    3.8.1.2 Example and error propagation in flexible precoding . 257
  3.8.2 Partial Response Channel Models ......................... 258
    3.8.2.1 Equalizers for the Partial-response channel. ........... 260
  3.8.3 Classes of Partial Response ................................ 261
  3.8.4 Simple Precoding ........................................... 264
    3.8.4.1 Precoding the DC Notch or Modified Duobinary Channels . 267
    3.8.4.2 Precoding EPR4 ........................................ 268
  3.8.5 General Precoding .......................................... 268
  3.8.6 Quadrature PR ............................................. 269
  3.9 Diversity Equalization ....................................... 272
  3.9.1 Multiple Received Signals and the RAKE ..................... 272
  3.9.2 Infinite-length MMSE Equalization Structures ............... 275
  3.9.3 Finite-length Multidimensional Equalizers .................. 276
  3.9.4 DFE RAKE Program ......................................... 278
  3.9.5 Multichannel Transmission ................................ 280
  3.10 MIMO/Matrix Equalizers .................................... 281
    3.10.1 The MIMO Channel Model ................................ 281
    3.10.2 Matrix SNRs ........................................... 283
    3.10.3 The MIMO MMSE DFE and Equalizers ....................... 285
    3.10.4 MIMO Precoding ........................................ 290
    3.10.5 MIMO Zero-Forcing ..................................... 290
  3.11 Information Theoretic Approach to Decision Feedback ........ 291
    3.11.1 MMSE Estimation and Conditional Entropy ................. 291
    3.11.2 The relationship of the MMSE-DFE to mutual information ... 293
    3.11.3 Canonical Channel Models ................................ 296
  3.12 Construction of the Optimized DFE Input ..................... 298
    3.12.1 The Discrete-Time Paley-Wiener Criterion (PWC) and Filter Synthesis . 298
    3.12.2 Theoretical multi-tone system tool for transmit optimization .... 301
    3.12.3 Discrete-Time Water-filling and the Paley-Wiener Criterion .... 303
    3.12.4 The baseband lowpass channel with a single contiguous water-fill band . 304
    3.12.5 The single-band bandpass case .......................... 309
    3.12.6 The general multiple band case and the Optimum MMSE-DFE .......... 310
    3.12.7 Relations between Zero-Forcing and MMSE DFE receivers .......... 316
    3.12.8 Optimized Transmit Filter with Linear Equalization .......... 320
Exercises - Chapter 3 ............................................. 325

A Linear Minimum Mean-Square Estimation, Normed Vector Spaces, Spectral Factorization, and Filter Realization 354
A.1 General MMSE estimation ....................................... 354
  A.1.1 Gaussian MMSE, autocorrelation, and cross-correlation ....... 355
A.2 The Orthogonality Principle and Linear MMSE estimation ........ 356
  A.2.1 Some expansion of linear MMSE to vector estimates ........... 357
A.3 Spectral Factorization and Scalar Filter Realization ............ 360
  A.3.1 Some Transform Basics: D-Transform and Laplace Transform .... 360
  A.3.2 Autocorrelation and Non-Negative Spectra Magnitudes ........... 361
  A.3.3 The Bilinear Transform and Spectral Factorization .......... 363
  A.3.4 The Paley-Wiener Criterion ................................ 365
    A.3.4.1 Illustrative Concepts ................................ 368
  A.3.5 Linear Prediction ......................................... 370
  A.3.6 Cholesky Factorization - Finite-Length Spectral Factorization ... 371
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.3.6.1</td>
<td>Cholesky Form 1 - Forward Prediction</td>
<td>371</td>
</tr>
<tr>
<td>A.3.6.1.1</td>
<td>Cholesky and finite-length linear prediction</td>
<td>372</td>
</tr>
<tr>
<td>A.3.6.1.2</td>
<td>Forward [Upper Triangular] Cholesky Algorithm</td>
<td>373</td>
</tr>
<tr>
<td>A.3.6.2</td>
<td>Backward [Lower Triangular] Cholesky Algorithm</td>
<td>374</td>
</tr>
<tr>
<td>A.3.6.3</td>
<td>Infinite-length convergence</td>
<td>374</td>
</tr>
<tr>
<td>A.4</td>
<td>MIMO Spectral Factorization</td>
<td>375</td>
</tr>
<tr>
<td>A.4.1</td>
<td>Vector Transforms</td>
<td>375</td>
</tr>
<tr>
<td>A.4.2</td>
<td>Autocorrelation and Power Spectra for vector sequences</td>
<td>376</td>
</tr>
<tr>
<td>A.4.3</td>
<td>Factorizability for MIMO Processes</td>
<td>378</td>
</tr>
<tr>
<td>A.4.4</td>
<td>Finite-Degree MIMO polynomial factorization</td>
<td>379</td>
</tr>
<tr>
<td>A.4.4.1</td>
<td>STEP ZERO - INITIALIZATION</td>
<td>380</td>
</tr>
<tr>
<td>A.4.4.2</td>
<td>STEP ONE - Polynomial Cholesky Factorization</td>
<td>380</td>
</tr>
<tr>
<td>A.4.4.3</td>
<td>STEP TWO: Left Extraction of Poles</td>
<td>381</td>
</tr>
<tr>
<td>A.4.4.4</td>
<td>STEP THREE: Compute and factor the determinant of $G$</td>
<td>381</td>
</tr>
<tr>
<td>A.4.4.5</td>
<td>STEP FOUR: Remove the maximum-phase zeros from $G(D)$</td>
<td>382</td>
</tr>
<tr>
<td>A.4.4.6</td>
<td>STEP FIVE: remove constants to the center</td>
<td>382</td>
</tr>
<tr>
<td>A.4.4.7</td>
<td>STEP SIX: Constant-Matrix Cholesky Factorization and adjustment</td>
<td>382</td>
</tr>
<tr>
<td>A.4.4.8</td>
<td>STEP SEVEN: Restore the original poles</td>
<td>382</td>
</tr>
<tr>
<td>A.4.4.9</td>
<td>Swenson’s Example</td>
<td>383</td>
</tr>
<tr>
<td>A.4.5</td>
<td>MIMO Paley Wiener Criterion and Matrix Filter Realization</td>
<td>387</td>
</tr>
<tr>
<td>A.4.5.1</td>
<td>Analytic Functions of Matrices</td>
<td>387</td>
</tr>
<tr>
<td>A.4.5.2</td>
<td>Necessity and the Sum-Rate Equivalent</td>
<td>387</td>
</tr>
<tr>
<td>A.5</td>
<td>The Matrix Inversion Lemma</td>
<td>390</td>
</tr>
<tr>
<td>B</td>
<td>Equalization for Partial Response</td>
<td>391</td>
</tr>
<tr>
<td>B.1</td>
<td>Controlled ISI with the DFE</td>
<td>391</td>
</tr>
<tr>
<td>B.1.1</td>
<td>ZF-DFE and the Optimum Sequence Detector</td>
<td>391</td>
</tr>
<tr>
<td>B.1.2</td>
<td>MMSE-DFE and sequence detection</td>
<td>393</td>
</tr>
<tr>
<td>B.2</td>
<td>Equalization with Fixed Partial Response $B(D)$</td>
<td>393</td>
</tr>
<tr>
<td>B.2.1</td>
<td>The Partial Response Linear Equalization Case</td>
<td>393</td>
</tr>
<tr>
<td>B.2.2</td>
<td>The Partial-Response Decision Feedback Equalizer</td>
<td>394</td>
</tr>
</tbody>
</table>
Chapter 3

Equalization

Chapter 1’s main focus was a single use of the channel, signal set, and detector to transmit one of \( M \) messages or symbols, commonly referred to as “one-shot” transmission. Chapter 2 expounded upon such a system to sequences or “codewords” of symbols that each passed the channel without overlap or interference from one another. That is the channel was memoryless. In practice, successive transmissions do often interfere with one another, especially as they are sent more closely together to increase the data transmission rate. The interference between successive transmissions is called intersymbol interference (ISI). ISI can severely complicate the implementation of an optimum detector.

Figure 3.1 illustrates a receiver for detection of a succession of transmitted messages. The matched filter outputs are processed by the receiver, which outputs samples, \( z_k \), that estimate the symbol transmitted at time \( k \), \( \hat{x}_k \). Each receiver output sample is the input to the same detector that would be used on an AWGN channel without ISI. This symbol-by-symbol (SBS) detector, while optimum for the AWGN channel, will not be a maximum-likelihood estimator for the message sequence. Nonetheless, if the receiver is well-designed, the receiver-detector combination may work nearly as well as an optimum detector with far less complexity. The objective of this chapter’s receiver will be to improve the simple SBS detector’s performance. For most of this chapter, the SBS detector will be a simple scalar, but developments will eventually encompass the MIMO channel with first the receiver allowing \( L_y > 1 \) spatial dimensions in what is called “diversity” and then further expanding to \( L_x > 1 \) for full MIMO.

Equalization methods are used by communication engineers to mitigate the effects of the intersymbol interference. An equalizer is essentially the content of Figure 3.1’s receiver box. This chapter studies both intersymbol interference and several equalization methods, which amount to different structures for the receiver box. The methods presented in this chapter are not optimal for detection, but rather are widely used sub-optimal cost-effective methods that reduce the ISI. These equalization methods try to convert a bandlimited channel with ISI into one that appears memoryless, hopefully synthesizing a new AWGN-like channel at the receiver output. The designer can then analyze the resulting memoryless, equalized channel using the methods of Chapters 1 and 2 with an appropriate equalized-system SNR, as if the channel were an AWGN with that SNR. For coded systems as in Chapter 2, the decoder may no longer be SBS but it is the same decoder as would be used on the AWGN for that same SNR and does not include channel effects, grace of the equalizer. With an appropriate choice of transmit signals, one of this chapter’s methods - the Decision Feedback Equalizer can be generalized into a canonical receiver (effectively achieves the highest possible transmission rate even though not an optimum receiver).

Section 3.1 models linear intersymbol interference between successive transmissions, thereby both illustrating and measuring the ISI. In practice, as shown by a simple example, distortion from overlapping symbols can be unacceptable, suggesting that some corrective action must be taken. Section 3.1 also refines the concept of signal-to-noise ratio, which is the method used in this text to quantify receiver performance. The SNR concept will be used consistently throughout the remainder of this text as a quick and accurate means of quantifying transmission performance, as opposed to probability of error, which can be more difficult to compute, especially for suboptimum designs. As Figure 3.1 shows, the receiver’s objective will be to convert the channel into an equivalent AWGN at each time \( k \), independent of all other times \( k \). An AWGN detector may then be applied to the derived channel, and performance computed
readily using the gap approximation or other known formulas of Chapters 1 and 2 with the SNR of the derived AWGN channel. There may be loss of optimality in creating such an equivalent AWGN, which will be measured by the equivalent AWGN’s SNR with respect to the best SNR that might be expected otherwise for an optimum detector. Section 3.3 discusses some desired types of channel responses that exhibit no intersymbol interference, specifically introducing the Nyquist Criterion for a linear channel (equalized or otherwise) to be free of intersymbol interference. Section 3.4 illustrates the basic concept of equalization through the zero-forcing equalizer (ZFE), which is simple to understand but often of limited effectiveness. The more widely used and higher performance, minimum mean square error linear (MMSE-LE) and decision-feedback equalizers (MMSE-DFE) are discussed in Sections 3.5 and 3.6. Section 3.7 discusses the design of finite-length equalizers, as is needed in practice. Section 3.8 discusses precoding, a method for eliminating error propagation in decision-feedback equalizers and the related concept of partial-response channels. Section 3.9 generalizes the equalization concepts to systems with one input, but several outputs (so \( L_y > 1 \)), such as wireless transmission systems with multiple receive antennas called “diversity” receivers.

Section 3.10 generalizes the developments to MIMO channel equalization with both \( L_x > 1 \) and/or \( L_y > 1 \). Section 3.11 introduces an information-theoretic infinite-length approach to interpreting and revisiting the SISO MMSE-DFE that is then useful for optimizing the transmit filters in Section 3.12. Appendix A provides significant results on minimum mean-square estimation, including the orthogonality principle, scalar and MIMO forms of the Paley Weiner Criterion, results on linear prediction, Cholesky Factorization, and both scalar and MIMO canonical factorization results. Information measures are also here related to MMSE estimation (at least until Chapter 2 is revised and then this will be removed from Appendix A. Appendix B generalizes equalization to partial-response channels.
3.1 Intersymbol Interference and Receivers for Successive Message Transmission

Intersymbol interference is a common practical impairment found in many transmission and storage systems, including voiceband modems, digital subscriber loop data transmission, storage disks, digital mobile radio channels, digital microwave channels, and even fiber-optic (where dispersion-limited) cables. This section introduces a model for intersymbol interference. This section then continues and revisits the equivalent AWGN of Figure 3.1 in view of various receiver corrective actions for ISI.

3.1.1 Transmission of Successive Messages

Most communication systems re-use the channel to transmit several messages in succession. From Section 1.1, the message transmissions are separated by $T$ units in time, where as in Chapter 1 $T$ is the symbol period, and $1/T$ is the symbol rate. Chapter 1’s data rate for a communication system that sends one of $M$ possible messages every $T$ time units is

$$R \Delta \frac{\log_2(M)}{T} = \frac{b}{T}.$$  \hspace{1cm} (3.1)

To increase the data rate in a design, either $b$ can be increased (which requires more signal energy to maintain $P_e$) or $T$ can be decreased. Decreasing $T$ narrows the time between message transmissions and thus increases intersymbol interference on any band-limited channel.

The transmitted signal $x(t)$ corresponding to $K$ successive transmissions is

$$x(t) = \sum_{k=0}^{K-1} x_k(t - kT).$$  \hspace{1cm} (3.2)

Equation (3.2) slightly abuses previous notation in that the subscript $k$ on $x_k(t - kT)$ refers to the index associated with the $k^{th}$ successive transmission. The $K$ successive transmissions could be considered an aggregate or “block” symbol, $x(t)$, conveying one of $M^K$ possible messages. The receiver could attempt to implement MAP or ML detection for this new transmission system with $M^K$ messages. A Gram-Schmidt decomposition on the set of $M^K$ signals would then be performed and an optimum detector designed accordingly. Such an approach has complexity that grows exponentially (in proportion to $M^K$) with the block message length $K$. That is, the optimal detector might need $M^K$ matched filters, one for each possible transmitted block symbol. As $K \to \infty$, the complexity can become too large for practical implementation. Chapter 9 addresses such “sequence detectors” in detail, and it may be possible to compute the à posteriori probability function with less than exponentially growing complexity.

An alternative (suboptimal) receiver can detect each of the successive $K$ messages independently. Such detection is called symbol-by-symbol (SBS) detection. Figure 3.2 contrasts the SBS detector with the block detector of Chapter 1. The bank of matched filters, presumably found by Gram-Schmidt decomposition of the set of (noiseless) channel output waveforms (of which it can be shown $K$ dimensions are sufficient only if $N = 1$, complex or real), precedes a block detector that determines the $K$-dimensional vector symbol transmitted. The complexity would become large or infinite as $K$ becomes large or infinite for the block detector. The lower system in Figure 3.2 has a single matched filter to the channel, with output sampled $K$ times, followed by a receiver and an SBS detector.

---

1The symbol rate is sometimes also called the “baud rate,” although abuse of the term baud (by equating it with data rate even when $M \neq 2$) has rendered the term archaic among communication engineers, and the term “baud” usually now only appears in trade journals and advertisements.
The later system has fixed (and lower) complexity per symbol/sample, but may not be optimum. Interference between successive transmissions, or intersymbol interference (ISI), can degrade the performance of symbol-by-symbol detection. This performance degradation increases as $T$ decreases (or the symbol rate increases) in most communication channels. The designer mathematically analyzes ISI by rewriting (3.2) as

$$x(t) = \sum_{k=0}^{K-1} \sum_{n=1}^{N} x_{kn} \varphi_n(t-kT),$$  \hspace{1cm} (3.3)

where the transmissions $x_k(t)$ are decomposed using a common orthonormal basis set $\{\varphi_n(t)\}$. In (3.3), $\varphi_n(t-kT)$ and $\varphi_m(t-lT)$ may be non-orthogonal when $k \neq l$. In some cases, translates of the basis functions are orthogonal. For instance, in QAM, the two bandlimited basis functions

$$\varphi_1(t) = \sqrt{2/T} \cos \left( \frac{m\pi t}{T} \right) \cdot \text{sinc} \left( \frac{t}{T} \right),$$  \hspace{1cm} (3.4)

$$\varphi_2(t) = -\sqrt{2/T} \sin \left( \frac{m\pi t}{T} \right) \cdot \text{sinc} \left( \frac{t}{T} \right),$$  \hspace{1cm} (3.5)

or from Chapter 2, the baseband equivalent

$$\varphi(t) = \frac{1}{\sqrt{T}} \text{sinc} \left( \frac{t}{T} \right).$$  \hspace{1cm} (3.6)

(with $m$ a positive integer) are orthogonal for all integer-multiple-of-$T$ time translations. In this case, the successive transmissions, when sampled at time instants $kT$, are free of ISI, and transmission is equivalent to a succession of “one-shot” uses of the channel. In this case symbol-by-symbol detection is optimal, and the MAP detector for the entire block of messages is the same as a MAP detector used separately for each of the $K$ independent transmissions. Signal sets for data transmission are usually designed to be orthogonal for any translation by an integer multiple of symbol periods. Most linear
AWGN channels, however, are more accurately modeled by a filtered AWGN channel as discussed in Section 1.7 and Chapter 2. The filtering of the channel alters the basis functions so that at the channel output the filtered basis functions are no longer orthogonal. The channel thus introduces ISI.

3.1.2 Bandlimited Channels

\[ h(t) \]

\[ x(t) \]

\[ x_p(t) \]

\[ y_p(t) \]

\[ n_p(t) \]

Figure 3.3: The bandlimited channel, and equivalent forms with pulse response.

Figure 3.3’s bandlimited linear ISI channel is the same as the filtered AWGN channel discussed in Section 1.3.7. This channel is used, however, for successive transmission of data symbols. The (noise-free) channel output, \( x_p(t) \), in Figure 3.3 is given by

\[
x_p(t) = \sum_{k=0}^{K-1} \sum_{n=1}^{N} x_{kn} \cdot \varphi_n(t - kT) * h(t) \quad (3.7)
\]

\[
= \sum_{k=0}^{K-1} \sum_{n=1}^{N} x_{kn} \cdot p_n(t - kT) \quad (3.8)
\]

where \( p_n(t) \triangleq \varphi_n(t) * h(t) \). When \( h(t) \neq \delta(t) \), the functions \( p_n(t - kT) \) do not necessarily form an orthonormal basis, nor are they even necessarily orthogonal. An optimum (MAP) detector would need to search a signal set of size \( M^K \), which is often too complex for implementation as \( K \) gets large. When \( N = 1 \) (or \( N = 2 \) with complex signals), there is only one pulse response \( p(t) \).
Equalization methods apply a processor, the “equalizer”, to the channel output to try to convert \( \{ p_n(t - kT) \} \) to an orthogonal set of functions. Symbol-by-symbol detection can then be used on the equalized channel output. Further discussion of such equalization filters is deferred to Section 3.4. The remainder of this chapter also presumes that the channel-input symbol sequence \( x_k \) is independent and identically distributed at each point in time. This presumption will be relaxed in later Chapters.

While a very general theory of ISI could be undertaken for any \( N \), such a theory would unnecessarily complicate the present development. This chapter handles the ISI-equalizer case for \( N = 1 \). Using the baseband-equivalent systems of Chapter 2, this chapter’s (Chapter 3’s) analysis will also apply to quadrature modulated systems modeled as complex (equivalent to two-dimensional real) channels. In this way, the developed theory of ISI and equalization will apply equally well to any one-dimensional (e.g. PAM) or two-dimensional (e.g. QAM or hexagonal) constellation. This was the main motivation for the introduction of bandpass analysis in Chapter 2.

The pulse response for the transmitter/channel fundamentally quantifies ISI:

**Definition 3.1.1 (Pulse Response)** The pulse response of a bandlimited channel is defined by

\[
p(t) = \varphi(t) * h(t) \quad .
\]

For the complex QAM case, \( p(t), \varphi(t), \) and \( h(t) \) can be complex time functions.

The one-dimensional noiseless channel output \( x_p(t) \) is

\[
x_p(t) = \sum_{k=0}^{K-1} x_k \cdot \varphi(t - kT) * h(t)
\]

\[
= \sum_{k=0}^{K-1} x_k \cdot p(t - kT) \quad .
\]

The signal in (3.11) is real for a one-dimensional system and complex for a baseband equivalent quadrature modulated system. The pulse response energy \( \| p \|^2 \) is not necessarily equal to 1, and this text introduces the normalized pulse response:

\[
\varphi_p(t) \triangleq \frac{p(t)}{\|p\|} \quad ,
\]

where

\[
\|p\|^2 = \int_{-\infty}^{\infty} p(t) p^*(t) dt = \langle p(t), p(t) \rangle \quad .
\]

The subscript \( p \) on \( \varphi_p(t) \) indicates that \( \varphi_p(t) \) is a normalized version of \( p(t) \). Using (3.12), Equation (3.11) becomes

\[
x_p(t) = \sum_{k=0}^{K-1} x_{p,k} \cdot \varphi_p(t - kT) \quad ,
\]

where

\[
x_{p,k} \triangleq x_k \cdot \|p\| \quad .
\]

\( x_{p,k} \) absorbs the channel gain/attenuation \( \|p\| \) into the definition of the input symbol, and thus has energy \( E_p = E [\|x_{p,k}\|^2] = E_x \cdot \|p\|^2 \). While the functions \( \varphi_p(t - kT) \) are normalized, they are not necessarily orthogonal, so symbol-by-symbol detection is not necessarily optimal for the signal in (3.14).

\footnote{Chapter 4 considers multidimensional signals and intersymbol interference, while Section 3.7 considers diversity receivers that may have several observations a single or multiple channel output(s).}
EXAMPLE 3.1.1 (Intersymbol interference and the pulse response) As an example of intersymbol interference, consider the pulse response $p(t) = \frac{1}{1+t^4}$ and two successive transmissions of opposite polarity ($-1$ followed by $+1$) through the corresponding channel.

Figure 3.4: Illustration of intersymbol interference for $p(t) = \frac{1}{1+t^4}$ with $T = 1$.

Figure 3.4 illustrates the two isolated pulses with correct polarity and also the waveform corresponding to the two transmissions separated by 1 unit in time. Clearly the peaks of the pulses have been displaced in time and also significantly reduced in amplitude. Higher transmission rates would force successive transmissions to be closer and closer together. Figure 3.5 illustrates the
resultant sum of the two waveforms for spacings of 1 unit in time, .5 units in time, and .1 units in time. Clearly, ISI has the effect of severely reducing pulse strength, thereby reducing immunity to noise.

**EXAMPLE 3.1.2 (Pulse Response Orthogonality - Modified Duobinary)** A PAM modulated signal using rectangular pulses is

\[
\varphi(t) = \frac{1}{\sqrt{T}}(u(t) - u(t - T)) \quad .
\]  

(3.16)

The channel introduces ISI, for example, according to

\[
h(t) = \delta(t) + \delta(t - T).
\]  

(3.17)

The resulting pulse response is

\[
p(t) = \frac{1}{\sqrt{T}}(u(t) - u(t - 2T))
\]  

(3.18)

and the normalized pulse response is

\[
\varphi_p(t) = \frac{1}{\sqrt{2T}}(u(t) - u(t - 2T)).
\]  

(3.19)

The pulse-response translates \(\varphi_p(t)\) and \(\varphi_p(t - T)\) are not orthogonal, even though \(\varphi(t)\) and \(\varphi(t - T)\) were originally orthonormal.
3.1.2.1 Noise Equivalent Pulse Response

Figure 3.1.2.1 models a channel with additive Gaussian noise that is not white, which often occurs in practice. The power spectral density of the noise is $\frac{N_0}{2} \cdot S_n(f)$.

When $S_n(f) \neq 0$ (noise is never exactly zero at any frequency in practice), the noise psd has an invertible square root as in Section 1.7. The invertible square-root can be realized as a filter in the
beginning of a receiver. Since this filter is invertible, by the reversibility theorem of Chapter 1, no information is lost. The designer can then construe this filter as being pushed back into, and thus a part of, the channel as shown in the lower part of Figure 3.1.2.1. The noise equivalent pulse response then has Fourier Transform \( P(f)/S_{1/2}^1(f) \) for an equivalent filtered-AWGN channel. The concept of noise equivalence allows an analysis for AWGN to be valid (using the equivalent pulse response instead of the original pulse response). Then also “colored noise” is equivalent in its effect to ISI, and furthermore the compensating equalizers that are developed later in this chapter can also be very useful on channels that originally have no ISI, but that do have “colored noise.” An AWGN channel with a notch in \( H(f) \) at some frequency is thus equivalent to a “flat channel” with \( H(f) = 1 \), but with narrow-band Gaussian noise at the same frequency as the notch, as illustrated in Figure 3.8.

![Figure 3.8: Two “noise-equivalent” channels.](image)

3.1.3 The ISI-Channel Model

A model for linear ISI channels is shown in Figure 3.1.3. In this model, \( x_k \) is scaled by \( \|p\| \) to form \( x_{p,k} \) so that \( E_x = E_{x_p} \cdot \|p\|^2 \). The additive noise is white Gaussian, although correlated Gaussian noise can be included by transforming the correlated-Gaussian-noise channel into an equivalent white Gaussian noise channel using the methods in the previous subsection and illustrated in Figure 3.1.2.1. The channel output \( y_p(t) \) is passed through a matched filter \( \varphi_p^*(t) \) to generate \( y(t) \). Then, \( y(t) \) is sampled at the symbol rate and subsequently processed by a discrete time receiver. The following theorem illustrates that there is no loss in performance that is incurred via the matched-filter/sampler combination.

**Theorem 3.1.1 (ISI-Channel Model Sufficiency)** The discrete-time signal samples \( y_k = y(kT) \) in Figure 3.1.3 are sufficient to represent the continuous-time ISI-model.
channel output \( y(t) \), if \( 0 < \| p \| < \infty \). (i.e., a receiver with minimum \( P_e \) can be designed that uses only the samples \( y_k \)).

**Sketch of Proof:**

Define

\[
\varphi_{p,k}(t) \triangleq \varphi_p(t - kT) ,
\]

(3.20)

where \( \{ \varphi_{p,k}(t) \}_{k \in (-\infty, \infty)} \) is a linearly independent set of functions. The set \( \{ \varphi_{p,k}(t) \}_{k \in (-\infty, \infty)} \) is related to a set of orthogonal basis functions \( \{ \phi_{p,k}(t) \}_{k \in (-\infty, \infty)} \) by an invertible transformation \( \Gamma \) (use Gram-Schmidt an infinite number of times). The transformation and its inverse are written

\[
\{ \varphi_{p,k}(t) \}_{k \in (-\infty, \infty)} = \Gamma(\{ \varphi_{p,k}(t) \}_{k \in (-\infty, \infty)})
\]

(3.21)

\[
\{ \phi_{p,k}(t) \}_{k \in (-\infty, \infty)} = \Gamma^{-1}(\{ \varphi_{p,k}(t) \}_{k \in (-\infty, \infty)})
\]

(3.22)

where \( \Gamma \) is the invertible transformation. In Figure 3.1.3, the transformation outputs are the filter samples \( y(kT) \). The infinite set of filters \( \{ \phi_{p,k}(t) \}_{k \in (-\infty, \infty)} \) followed by \( \Gamma^{-1} \) is equivalent to an infinite set of matched filters to \( \{ \varphi_{p,k}(t) \}_{k \in (-\infty, \infty)} \). By (3.20) this last set is equivalent to a single matched filter \( \varphi_p^*(-t) \), whose output is sampled at \( t = kT \) to produce \( y(kT) \). Since the set \( \{ \varphi_{p,k}(t) \}_{k \in (-\infty, \infty)} \) is orthonormal, the set of sampled filter outputs in Figure 3.1.3 are sufficient to represent \( y_p(t) \). Since \( \Gamma^{-1} \) is invertible (inverse is \( \Gamma \)), then by the theorem of reversibility in Chapter 1, the sampled matched filter output \( y(kT) \) is a sufficient representation of the ISI-channel output \( y_p(t) \).

QED.
Referring to Figure 3.1.3,

\[ y(t) = \sum_k \|p\| \cdot x_k q(t - kT) + n_p(t) * \varphi_p^*(-t) \quad , \]

where

\[ q(t) \triangleq \varphi_p(t) * \varphi_p^*(-t) = \frac{p(t) * p^*(-t)}{\|p\|^2} . \]

The deterministic autocorrelation function \( q(t) \) is Hermitian \((q^*(-t) = q(t))\). Also, \( q(0) = 1 \), so the symbol \( x_k \) passes at time \( kT \) to the output with amplitude scaling \( \|p\| \). The function \( q(t) \) can also exhibit ISI, as illustrated in Figure 3.11. The plotted \( q(t) \) corresponds to \( q_k = [-1.159 \quad 0.2029 \quad 1 \quad 0.2029 \quad -1.159] \) or, equivalently, to the channel \( p(t) = \sqrt{\frac{1}{T}} \cdot \left( \text{sinc}(t/T) + 0.25 \cdot \text{sinc}((t - T)/T) - 0.125 \cdot \text{sinc}((t - 2T)/T) \right) \), or \( P(D) = \frac{1}{\sqrt{T}} \cdot (1 + 0.5D)(1 - 0.25D) \) (the notation \( P(D) \) is defined in Appendix A). (The values for \( q_k \) can be confirmed by convolving \( p(t) \) with its time reverse, normalizing, and sampling.) For notational brevity, let \( y_k \triangleq y(kT), q_k \triangleq q(kT), n_k \triangleq n(kT) \) where \( n(t) \triangleq n_p(t) * \varphi_p^*(-t) \). Thus

\[ y_k = \underbrace{\|p\| \cdot x_k}_{\text{scaled input (desired)}} + \underbrace{n_k}_{\text{noise}} + \underbrace{\|p\| \cdot \sum_{m \neq k} x_m q_{k-m}}_{\text{ISI}} \quad . \]

173
The output $y_k$ consists of the scaled input, noise and ISI. The scaled input is the desired information-bearing signal. The ISI and noise are unwanted signals that act to distort the information being transmitted. The ISI represents a new distortion component not previously considered in the analysis of Chapters 1 and 2 for a suboptimum SBS detector. This SBS detector is the same detector as in Chapters 1 and 2, except used under the (false) assumption that the ISI is just additional AWGN. Such a receiver can be decidedly suboptimum when the ISI is nonzero.

Using $D$-transform notation, (3.25) becomes

$$Y(D) = X(D) \cdot \|p\| \cdot Q(D) + N(D)$$

(3.26)

where $Y(D) = \sum_{k=-\infty}^{\infty} y_k \cdot D^k$. If the receiver uses symbol-by-symbol detection on the sampled output $y_k$, then the noise sample $n_k$ of (one-dimensional) variance $\frac{N_0}{2}$ at the matched-filter output combines with ISI from the sample times $mT$ ($m \neq k$) in corrupting $\|p\| \cdot x_k$.

There are two common measures of ISI distortion. The first is **Peak Distortion**, which only has meaning for real-valued $q(t)$:

**Definition 3.1.2 (Peak Distortion Criterion)** If $|x_{max}|$ is the maximum value for $|x_k|$, then the peak distortion is:

$$D_p \triangleq |x|_{max} \cdot \|p\| \cdot \sum_{m \neq 0} |q_m|$$

(3.27)

---

3In the real case, the magnitudes correspond to actual values. However, for complex-valued terms, the ISI is characterized by both its magnitude and phase. So, addition of the magnitudes of the symbols ignores the phase components, which may significantly change the ISI term.
For $q(t)$ in Figure 3.11 with $x_{max} = 3$, $D_p = 3 \cdot \|p\|(1.1159 + 0.2029 + 0.2029 + 0.1159) \approx 3 \cdot \sqrt{1.078 \cdot 0.6376} \approx 1.99$.

The peak distortion represents a worst-case loss in minimum distance between signal points in the signal constellation for $x_k$, or equivalently

$$P_e \leq N_e Q \left[ \frac{\|p\|^2}{2} - D_p \right], \quad (3.28)$$

for symbol-by-symbol detection. Consider two matched filter outputs $y_k$ and $y_k'$ that the receiver attempts to distinguish by suboptimally using symbol-by-symbol detection. These outputs are generated by two different sequences $\{x_k\}$ and $\{x'_k\}$. Without loss of generality, assume $y_k > y_k'$, and consider the difference

$$y_k - y_k' = \|p\| \left[ (x_k - x'_k) + \sum_{m \neq k} (x_m - x'_m)q_{k-m} \right] + \tilde{n} \quad (3.29)$$

The summation term inside the brackets in (3.29) represents the change in distance between $y_k$ and $y_k'$ caused by ISI. Without ISI the distance is

$$y_k - y_k' \geq \|p\| \cdot d_{min}, \quad (3.30)$$

while with ISI the distance can decrease to

$$y_k - y_k' \geq \|p\| \left[ d_{min} - 2|x|_{max} \sum_{m \neq 0} |q_m| \right]. \quad (3.31)$$

Implicitly, the distance interpretation in (3.31) assumes $2D_p \leq \|p\|d_{min}$.

While peak distortion represents the worst-case ISI, this worst case might not occur very often in practice. For instance, with an input alphabet size $M = 4$ and a $q(t)$ that spans 15 symbol periods, the probability of occurrence of the worst-case value (worst level occurring in all 14 ISI contributors) is $4^{-14} = 3.7 \times 10^{-9}$, well below typical channel $P_e$’s in data transmission. Nevertheless, there may be other ISI patterns of nearly just as bad interference that can also occur. Rather than separately compute each possible combination’s reduction of minimum distance, its probability of occurrence, and the resulting error probability, data transmission engineers more often use a measure of ISI called Mean-Square Distortion (valid for 1 or 2 dimensions):

**Definition 3.1.3 (Mean-Square Distortion)** The Mean-Square Distortion is defined by:

$$D_{ms} \triangleq E \left\{ \left| \sum_{m \neq k} x_{p,m} \cdot q_{k-m} \right|^2 \right\} \quad (3.32)$$

$$= \mathcal{E} \cdot \|p\|^2 \cdot \sum_{m \neq 0} |q_m|^2, \quad (3.33)$$

where (3.33) is valid when the successive data symbols are independent and identically distributed with zero mean.

In the example of Figure 3.11, the mean-square distortion (with $\mathcal{E} = 5$) is $D_{ms} = 5\|p\|^2(1.1159^2 + 0.2029^2 + 0.2029^2 + 0.1159^2) \approx 5(1.078).109 \approx .588$. The fact $\sqrt{.588} = .767 < 1.99$ illustrates that $D_{ms} \leq D_p$. (The proof of this fact is left as an exercise to the reader.)

---

4On channels for which $2D_p \geq \|p\|d_{min}$, the worst-case ISI occurs when $|2D_p - \|p\|d_{min}|$ is maximum.
The mean-square distortion criterion assumes (erroneously\(^5\)) that \(D_{ms}\) is the variance of an uncorrelated Gaussian noise that is added to \(n_k\). With this assumption, \(P_e\) is approximated by

\[
P_e \approx N_e \cdot Q \left( \frac{\|p\|d_{\text{min}}}{2\sqrt{\sigma^2 + D_{ms}}} \right).
\]

(3.34)

One way to visualize ISI is through the “eye diagram”, some examples of which are shown in Figures 3.12 and 3.13. The eye diagram is similar to what would be observed on an oscilloscope, when the trigger is synchronized to the symbol rate. The eye diagram is produced by overlaying several successive symbol intervals of the modulated and filtered continuous-time waveform (except Figures 3.12 and 3.13 do not include noise). The Lorentzian pulse response \(p(t) = 1/(1 + (3t/T)^2)\) is used in both plots. For binary transmission on this channel, there is a significant opening in the eye in the center of the plot in Figure 3.12. With 4-level PAM transmission, the openings are much smaller, leading to less noise immunity. The ISI causes the spread among the path traces; more ISI results in a narrower eye opening. Clearly increasing \(M\) reduces the eye opening.

![Figure 3.12: Binary eye diagram for a Lorentzian pulse response.](image)

---

\(^5\)This assumption is only true when \(x_k\) is Gaussian. In very well-designed data transmission systems, \(x_k\) is approximately i.i.d. and Gaussian, see Chapter 6, so that this approximation of Gaussian ISI becomes accurate.
Figure 3.13: 4-Level eye diagram for a Lorentzian pulse response.
3.2 Basics of the Receiver-generated Equivalent AWGN

Figure 3.2 focuses upon the receiver and specifically the device shown generally as $R$. When channels have ISI, such a receiving device is inserted at the sampler output. The purpose of the receiver is to attempt to convert the channel into an equivalent AWGN that is also shown below the dashed line. Such an AWGN is not always exactly achieved, but nonetheless any deviation between the receiver output $z_k$ and the channel input symbol $x_k$ is viewed as additive white Gaussian noise. An SNR, as in Subsection 3.2.1 can be used then to analyze the performance of the symbol-by-symbol detector that follows the receiver $R$. Usually, smaller deviation from the transmitted symbol means better performance, although not exactly so as Subsection 3.2.2 discusses.

![Diagram of receiver and equivalent AWGN](image)

Subsection 3.2.3 finishes this section with a discussion of the highest possible SNR that a designer could expect for any filtered AWGN channel, the so-called “matched-filter-bound” SNR, $\text{SNR}_{\text{MFB}}$. This section shall not be specific as to the content of the box shown as $R$, but later sections will allow both linear and slightly nonlinear structures that may often be good choices because their performance can be close to $\text{SNR}_{\text{MFB}}$.

### 3.2.1 Receiver Signal-to-Noise Ratio

**Definition 3.2.1 (Receiver SNR)** The receiver SNR, $\text{SNR}_R$ for any receiver $R$ with (pre-decision) output $z_k$, and decision regions based on $x_k$ (see Figure 3.2) is

$$\text{SNR}_R = \frac{E_x}{E_{\mathbb{E}[e_k]}^2} \quad (3.35)$$

where $e_k = x_k - z_k$ is the receiver error. The denominator of (3.35) is the mean-square error $\text{MSE}=E|e_k|^2$. When $E[z_k|x_k] = x_k$, the receiver is unbiased (otherwise biased) with respect to the decision regions for $x_k$.

The concept of a receiver SNR facilitates evaluation of the performance of data transmission systems with various compensation methods (i.e., equalizers) for ISI. Use of SNR as a performance measure builds upon the simplifications of considering mean-square distortion, that is both noise and ISI are
jointly considered in a single measure. The two right-most terms in (3.25) have normalized mean-square value $\sigma^2 + \overline{D_{ms}}$. The SNR for the matched filter output $y_k$ in Figure 3.2 is the ratio of channel output sample energy $\overline{E_x}p^2$ to the mean-square distortion $\sigma^2 + \overline{D_{ms}}$. This SNR is often directly related to probability of error and is a function of both the receiver and the decision regions for the SBS detector. This text uses SNR consistently, replacing probability of error as a measure of comparative performance. SNR is easier to compute than $P_e$, independent of $M$ at constant $\overline{E_x}$, and a generally good measure of performance: higher SNR means lower probability of error. The probability of error is difficult to compute exactly because the distribution of the ISI-plus-noise is not known or is difficult to compute. The SNR is easier to compute and this text assumes that the insertion of the appropriately scaled SNR (see Chapter 1 - Sections 1.4 - 1.6)) into the argument of the Q-function approximates the probability of error for the suboptimum SBS detector. Even when this insertion into the Q-function is not sufficiently accurate, comparison of SNR’s for different receivers usually relates which receiver is better.

### 3.2.2 Receiver Biases

Figure 3.15: Receiver SNR concept.

Figure 3.2.2 illustrates a receiver that somehow has tried to reduce the combination of ISI and noise. Any time-invariant receiver’s output samples, $z_k$, satisfy

$$z_k = \alpha \cdot (x_k + u_k)$$

(3.36)

where $\alpha$ is some positive scale factor that may have been introduced by the receiver and $u_k$ is an uncorrelated distortion

$$u_k = \sum_{m \neq 0} r_m \cdot x_{k-m} + \sum_{m} f_m \cdot n_{k-m}$$

(3.37)

The coefficients for residual intersymbol interference $r_m$ and the coefficients of the filtered noise $f_k$ will depend on the receiver and generally determine the level of mean-square distortion. The uncorrelated distortion has no remnant of the current symbol being decided by the SBS detector, so that $E[u_k/x_k] = 0$. However, the receiver may have found that by scaling (reducing) the $x_k$ component in $z_k$ by $\alpha$ that the SNR improves (small signal loss in exchange for larger uncorrelated distortion reduction). When $E[z_k/x_k] = \alpha \cdot x_k \neq x_k$, the decision regions in the SBS detector are “biased.” Removal of the bias is easily achieved by scaling by $1/\alpha$ as also in Figure 3.2.2. If the distortion is assumed to be Gaussian noise, as is the assumption with the SBS detector, then removal of bias by scaling by $1/\alpha$ improves the probability of error of such a detector as in Chapter 1. (Even when the noise is not Gaussian as is the case with the ISI component, scaling the signal correctly improves the probability of error on the average if the input constellation has zero mean.)
The following theorem relates the SNR’s of the unbiased and biased decision rules for any receiver \( R \):

**Theorem 3.2.1 (Unconstrained and Unbiased Receivers)** Given an unbiased receiver \( R \) for a decision rule based on a signal constellation corresponding to \( x_k \), the maximum unconstrained SNR corresponding to that same receiver with any biased decision rule is

\[
SNR_R = SNR_{R,U} + 1,
\]

where \( SNR_{R,U} \) is the SNR using the unbiased decision rule.

**Proof:** From Figure 3.2.2, the SNR after scaling is easily

\[
SNR_{R,U} = \frac{\mathcal{E}_x}{\bar{\sigma}_e^2}.
\]

The maximum SNR for the biased signal \( z_k \) prior to the scaling occurs when \( \alpha \) is chosen to maximize the unconstrained SNR

\[
SNR_R = \frac{\mathcal{E}_x}{|\alpha|^2 \sigma_u^2 + |1 - \alpha|^2 \bar{\sigma}_e^2}.
\]

Allowing for complex \( \alpha \) with phase \( \theta \) and magnitude \( |\alpha| \), the SNR maximization over alpha is equivalent to minimizing

\[
1 - 2|\alpha|\cos(\theta) + |\alpha|^2 (1 + \frac{1}{SNR_{R,U}}).
\]

Clearly \( \theta = 0 \) for a minimum and differentiating with respect to \( |\alpha| \) yields

\[
-2 + 2|\alpha|(1 + \frac{1}{SNR_{R,U}}) = 0
\]

or \( \alpha_{opt} = 1/(1 + (SNR_{R,U})^{-1}) \). Substitution of this value into the expression for \( SNR_R \) finds

\[
SNR_R = SNR_{R,U} + 1.
\]

Thus, a receiver \( R \) and a corresponding SBS detector that have zero bias will not correspond to a maximum SNR – the SNR can be improved by scaling (reducing) the receiver output by \( \alpha_{opt} \). Conversely, a receiver designed for maximum SNR can be altered slightly through simple output scaling by \( 1/\alpha_{opt} \) to a related receiver that has no bias and has SNR thereby reduced to \( SNR_{R,U} = SNR_R - 1 \). QED.

To illustrate the relationship of unbiased and biased receiver SNRs, suppose an ISI-free AWGN channel has an SNR=10 with \( \mathcal{E}_x = 1 \) and \( \sigma_e^2 = N_0^2 = .1 \). Then, a receiver could scale the channel output by \( \alpha = 10/11 \). The resultant new error signal is \( e_k = x_k(1 - \frac{10}{11}) - \frac{10}{11}n_k \), which has MSE=\( E[|e_k|^2] = \frac{1}{11} + \frac{100}{11}(1) = \frac{1}{11} \) and SNR=11. Clearly, the biased SNR is equal to the unbiased SNR plus 1. The scaling has done nothing to improve the system, and the appearance of an improved SNR is an artifact of the SNR definition, which allows noise to be scaled down without taking into account the fact that actual signal power after scaling has also been reduced. Removing the bias corresponds to using the actual signal power, and the corresponding performance-characterizing SNR can always be found by subtracting 1 from the biased SNR. A natural question is then “Why compute the biased SNR?” The answer is that the biased receiver corresponds directly to minimizing the mean-square distortion, and the SNR for the “MMSE” case will often be easier to compute. Figure 3.5.2 in Section 3.5 illustrates the usual situation of removing a bias (and consequently reducing SNR, but not improving \( P_e \) since the SBS detector works best when there is no bias) from a receiver that minimizes mean-square distortion (or error) to get an unbiased decision. The bias from a receiver that maximizes SNR by equivalently minimizing mean-square error can then be removed by simple scaling and the resultant more accurate SNR is thus found for the unbiased receiver by subtracting 1 from the more easily computed biased receiver. This concept will be very useful in evaluating equalizer performance in later sections of this chapter, and is formalized in Theorem 3.2.2 below.
Theorem 3.2.2 (Unbiased MMSE Receiver Theorem) Let \( \mathcal{R} \) be any allowed class of receivers \( R \) producing outputs \( z_k \), and let \( R_{\text{opt}} \) be the receiver that achieves the maximum signal-to-noise ratio \( \text{SNR}(R_{\text{opt}}) \) over all \( R \in \mathcal{R} \) with an unconstrained decision rule. Then the receiver that achieves the maximum SNR with an unbiased decision rule is also \( R_{\text{opt}} \), and
\[
\max_{R \in \mathcal{R}} \text{SNR}_{R,U} = \text{SNR}(R_{\text{opt}}) - 1.
\]

Proof. From Theorem 3.2.1, for any \( R \in \mathcal{R} \), the relation between the signal-to-noise ratios of unbiased and unconstrained decision rules is \( \text{SNR}_{R,U} = \text{SNR}_R - 1 \), so
\[
\max_{R \in \mathcal{R}} [\text{SNR}_{R,U}] = \max_{R \in \mathcal{R}} [\text{SNR}_R] - 1 = \text{SNR}_{R_{\text{opt}}} - 1.
\]

QED.

This theorem implies that the optimum unbiased receiver and the optimum biased receiver settings are identical except for any scaling to remove bias; only the SNR measures are different. For any SBS detector, \( \text{SNR}_{R,U} \) is the SNR that corresponds to best \( P_e \). The quantity \( \text{SNR}_{R,U} + 1 \) is artificially high because of the bias inherent in the general SNR definition.

3.2.3 The Matched-Filter Bound

The Matched-Filter Bound (MFB), also called the “one-shot” bound, specifies an upper SNR limit on the performance of data transmission systems with ISI.

**Lemma 3.2.1 (Matched-Filter Bound SNR)** The SNR\(_{\text{MFB}}\) is the SNR that characterizes the best achievable performance for a given pulse response \( p(t) \) and signal constellation (on an AWGN channel) if the channel is used to transmit only one message. This SNR is
\[
\text{SNR}_{\text{MFB}} = E_x [\|p\|^2] / N_0 / 2
\]

MFB denotes the square of the argument to the Q-function that arises in the equivalent “one-shot” analysis of the channel.

Proof: Given a channel with pulse response \( p(t) \) and isolated input \( x_0 \), the maximum output sample of the matched filter is \( \|p\| \cdot x_0 \). The normalized average energy of this sample is \( \|p\|^2 \tilde{E}_x \), while the corresponding noise sample energy is \( N_0 / 2 \), \( \text{SNR}_{\text{MFB}} = \tilde{E}_x / N_0 / 2 \). QED.

The probability of error, measured after the matched filter and prior to the symbol-by-symbol detector, satisfies \( P_e \geq N_e \cdot Q(\sqrt{\text{MFB}}) \). When \( \tilde{E}_x \) equals \( (d_{\text{min}}^2/4) / \kappa \), then MFB equals \( \text{SNR}_{\text{MFB}} \cdot \kappa \). In effect the MFB forces no ISI by disallowing preceding or successive transmitted symbols. An optimum detector is used for this “one-shot” case. The performance is tacitly a function of the transmitter basis functions, implying performance is also a function of the symbol rate \( 1/T \). No other (for the same input constellation) receiver for continuous transmission could have better performance, if \( x_k \) is an i.i.d. sequence, since the sequence must incur some level of ISI. The possibility of correlating the input sequence \{\( x_k \)\} to take advantage of the channel correlation will be considered in Chapters 4 and 5.

The following example illustrates computation of the MFB for several cases of practical interest:

**Example 3.2.1 (Binary PAM)** For binary PAM,
\[
x_p(t) = \sum_k x_k \cdot p(t - kT)
\]
where \( x_k = \pm \sqrt{E_x} \). The minimum distance at the matched-filter output is \( \| p \| \cdot d_{\text{min}} = \| p \| \cdot d = 2 \cdot \| p \| \cdot \sqrt{E_x} \), so \( E_x = \frac{d_{\text{min}}^2}{4} \) and \( \kappa = 1 \). Then,

\[
\text{MFB} = \text{SNR}_{MFB} \ .
\]

Thus for a binary PAM channel, the MFB (in dB) is just the “channel-output” SNR, SNR_{MFB}. If the transmitter symbols \( x_k \) are equal to \( \pm 1 \) (\( E_x = 1 \)), then

\[
\text{MFB} = \frac{\| p \|^2}{\sigma^2} \ ,
\]
where, again, \( \sigma^2 = \frac{N_0}{2} \). The binary-PAM \( P_e \) is then bounded by

\[
P_e \geq Q\left( \sqrt{\text{SNR}_{MFB}} \right) \ .
\]

**EXAMPLE 3.2.2 (M-ary PAM)** For M-ary PAM, \( x_k = \pm \frac{d}{2}, \pm \frac{3d}{2}, \ldots, \pm \frac{(M-1)d}{2} \) and

\[
d = \sqrt{\frac{3E_x}{M^2 - 1}} \ ,
\]
so \( \kappa = 3/(M^2 - 1) \). Thus,

\[
\text{MFB} = \frac{3}{M^2 - 1} \text{SNR}_{MFB} \ ,
\]
for \( M \geq 2 \). If the transmitter symbols \( x_k \) are equal to \( \pm 1, \pm 3, \ldots, \pm (M - 1) \), then

\[
\text{MFB} = \frac{\| p \|^2}{\sigma^2} \ .
\]
Equation (3.52) is the same result as (3.48), which should be expected since the minimum distance is the same at the transmitter, and thus also at the channel output, for both (3.48) and (3.52). The M′ary-PAM \( P_e \) is then bounded by

\[
P_e \geq 2(1 - 1/M) \cdot Q\left( \sqrt{\frac{3 \cdot \text{SNR}_{MFB}}{M^2 - 1}} \right) \ .
\]

**EXAMPLE 3.2.3 (QPSK)** For QPSK, \( x_k = \pm \frac{d}{2} \pm j \frac{d}{2}, \) and \( d = 2 \sqrt{E_x} \), so \( \kappa = 1 \). Thus

\[
\text{MFB} = \text{SNR}_{MFB} \ .
\]
Thus, for a QPSK (or 4SQ QAM) channel, MFB (in dB) equals the channel output SNR. If the transmitter symbols \( x_k \) are \( \pm 1 \pm j \), then

\[
\text{MFB} = \frac{\| p \|^2}{\sigma^2} \ .
\]
The best QPSK \( P_e \) is then approximated by

\[
P_e \approx Q\left( \sqrt{\text{SNR}_{MFB}} \right) \ .
\]

**EXAMPLE 3.2.4 (M-ary QAM Square)** For M-ary QAM, \( \Re\{x_k\} = \pm \frac{d}{2}, \pm \frac{3d}{2}, \ldots, \pm \frac{(M-1)d}{2} \), \( \Im\{x_k\} = \pm \frac{d}{2}, \pm \frac{3d}{2}, \ldots, \pm \frac{(M-1)d}{2} \), (recall that \( \Re \) and \( \Im \) denote real and imaginary parts, respectively) and

\[
d = \frac{3E_x}{M - 1} \ ,
\]
so $\kappa = 3/(M - 1)$. Thus

$$MFB = \frac{3}{M - 1} \text{SNR}_{MFB},$$

(3.58)

for $M \geq 4$. If the real and imaginary components of the transmitter symbols $x_k$ equal $\pm 1, \pm 3, \ldots, \pm(\sqrt{M} - 1)$, then

$$MFB = \frac{||p||^2}{\sigma^2}.$$  

(3.59)

The best M'ary QAM $P_e$ is then approximated by

$$\tilde{P}_e \approx 2(1 - 1/\sqrt{M}) \cdot Q(\sqrt{\frac{3 \cdot \text{SNR}_{MFB}}{M - 1}}).$$

(3.60)

In general for square QAM constellations,

$$MFB = \frac{3}{4^b - 1} \text{SNR}_{MFB}.$$  

(3.61)

For the QAM Cross constellations,

$$MFB = \frac{2E_x ||p||^2}{\frac{4}{M - \frac{4}{3}} \sigma^2} = \frac{96}{31 \cdot 4^b - 32} \text{SNR}_{MFB}.$$  

(3.62)

For the suboptimum receivers to come in later sections, $\text{SNR}_U \leq \text{SNR}_{MFB}$. As $\text{SNR}_U \to \text{SNR}_{MFB}$, then the receiver is approaching the bound on performance. It is not always possible to design a receiver that attains $\text{SNR}_{MFB}$, even with infinite complexity, unless one allows co-design of the input symbols $x_k$ in a channel-dependent way (see Chapters 4 and 5). The loss with respect to matched filter bound will be determined for any receiver by $\text{SNR}_u/\text{SNR}_{MFB} \leq 1$, in effect determining a loss in signal power because successive transmissions interfere with one another – it may well be that the loss in signal power is an acceptable exchange for a higher rate of transmission.
3.3 Nyquist Criterion

The Nyquist Criterion specifies the conditions on \( q(t) = \varphi_p(t) * \varphi_p^*(-t) \) for an ISI-free channel on which a symbol-by-symbol detector is optimal. This section first reviews some fundamental relationships between \( q(t) \) and its samples \( q_k = q(kT) \) in the frequency domain.

\[
q(kT) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Q(\omega)e^{j\omega kT} d\omega
\]

(3.63)

\[
= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{\frac{(2n+1)\pi}{T}}^{\frac{(2n-1)\pi}{T}} Q(\omega)e^{j\omega kT} d\omega
\]

(3.64)

\[
= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} Q(\omega + \frac{2\pi n}{T})e^{j(\omega + \frac{2\pi n}{T})kT} d\omega
\]

(3.65)

\[
= \frac{1}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} Q_{eq}(\omega)e^{j\omega kT} d\omega ,
\]

(3.66)

where \( Q_{eq}(\omega) \), the equivalent frequency response becomes

\[
Q_{eq}(\omega) \triangleq \sum_{n=-\infty}^{\infty} Q(\omega + \frac{2\pi n}{T}) .
\]

(3.67)

The function \( Q_{eq}(\omega) \) is periodic in \( \omega \) with period \( \frac{2\pi}{T} \). This function is also known as the folded or aliased spectrum of \( Q(\omega) \) because the sampling process causes the frequency response outside of the fundamental interval \((-\frac{\pi}{T}, \frac{\pi}{T})\) to be added (i.e. “folded in”). Writing the Fourier Transform of the sequence \( q_k \) as \( Q(e^{-j\omega T}) = \sum_{k=-\infty}^{\infty} q_k e^{-j\omega kT} \) leads to

\[
\frac{1}{T} \cdot Q_{eq}(\omega) = Q(e^{-j\omega T}) \triangleq \sum_{k=-\infty}^{\infty} q_k e^{-j\omega kT} ,
\]

(3.68)

a well-known relation between the discrete-time and continuous-time representations of any waveform in digital signal processing.

It is now straightforward to specify Nyquist’s Criterion:

**Theorem 3.3.1 (Nyquist’s Criterion)** A channel specified by pulse response \( p(t) \) (and resulting in \( q(t) = \frac{p(t) * p^*(-t)}{\|p\|^2} \)) is ISI-free if and only if

\[
Q(e^{-j\omega T}) = \frac{1}{T} \sum_{n=-\infty}^{\infty} Q(\omega + \frac{2\pi n}{T}) = 1 .
\]

(3.69)

**Proof:**

By definition the channel is ISI-free if and only if \( q_k = 0 \) for all \( k \neq 0 \) (recall \( q_0 = 1 \) by definition). The proof follows directly by substitution of \( q_k = \delta_k \) into (3.68). QED.

Functions that satisfy (3.69) are called “Nyquist pulses.” One function that satisfies Nyquist’s Criterion is

\[
q(t) = \text{sinc} \left( \frac{t}{T} \right) ,
\]

(3.70)

which corresponds to normalized pulse response

\[
\varphi_p(t) = \frac{1}{\sqrt{T}} \text{sinc} \left( \frac{t}{T} \right) .
\]

(3.71)
The function $q(kT) = \text{sinc}(k) = \delta_k$ satisfies the ISI-free condition. One feature of $\text{sinc}(t/T)$ is that it has minimum bandwidth for no ISI.

![Figure 3.16: The $\text{sinc}(t/T)$ function.](image)

No other function has this same minimum bandwidth and also satisfies the Nyquist Criterion. (Proof is left as an exercise to the reader.) The $\text{sinc}(t/T)$ function is plotted in Figure 3.16 for $-20T \leq t \leq 20T$.

The frequency $\frac{1}{2T}$ (1/2 the symbol rate) is often construed as the maximum frequency of a sampled signal that can be represented by samples at the sampling rate. In terms of positive-frequencies, $1/2T$ represents a minimum bandwidth necessary to satisfy the Nyquist Criterion, and thus has a special name in data transmission:

**Definition 3.3.1 (Nyquist Frequency)** The frequency $\omega = \frac{\pi}{T}$ or $f = \frac{1}{2T}$ is called the Nyquist frequency.

### 3.3.1 Vestigial Symmetry

In addition to the $\text{sinc}(t/T)$ Nyquist pulse, data-transmission engineers use responses with up to twice the minimum bandwidth. For all these pulses, $Q(\omega) = 0$ for $|\omega| > \frac{\pi}{2T}$. These wider bandwidth responses will provide more immunity to timing errors in sampling as follows: The $\text{sinc}(t/T)$ function decays in amplitude only linearly with time. Thus, any sampling-phase error in the sampling process of Figure 3.1.3 introduces residual ISI with amplitude that only decays linearly with time. In fact for $q(t) = \text{sinc}(t/T)$, the ISI term $\sum_{k \neq 0} q(\tau + kT)$ with a sampling timing error of $\tau \neq 0$ is not absolutely summable, resulting in infinite peak distortion. The envelope of the time domain response decays more rapidly if the frequency response is smooth (i.e. continuously differentiable). To meet this smoothness condition and also satisfy Nyquist’s Criterion, the response must occupy a larger than minimum bandwidth that is between $1/2T$ and $1/T$. A $q(t)$ with higher bandwidth can exhibit significantly faster decay as $|t|$ increases, thus reducing sensitivity to timing phase errors. Of course, any increase in bandwidth should be as small as possible, while still meeting other system requirements. The percent excess bandwidth\(^7\), or percent roll-off, is a measure of the extra bandwidth.

---

\(^6\)This text distinguishes the Nyquist Frequency from the Nyquist Rate in sampling theory, where the latter is twice the highest frequency of a signal to be sampled and is not the same as the Nyquist Frequency here.

\(^7\)The quantity alpha used here is not a bias factor, and similarly $Q(\cdot)$ is a measure of ISI and not the integral of a unit-variance Gaussian function – uses should be clear to reasonable readers who’ll understand that sometimes symbols are re-used in obviously different contexts.
Definition 3.3.2 (Percent Excess Bandwidth) The percent excess bandwidth \( \alpha \) is determined from a strictly bandlimited \( Q(\omega) \) by finding the highest frequency in \( Q(\omega) \) for which there is nonzero energy transfer. That is
\[
Q(\omega) = \begin{cases} 
\text{nonzero} & |\omega| \leq (1 + \alpha) \frac{\pi}{T} \\
0 & |\omega| > (1 + \alpha) \frac{\pi}{T} 
\end{cases} .
\] (3.72)

Thus, if \( \alpha = 0.15 \) (a typical value), the pulse \( q(t) \) is said to have “15% excess bandwidth.” Usually, data transmission systems have \( 0 \leq \alpha \leq 1 \). In this case in equation (3.69), only the terms \( n = -1, 0, +1 \) contribute to the folded spectrum and the Nyquist Criterion becomes
\[
1 = \frac{1}{T} \left\{ Q(\omega + \frac{2\pi}{T}) + Q(\omega) + Q\left(\omega - \frac{2\pi}{T}\right) \right\} - \frac{T}{\pi} \leq \omega \leq \frac{T}{\pi} .
\] (3.73)

Further, recalling that \( q(t) = \varphi_p(t) \ast \varphi_p^*(-t) \) is Hermitian and has the properties of an autocorrelation function, then \( Q(\omega) \) is real and \( Q(\omega) \geq 0 \). For the region \( 0 \leq \omega \leq \frac{T}{\pi} \) (for real signals), (3.74) reduces to
\[
1 = \frac{1}{T} \left\{ Q(\omega) + Q\left(\omega - \frac{2\pi}{T}\right) \right\} .
\] (3.75)

For complex signals, the negative frequency region \( \left( -\frac{T}{\pi} \leq \omega \leq 0 \right) \) should also have
\[
1 = \frac{1}{T} \left\{ Q(\omega) + Q\left(\omega + \frac{2\pi}{T}\right) \right\} .
\] (3.76)

Figure 3.17: The \( \text{sinc}^2(t/T) \) Function – time-domain

Any \( Q(\omega) \) satisfying (3.76) (and (3.78) in the complex case) is said to be **vestigially symmetric** with respect to the Nyquist Frequency. An example of a vestigially symmetric response with 100% excess bandwidth is \( q(t) = \text{sinc}^2(t/T) \), which is shown in Figures 3.17 and 3.3.1.
Figure 3.18: The $sinc^2(t/T)$ function – frequency-domain
3.3.2 Raised Cosine Pulses

![Raised Cosine Pulse Shapes – Time-Domain](image)

Figure 3.19: Raised cosine pulse shapes – time-domain

The most widely used set of functions that satisfy the Nyquist Criterion are the **raised-cosine** pulse shapes:

**Definition 3.3.3 (Raised-Cosine Pulse Shapes)** The raised cosine family of pulse shapes (indexed by $0 \leq \alpha \leq 1$) is given by

$$q(t) = \operatorname{sinc} \left( \frac{t}{T} \right) \cdot \frac{\cos \left( \frac{\alpha \pi t}{T} \right)}{1 - \left( \frac{2\alpha t}{T} \right)^2},$$

(3.79)

and have Fourier Transforms

$$Q(\omega) = \begin{cases} \frac{T}{\pi} \left[ 1 - \sin \left( \frac{T}{\pi \alpha} (|\omega| - \frac{\pi}{4}) \right) \right] & \frac{T}{\pi} (1 - \alpha) \leq |\omega| \leq \frac{T}{\pi} (1 + \alpha) \\ 0 & \text{otherwise} \end{cases}.$$  

(3.80)
The raised cosine pulse shapes are shown in Figure 3.19 (time-domain) and Figure 3.20 (frequency-domain) for $\alpha = 0, .5$, and 1. When $\alpha = 0$, the raised cosine reduces to a sinc function, which decays asymptotically as $\frac{1}{t}$ for $t \to \infty$, while for $\alpha \neq 0$, the function decays as $\frac{1}{t^3}$ for $t \to \infty$.

### 3.3.2.0.1 Raised Cosine Derivation

Many texts simply provide the form of the raised-cosine function, but for the intrigued student, this text does the inverse transform

$$q_{RCR}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Q_{RCR}(\omega) \cdot e^{j\omega t} \, d\omega .$$

Before inserting the expression for $Q_{RCR}(\omega)$, the inverse transform of this even function simplifies to

$$q_{RCR}(t) = \frac{1}{\pi} \int_{0}^{\infty} Q_{RCR}(\omega) \cdot \cos(\omega t) \, d\omega ,$$

which is separated into 2 integrals over successive positive frequency ranges when the formula from Section 3.3.3 is inserted as

$$q_{RCR}(t) = \frac{1}{\pi} \int_{0}^{(1-\alpha)\pi/T} T \cdot \cos(\omega t) \, d\omega + \frac{1}{2\pi} \int_{(1-\alpha)\pi/T}^{(1+\alpha)\pi/T} \left[ 1 - \sin\left(\frac{T}{2\alpha}(\omega - \pi/T)\right) \right] \cdot \cos(\omega t) \, d\omega .$$

(3.81)

The first integral in Equation (3.81) easily is performed to obtain

$$q_{RCR,1}(t) = (1 - \alpha) \cdot \text{sinc} \left[ \frac{(1 - \alpha)t}{T} \right] ,$$

Equation (3.3.2.0.1) also can be rewritten using $\sin(A - B) = \sin(A) \cdot \cos(B) - \cos(A) \sin(B)$ as

$$q_{RCR,1}(t) = \frac{\sin\left[ \frac{(1-\alpha)\pi t}{T} \right]}{\frac{\pi t}{T}} = \frac{1}{\pi t} \cdot \left[ \sin\left( \frac{\pi t}{T} \right) \cdot \cos\left( \frac{\alpha\pi t}{T} \right) - \cos\left( \frac{\pi t}{T} \right) \cdot \sin\left( \frac{\alpha\pi t}{T} \right) \right]$$

$$= \text{sinc}\left( \frac{\pi t}{T} \right) \cdot \cos\left( \frac{\alpha\pi t}{T} \right) - \frac{\cos\left( \frac{\pi t}{T} \right)}{\frac{\pi t}{T}} \cdot \sin\left( \frac{\alpha\pi t}{T} \right) ,$$

189
which will be a convenient form later in this derivation.

The second integral in Equation (3.81) follows as the sum of two more integrals for the constant term \((2a)\) and the sinusoidal term \((2b)\) as

\[
q_{R\!\!R\!\!C,2a}(t) = -\frac{T}{2\pi t} \left[ \sin \left( \frac{(1 + \alpha)\pi t}{T} \right) - \sin \left( \frac{(1 - \alpha)\pi t}{T} \right) \right]
\]

and

\[
q_{R\!\!R\!\!C,2b}(t) = \cos \left( \frac{\pi t}{T} \right) \cdot \cos \left( \frac{\alpha \pi t}{T} \right)
\]

Using the formula \(\cos(A) \sin(B) = \frac{1}{2} [\sin(A + B) - \sin(A - B)]\), then \(q_{R\!\!R\!\!C,2a}(t)\) becomes:

\[
q_{R\!\!R\!\!C,2a}(t) = \frac{\cos \left( \frac{\pi t}{T} \right)}{\pi t/T} \cdot \sin \left( \frac{\alpha \pi t}{T} \right)
\]

It follows then that

\[
q_{R\!\!R\!\!C,1}(t) + q_{R\!\!R\!\!C,2a}(t) = \frac{\cos \left( \frac{\pi t}{T} \right)}{\pi t/T} \cdot \cos \left( \frac{\alpha \pi t}{T} \right)
\]

\[
q_{R\!\!R\!\!C,2b}(t) = -\frac{T}{2\pi t} \int_{(1-\alpha)\pi/T}^{(1+\alpha)\pi/T} \sin \left[ \frac{T}{2\alpha} \left( \omega - \frac{\pi}{T} \right) \right] \cdot \cos(\omega t) \cdot d\omega
\]

Using the formula \(\sin(A) \cos(B) = \frac{1}{2} [\sin(A + B) + \sin(A - B)]\), then \(q_{R\!\!R\!\!C,2b}(t)\) becomes:

\[
q_{R\!\!R\!\!C,2b}(t) = \frac{T}{4\pi} \left\{ \frac{1}{\frac{T}{2\alpha} + t} \cos \left[ \frac{(1 + \alpha)\pi}{T} \left( \frac{T}{2\alpha} + t \right) - \frac{\pi}{2\alpha} \right] - \frac{1}{\frac{T}{2\alpha} - t} \cos \left[ \frac{(1 - \alpha)\pi}{T} \left( \frac{T}{2\alpha} - t \right) - \frac{\pi}{2\alpha} \right] \right\}
\]

Then, the sum is the RCR so

\[
q_{RCR}(t) = q_{R\!\!R\!\!C,1}(t) + q_{R\!\!R\!\!C,2a}(t) + q_{R\!\!R\!\!C,2b}(t)
\]

\[
= \left[ 1 + \frac{4\alpha^2 t^2}{1 - 4\alpha^2(t/T)^2} \right] \cdot \operatorname{sinc} \left( \frac{t}{T} \right) \cdot \cos \left( \frac{\alpha \pi t}{T} \right)
\]
However, Equation (3.84) seems best to see the limiting value at
\[ t = \phi \]
filtering \( \phi \).

### 3.3.3 Square-Root Splitting of the Nyquist Pulse

The optimum ISI-free transmission system that this section has described has transmit-and-channel filtering \( \varphi_p(t) \) and receiver matched filter \( \varphi_p^*(-t) \) such that \( q(t) = \varphi_p(t) * \varphi_p^*(-t) \) or equivalently

\[
\Phi_p(\omega) = Q^{1/2}(\omega)
\]

so that the matched filter and transmit/channel filters are “square-roots” of the Nyquist pulse. When \( h(t) = \delta(t) \) or equivalently \( \varphi_p(t) = \varphi(t) \), the transmit filter (and the receiver matched filters) are square-roots of a Nyquist pulse shape. Such square-root transmit filtering is often used even when \( \varphi_p(t) \neq \varphi(t) \) and the pulse response is the convolution of \( h(t) \) with the square-root filter.

The raised-cosine pulse shapes are then often used in square-root form, which is

\[
\sqrt{Q(\omega)} = \begin{cases} 
\sqrt{T} & |\omega| \leq \frac{\pi}{4}(1-\alpha) \\
\sqrt{T} \left[ 1 - \sin \left( \frac{\pi}{4T} |\omega| \right) \right]^{1/2} & \frac{\pi}{4}(1-\alpha) \leq |\omega| \leq \frac{\pi}{4}(1+\alpha) \\
\frac{\pi}{4}(1+\alpha) \leq |\omega| 
\end{cases}
\]

This expression can be inverted to the time-domain via use of the identity \( \sin^2(\theta) = .5(1 - \cos(2\theta)) \), as in Problem 3.26, to obtain

\[
\varphi_p(t) = \frac{4\alpha}{\pi \sqrt{T}} \cdot \cos \left[ \frac{1+\alpha}{\pi} \frac{\pi}{4T} t \right] + \frac{4\alpha t \cdot \cos \left( \frac{1+\alpha}{\pi} \frac{\pi}{4T} t \right)}{1 - \left( \frac{4\alpha t}{\pi T} \right)^2} .
\]

Another simpler form for seeing the value at time 0 (which is \( 1/\sqrt{T} \cdot (1 - \alpha + 4\alpha/\pi) \)) is

\[
\varphi_p(t) = \frac{(1-\alpha)}{\sqrt{T}} \cdot \frac{\sqrt{\pi}}{\sqrt{2T}} \cdot \frac{\sin \left( \frac{(1-\alpha)t}{\sqrt{T}} \right)}{\left( 1 - \frac{4\alpha t}{\pi T} \right)^2} + \frac{4\alpha t \cdot \cos \left( \frac{(1+\alpha)}{\pi} \frac{\pi}{4T} t \right)}{\sqrt{T} \pi t \left( \frac{4\alpha t}{\pi T} \right)^2 - 1} .
\]

However, Equation (3.84) seems best to see the limiting value at \( t = \pm \frac{T}{4\alpha} \), where the denominator is readily seen as zero. However, the numerator is also zero if one recalls that \( \cos(x) = \sin(\pi/2 - x) \). The value at these two times (through use of L’Hôpital’s rule) is

\[
\varphi_{SR}(\pm \frac{T}{4\alpha}) = \frac{\alpha}{2\sqrt{T}} \left[ \left( 1 + \frac{2}{\pi} \right) \sin \left( \frac{\pi}{4\alpha} \right) + \left( 1 - \frac{2}{\pi} \right) \cos \left( \frac{\pi}{4\alpha} \right) \right] .
\]

The reader is reminded that the square-root of a Nyquist-satisfying pulse shape does not need to be Nyquist by itself, only in combination with its matched square-root filter. Thus, this pulse is not zero at integer sampling instants.

### 3.3.3.0.1 Square-Root Raised Cosine Derivation

The square-root of a raised-cosine function has inverse transform

\[
\varphi_{SR}(t) \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} \sqrt{Q_{RCR}(\omega)} \cdot e^{j\omega t} d\omega .
\]

Before inserting the expression for \( Q_{RCR}(\omega) \), the inverse transform of this even function simplifies to

\[
\varphi_{SR}(t) \triangleq \frac{1}{\pi} \int_{0}^{\infty} \sqrt{Q_{RCR}(\omega)} \cdot \cos(\omega t) \, d\omega ,
\]

which is the desired result.
which is separated into 2 integrals over successive positive frequency ranges when the formula from Section 3.3.3 is inserted as

\[
\varphi_{SR}(t) = \frac{1}{\pi} \int_{0}^{(1-\alpha)\pi/T} \sqrt{T} \cdot \cos(\omega t) \, d\omega + \frac{1}{\pi} \int_{(1-\alpha)\pi/T}^{(1+\alpha)\pi/T} \sqrt{T/2} \cdot \left[ 1 - \sin \left( \frac{T}{2\alpha} (\omega - \pi/T) \right) \right] \cdot \cos(\omega t) \, d\omega .
\]

(3.87)

The first integral in Equation (3.87) easily is performed to obtain

\[
\varphi_{SR,1}(t) = \frac{(1-\alpha)}{\sqrt{T}} \cdot \sin \left( \frac{(1-\alpha)t}{T} \right) .
\]

(3.88)

Using the half-angle formula \(1 - \sin(x) = 1 - \cos(x - \pi/2) = 2\sin^2 \left( \frac{\xi}{2} - \frac{x}{2} \right)\), the 2nd integral in Equation (3.87) expands into integrable terms as

\[
\varphi_{SR,2}(t) = \frac{\sqrt{T}}{\pi} \int_{(1-\alpha)\pi/T}^{(1+\alpha)\pi/T} \sin \left( \frac{T}{4\alpha} \left[ \omega - \frac{(1+\alpha)\pi}{T} \right] + \omega t \right) + \sin \left( \frac{T}{4\alpha} \left[ \omega - \frac{(1-\alpha)\pi}{T} \right] - \omega t \right) \cdot \cos(\omega t) \, d\omega
\]

= \frac{\sqrt{T}}{\pi} \int_{(1-\alpha)\pi/T}^{(1+\alpha)\pi/T} \sin \left[ \frac{\omega(t + \frac{T}{4\alpha}) - \pi(1+\alpha)}{4\alpha} \right] + \sin \left[ \frac{\omega(t - \frac{T}{4\alpha}) - \pi(1-\alpha)}{4\alpha} \right] \cdot \cos(\omega t) \, d\omega
\]

= \sqrt{T} \cdot \frac{-4\alpha}{2\pi(4\alpha t + T)} \cdot \cos \left[ \frac{\omega(t + \frac{T}{4\alpha}) - \pi(1+\alpha)}{4\alpha} \right] + \sqrt{T} \cdot \frac{-4\alpha}{2\pi(T - 4\alpha)} \cdot \cos \left[ \frac{\omega(t - \frac{T}{4\alpha}) - \pi(1-\alpha)}{4\alpha} \right]

\]

Continuing with algebra at this point:

\[
\varphi_{SR,2}(t) = \frac{-2\alpha \sqrt{T}}{\pi(4\alpha t + T)} \left\{ \cos \left( \frac{(1+\alpha)\pi t}{T} \right) + \sin \left( \frac{(1-\alpha)\pi t}{T} \right) \right\} + \frac{2\alpha \sqrt{T}}{\pi(T - 4\alpha)} \left\{ \cos \left( \frac{(1+\alpha)\pi t}{T} \right) - \sin \left( \frac{(1-\alpha)\pi t}{T} \right) \right\}
\]

= \frac{2\alpha \sqrt{T}}{\pi} \left[ \frac{-1}{4\alpha t + T} + \frac{1}{T - 4\alpha} \right] \cos \left( \frac{(1+\alpha)\pi t}{T} \right)

\]

Continuing with algebra at this point:

\[
\varphi_{SR,2}(t) = \frac{2\alpha \sqrt{T}}{\pi} \left[ \frac{-2T}{(4\alpha t)^2 - T^2} \right] \cos \left( \frac{(1+\alpha)\pi t}{T} \right)
\]

192
\[
\begin{align*}
\varphi_{SR}(t) &= \frac{1}{\sqrt{T}} \cdot \sin \left( \frac{(1 - \alpha)t}{T} \right) + \frac{4\alpha \cdot \cos \left( \frac{(1 + \alpha)t}{T} \right)}{\sqrt{T} \pi t \left( \frac{4\alpha t}{T} \right)^2 - 1} + \frac{16\alpha^2 t^2}{T^2} \cdot \frac{T \cdot \sin \left( \frac{(1 - \alpha)t}{T} \right)}{\left( \frac{4\alpha t}{T} \right)^2 - 1} \\
&= \frac{2\alpha \sqrt{T}}{\pi} \left[ \frac{8\alpha t}{(4\alpha t)^2 - T^2} \right] \sin \left( \frac{(1 - \alpha)t}{T} \right) \\
&= \frac{4\alpha \cdot \cos \left( \frac{(1 + \alpha)t}{T} \right)}{\sqrt{T} \pi t \left( \frac{4\alpha t}{T} \right)^2 - 1} + \frac{16\alpha^2 t^2}{T^2} \cdot \frac{T \cdot \sin \left( \frac{(1 - \alpha)t}{T} \right)}{\left( \frac{4\alpha t}{T} \right)^2 - 1} + \frac{16\alpha^2 t^2}{T^2} \cdot \frac{T \cdot \sin \left( \frac{(1 - \alpha)t}{T} \right)}{\left( \frac{4\alpha t}{T} \right)^2 - 1} \\
&= \frac{4\alpha \cdot \cos \left( \frac{(1 + \alpha)t}{T} \right)}{\sqrt{T} \pi t \left( \frac{4\alpha t}{T} \right)^2 - 1} + \frac{16\alpha^2 t^2}{T^2} \cdot \frac{T \cdot \sin \left( \frac{(1 - \alpha)t}{T} \right)}{\left( \frac{4\alpha t}{T} \right)^2 - 1}.
\end{align*}
\]

The first term from Equation (3.87) now is added so that

\[
\varphi_{SR}(t) = \frac{(1 - \alpha)}{\sqrt{T}} \cdot \sin \left[ \frac{(1 - \alpha)t}{T} \right] + \frac{4\alpha \cdot \cos \left( \frac{(1 + \alpha)t}{T} \right)}{\sqrt{T} \pi t \left( \frac{4\alpha t}{T} \right)^2 - 1} + \frac{16\alpha^2 t^2}{T^2} \cdot \frac{T \cdot \sin \left( \frac{(1 - \alpha)t}{T} \right)}{\left( \frac{4\alpha t}{T} \right)^2 - 1}.
\]

Combining the sinc terms provides one convenient form of the square-root raised cosine of

\[
\varphi_{SR}(t) = \frac{(1 - \alpha)}{\sqrt{T}} \cdot \text{sinc} \left[ \frac{(1 - \alpha)t}{T} \right] + \frac{4\alpha \cdot \cos \left( \frac{(1 + \alpha)t}{T} \right)}{\sqrt{T} \pi t \left( \frac{4\alpha t}{T} \right)^2 - 1} + \frac{16\alpha^2 t^2}{T^2} \cdot \frac{T \cdot \sin \left( \frac{(1 - \alpha)t}{T} \right)}{\left( \frac{4\alpha t}{T} \right)^2 - 1},
\]

from which limiting results like \( \varphi_{SR}(0) = 1/sqrt{T} \) follow easily. The preferred form for most engineering texts follows by combining the two fractional terms and simplifying.

\[
\varphi_{SR}(t) = \varphi_p(t) = \frac{4\alpha}{\pi \sqrt{T}} \cdot \cos \left( \frac{(1 + \alpha)^2}{4\alpha t} \right) + \frac{T \cdot \sin \left[ \frac{(1 - \alpha)^2}{4\alpha t} \right]}{4\alpha t} \cdot \frac{1 - \left( \frac{4\alpha t}{T} \right)^2}{1 - \left( \frac{4\alpha t}{T} \right)^2}.
\]
3.4 Linear Zero-Forcing Equalization

This section examines the Zero-Forcing Equalizer (ZFE), which is the easiest type of equalizer to analyze and understand, but has inferior performance to some other equalizers to be introduced in later sections. The ISI model in Figure 3.1.3 is used to describe the zero-forcing version of the discrete-time receiver. The ZFE is often a first non-trivial attempt at a receiver in Figure 3.2 of Section 3.2.

The ZFE sets $R$ equal to a linear time-invariant filter with discrete impulse response $w_k$ that acts on $y_k$ to produce $z_k$, which is an estimate of $x_k$. Ideally, for symbol-by-symbol detection to be optimal, $q_k = \delta_k$ by Nyquist’s Criterion. The equalizer tries to restore this Nyquist Pulse character to the channel. In so doing, the ZFE ignores the noise and shapes the signal $y_k$ so that it is free of ISI. From the ISI-channel model of Section 3.1 and Figure 3.1.3,

$$y_k = \|p\| \cdot x_k * q_k + n_k .$$

In the ZFE case, $n_k$ is initially viewed as being zero. The $D$-transform of $y_k$ is (See Appendix A.2)

$$Y(D) = \|p\| \cdot X(D) \cdot Q(D) .$$

The ZFE output, $z_k$, has Transform

$$Z(D) = W(D) \cdot Y(D) = W(D) \cdot Q(D) \cdot \|p\| \cdot X(D) ,$$

and will be free of ISI if $Z(D) = X(D)$, leaving the ZFE filter characteristic:

**Definition 3.4.1 (Zero-Forcing Equalizer)** The ZFE transfer characteristic is

$$W(D) = \frac{1}{Q(D) \cdot \|p\|} .$$

This discrete-time filter processes the discrete-time sequence corresponding to the matched-filter output. The ZFE is so named because ISI is “forced to zero” at all sampling instants $kT$ except $k = 0$. The receiver uses symbol-by-symbol detection, based on decision regions for the constellation defined by $x_k$, on the output of the ZFE.

3.4.1 Performance Analysis of the ZFE

The variance of the noise, which is not zero in practice, at the output of the ZFE is important in determining performance, even if ignored in the ZFE design of Equation (3.92). As this noise is produced by a linear filter acting on the discrete Gaussian noise process $n_k$, it is also Gaussian. The designer can compute the discrete autocorrelation function (the bar denotes normalized to one dimension, so $N = 2$ for the complex QAM case) for the noise $n_k$ as

$$\bar{r}_{nn,k} = E [n_l n_{l-k}^*] / N$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E [n_l(t) n_{l-k}^*(s)] \varphi_p^*(t - lT) \varphi_p(s - (l-k)T) dt ds$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{N_0}{2} \delta(t - s) \varphi_p^*(t - lT) \varphi_p(s - (l-k)T) dt ds$$

$$= \frac{N_0}{2} \int_{-\infty}^{\infty} \varphi_p^*(t - lT) \varphi_p(t - (l-k)T) dt$$

$$= \frac{N_0}{2} \int_{-\infty}^{\infty} \varphi_p^*(u) \varphi_p(u + kT) du \quad \text{(letting } u = t - lT)$$

$$= \frac{N_0}{2} q_{-k} = \frac{N_0}{2} q_k ,$$
or more simply
\[ \tilde{R}_{nn}(D) = \frac{N_0}{2} Q(D), \]  
(3.99)

where the analysis uses the substitution \( u = t - lT \) in going from (3.96) to (3.97), and assumes baseband equivalent signals for the complex case. The complex baseband-equivalent noise, has (one-dimensional) sample variance \( \frac{N_0}{2} \) at the normalized matched filter output. The noise at the output of the ZFE, \( n_{ZFE,k} \), has autocorrelation function
\[ \tilde{r}_{ZFE,k} = \tilde{r}_{nn,k} \ast w_k \ast w_k^\ast. \]  
(3.100)

The \( D \)-Transform of \( \tilde{r}_{ZFE,k} \) is then
\[ \tilde{R}_{ZFE}(D) = \frac{N_0}{2 \|p\|^2} \cdot Q(D) = \frac{\frac{N_0}{2}}{\|p\|^2} \cdot Q(D). \]  
(3.101)

The power spectral density of the noise samples \( n_{ZFE,k} \) is then \( \tilde{R}_{ZFE}(e^{-j\omega T}) \). The (per-dimensional) variance of the noise samples at the output of the ZFE is the (per-dimensional) mean-square error between the desired \( x_k \) and the ZFE output \( z_k \). Since
\[ z_k = x_k + n_{ZFE,k}, \]  
(3.102)

then \( \sigma_{ZFE}^2 \) is computed as
\[ \sigma_{ZFE}^2 = \frac{T}{2\pi} \int_{\pi T}^{\pi} \tilde{R}_{ZFE}(e^{-j\omega T}) d\omega = \frac{T}{2\pi} \int_{\pi T}^{-\pi T} \frac{N_0}{\|p\|^2} \cdot Q(e^{-j\omega T}) d\omega = \frac{N_0}{\|p\|^2} \gamma_{ZFE} = \frac{N_0}{\|p\|^2} w_0, \]  
(3.103)

where
\[ \gamma_{ZFE} = \frac{T}{2\pi} \int_{\pi T}^{\pi} \frac{1}{Q(e^{-j\omega T})} d\omega = w_0 \cdot \|p\|. \]  
(3.104)

The center tap of a linear equalizer is \( w_0 \), or if the ZFE is explicitly used, then \( w_{ZFE,0} \). This tap is easily shown to always have the largest magnitude for a ZFE and thus spotted readily when plotting the impulse response of any linear equalizer. From (3.102) and the fact that \( E[n_{ZFE,k}] = 0 \), \( E[z_k/x_k] = x_k \), so that the ZFE is unbiased for a detection rule based on \( x_k \). The SNR at the output of the ZFE is
\[ \text{SNR}_{ZFE} = \frac{\tilde{\xi}_x}{\sigma_{ZFE}^2} = \text{SNR}_{MFB} \cdot \frac{1}{\gamma_{ZFE}}. \]  
(3.105)

Computation of \( d_{\text{min}} \) over the signal set corresponding to \( x_k \) leads to a relation between \( \tilde{\xi}_x, d_{\text{min}}, \) and \( M \), and the NNUB on error probability for a symbol-by-symbol detector at the ZFE output is (please, do not confuse the \( Q \) function with the channel autocorrelation \( Q(D) \))
\[ P_{ZFE,e} \approx N_e \cdot Q \left( \frac{d_{\text{min}}}{2\sigma_{ZFE}} \right). \]  
(3.106)

Since symbol-by-symbol detection can never have performance that exceeds the MFB,
\[ \sigma^2 \leq \sigma_{ZFE}^2 \|p\|^2 = \frac{\sigma^2}{2\pi} \int_{\pi T}^{\pi T} \frac{d\omega}{Q(e^{-j\omega T})} = \sigma^2 \cdot \gamma_{ZFE} \]  
(3.107)

\[ \text{SNR}_{MFB} \geq \text{SNR}_{ZFE} \]  
(3.108)

\[ \gamma_{ZFE} \geq 1 \]  
(3.109)

with equality if and only if \( Q(D) = 1 \).

Finally, the ZFE equalization loss, \( 1/\gamma_{ZFE} \), defines the SNR reduction from the MFB:
Definition 3.4.2 (ZFE Equalization Loss) The ZFE Equalization Loss, $\gamma_{ZFE}$ in decibels is

$$\gamma_{ZFE} \triangleq 10 \cdot \log_{10} \left( \frac{SNR_{MFB}}{SNR_{ZFE}} \right) = 10 \log_{10} \frac{\|p\|^2 \cdot \sigma^2_{ZFE}}{\sigma^2} = 10 \log_{10} \|p\| \cdot w_{ZFE,0} \ , \ (3.110)$$

and is a measure of the loss (in dB) with respect to the MFB for the ZFE. The sign of the loss is often ignored since it is always a negative (or 0) number and only its magnitude is of concern.

Equation (3.110) always thus provides a non-negative number.

### 3.4.2 Noise Enhancement

The design of $W(D)$ for the ZFE ignored the effects of noise. This oversight can lead to noise enhancement, a $\sigma^2_{ZFE}$ that is unacceptably large, and the consequent poor performance.

Figure 3.21 hypothetically illustrates an example of a lowpass channel with a notch at the Nyquist Frequency, that is, $Q(e^{-j\omega T})$ is zero at $\omega = \frac{\pi}{T}$. Since $W(e^{-j\omega T}) = 1/(\|p\| \cdot Q(e^{-j\omega T}))$, then any noise energy near the Nyquist Frequency is enhanced (increased in power or energy), in this case so much that $\sigma^2_{ZFE} \to \infty$. Even when there is no channel notch at any frequency, $\sigma^2_{ZFE}$ can be finite, but large, leading to unacceptable performance degradation.

In actual computation of examples, the author has found that the table in Table 3.1 is useful in recalling basic digital signal processing equivalences related to scale factors between continuous time convolution and discrete-time convolution.

The following example clearly illustrates the noise-enhancement effect:
Table 3.1: Conversion factors of $T$.

<table>
<thead>
<tr>
<th>Time $p(t)$ ↔ $P(\omega)$</th>
<th>Frequency $P_k = p(kT)$ ↔ $P(e^{-j\omega T}) = \frac{1}{T} \sum_{n=-\infty}^{\infty} P\left(\omega + \frac{2\pi m}{T}\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y(t) = x(t) \ast p(t)$ ↔ $Y(\omega) = X(\omega) \cdot P(\omega)$</td>
<td>$y'(kT) = T \cdot x_k \ast p_k$ ↔ $Y'(e^{-j\omega T}) = T \cdot X(e^{-j\omega T}) \cdot P(e^{-j\omega T})$</td>
</tr>
<tr>
<td>$|p|^2 = \int_{-\infty}^{\infty}</td>
<td>p(t)</td>
</tr>
</tbody>
</table>

| continuous-time | discrete-time |
| sampling theorem satisfied | (sampling thm satisfied) |
EXAMPLE 3.4.1 (1 + .9D⁻¹ - ZFE) Suppose the pulse response of a (strictly bandlimited) channel used with binary PAM is

\[ P(\omega) = \begin{cases} \sqrt{T} (1 + .9e^{j\omega T}) & |\omega| \leq \frac{\pi}{T} \\ 0 & |\omega| > \frac{\pi}{T} \end{cases} \quad (3.111) \]

Then \( ||p||^2 \) is computed as

\[
||p||^2 = \frac{1}{2\pi} \int_{-\frac{T}{2}}^{\frac{T}{2}} |P(\omega)|^2 d\omega \quad (3.112)
\]

\[
= \frac{1}{2\pi} \int_{-\frac{T}{2}}^{\frac{T}{2}} T [1.81 + 1.8 \cos(\omega T)] d\omega \quad (3.113)
\]

\[
= \frac{T}{2\pi} 1.81 \frac{2\pi}{T} = 1.81 = 1^2 + .9^2 \quad (3.114)
\]

The magnitude of the Fourier transform of the pulse response appears in Figure 3.22. Thus, with \( \Phi_p(\omega) \) as the Fourier transform of \( \{\phi_p(t)\} \),

\[
\Phi_p(\omega) = \begin{cases} \sqrt{T/1.81} (1 + .9e^{j\omega T}) & |\omega| \leq \frac{\pi}{T} \\ 0 & |\omega| > \frac{\pi}{T} \end{cases} \quad (3.115)
\]
Then, (using the bandlimited property of $P(\omega)$ in this example)

$$Q(e^{-\omega T}) = \frac{1}{T} |\Phi_p(\omega)|^2$$

$$= \frac{1}{1.81} |1 + .9e^{\omega T}|^2$$

$$= \frac{1.81 + 1.8 \cos(\omega T)}{1.81}$$

$$Q(D) = \frac{.9D^{-1} + 1.81 + .9D}{1.81}.$$  \hspace{1cm} (3.116)

$$= .9 + 1.81 + .9D + 9D^2.$$

Then

$$W(D) = \sqrt{\frac{1.81D}{.9 + 1.81 + .9D}}.$$  \hspace{1cm} (3.117)

The magnitude of the Fourier transform of the ZFE response appears in Figure 3.23.

![Figure 3.23: ZFE magnitude for example.](image)

The time-domain sample response of the equalizer is shown in Figure 3.24.
Figure 3.24: ZFE time-domain response.

Computation of $\sigma^2_{ZFE}$ allows performance evaluation, \(^8\)

\[
\sigma^2_{ZFE} = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{N_0}{2} Q(e^{-j\omega T}) \cdot ||p||^2 d\omega
\]

(3.121)

\[
= \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{N_0}{2} (1.81/1.81) \cdot \frac{N_0}{2} d\omega
\]

(3.122)

\[
= \frac{N_0}{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{1.81 + 1.8 \cos(\omega T)} du
\]

(3.123)

\[
= \left( \frac{N_0}{2} \right) \frac{2}{2\pi} \left[ \frac{2}{\sqrt{1.81^2 - 1.8^2}} \tan^{-1} \frac{1.81^2 - 1.8^2 \tan \frac{u}{2}}{1.81 + 1.8} \right]_0^\pi
\]

(3.124)

\[
= \left( \frac{N_0}{2} \right) \frac{4}{2\pi\sqrt{1.81^2 - 1.8^2}} \cdot \frac{\pi}{2}
\]

(3.125)

\[
= \left( \frac{N_0}{2} \right) \frac{1}{\sqrt{1.81^2 - 1.8^2}}
\]

(3.126)

\[
= \frac{N_0}{2} (5.26)
\]

(3.127)

\(^8\)From a table of integrals

\[
\int_{\frac{1}{a+b \cos u}}^1 du = \frac{2}{\sqrt{a^2 - b^2}} \tan^{-1} \left( \frac{\sqrt{a^2 - b^2} \tan(\frac{u}{2})}{a+b} \right).
\]
The ZFE-output SNR is
\[ \text{SNR}_{ZFE} = \frac{\tilde{\mathcal{E}}_x}{5.26 N_0 T}. \] (3.128)
This SNR\textsubscript{ZFE} is also the argument of the Q-function for a binary detector at the ZFE output. Assuming the two transmitted levels are \( x_k = \pm 1 \), then we get
\[ \text{MFB} = \frac{\|p\|^2}{\sigma^2} = \frac{1.81}{\sigma^2}, \] (3.129)
leaving
\[ \gamma_{ZFE} = 10 \log_{10}(1.81 \cdot 5.26) \approx 9.8 \text{dB} \] (3.130)
for any value of the noise variance. This is very poor performance on this channel. No matter what \( b \) is used, the loss is almost 10 dB from best performance. Thus, noise enhancement can be a serious problem. Chapter 9 demonstrates a means by which to ensure that there is no loss with respect to the matched filter bound for this channel. With alteration of the signal constellation, Chapter 10 describes a means that ensures an error rate of \( 10^{-5} \) on this channel, with no information rate loss, which is far below the error rate achievable even when the MFB is attained. Thus, there are good solutions, but the simple concept of a ZFE is not a good solution.

Another example illustrates the generalization of the above procedure when the pulse response is complex (corresponding to a QAM channel):

**EXAMPLE 3.4.2 (QAM: \(-.5 D^{-1} + (1 + .25j) - .5j D\text{ Channel})** Given a baseband equivalent channel
\[ p(t) = \frac{1}{\sqrt{T}} \left\{ -\frac{1}{2} \text{sinc} \left( \frac{t + T}{T} \right) + \left( 1 + \frac{j}{4} \right) \text{sinc} \left( \frac{t}{T} \right) - \frac{j}{2} \text{sinc} \left( \frac{t - T}{T} \right) \right\}, \] (3.131)
the discrete-time channel samples are
\[ p_k = \frac{1}{\sqrt{T}} \left[ -\frac{1}{2} \left( 1 + \frac{j}{4} \right), -\frac{j}{2} \right]. \] (3.132)
This channel has the transfer function of Figure 3.25.
The pulse response norm (squared) is

\[
\|p\|^2 = \frac{T}{T} (.25 + 1 + .0625 + .25) = 1.5625.
\]  

(3.133)

Then \(q_k\) is given by

\[
q_k = q(kT) = T (\varphi_{p,k} \ast \varphi_{p,-k}^*) = \frac{1}{1.5625} \left[ -\frac{j}{4}, \frac{5}{8}(-1+j), 1.5625, -\frac{5}{8}(1+j), \frac{j}{4} \right]
\]  

or

\[
Q(D) = \frac{1}{1.5625} \left[ -\frac{j}{4}D^{-2} + \frac{5}{8}(-1+j)D^{-1} + 1.5625 - \frac{5}{8}(1+j)D + \frac{j}{4}D^2 \right]
\]  

(3.134)

(3.135)

Then, \(Q(D)\) factors into

\[
Q(D) = \frac{1}{1.5625} \left\{ (1 - .5jD) (1 - .5D^{-1}) (1 + .5jD^{-1}) (1 - .5D) \right\}
\]  

(3.136)

and

\[
W_{ZFE}(D) = \frac{1}{Q(D)\|p\|}
\]  

(3.137)
or

\[ W_{ZFE}(D) = \frac{\sqrt{1.5625}}{(-.25jD^{-2} + .625(-1 + j)D^{-1} + 1.5625 - .625(1 + j)D + .25jD^2)}. \quad (3.138) \]

The Fourier transform of the ZFE is in Figure 3.26.

![Fourier transform magnitude of the ZFE for the complex baseband channel example.](image)

Figure 3.26: Fourier transform magnitude of the ZFE for the complex baseband channel example.

The real and imaginary parts of the equalizer response in the time domain are shown in Figure 3.27.
Then
\[ \sigma_{ZFE}^2 = \frac{T}{2\pi} \int_{\pi/2}^{\pi} \frac{N_0}{\|p\|^2} Q(e^{-j\omega T}) d\omega , \]  
(3.139)

or
\[ \sigma_{ZFE}^2 = \frac{N_0}{2} \frac{w_0}{\|p\|} , \]  
(3.140)

where \( w_0 \) is the zero (center) coefficient of the ZFE. This coefficient can be extracted from
\[
W_{ZFE}(D) = \sqrt{1.5625} \left[ \frac{D^2(-4j)}{(D+2j)(D-2)(D-.5)(D+.5j)} \right] \]
\[
= \sqrt{1.5625} \left[ \frac{A}{D-2} + \frac{B}{D+2j} + \text{terms not contributing to } w_0 \right] ,
\]

where \( A \) and \( B \) are coefficients in the partial-fraction expansion (the residuez and residue commands in Matlab can be very useful here):
\[
A = \frac{4(-4j)}{(2+2j)(2+.5j)(1.5)} = -1.5686 - j.9412 = 1.8293\angle -2.601 \]  
(3.141)
\[
B = \frac{-4(-4j)}{(-2j-.5)(-2j-2)(-1.5j)} = .9412 + j1.5685 = 1.8293\angle 1.030 . \]  
(3.142)

Finally
\[
W_{ZFE}(D) = \sqrt{1.5625} \left[ \frac{A(-.5)}{1-.5D} + \frac{B(-.5j)}{1-.5jD} + \text{terms} \right] ,
\]
(3.143)

and
\[
w_0 = \sqrt{1.5625} \left[ \left( -1.5686 - j.9412 \right) - .5j(\cdot9412 + j1.5686) \right] \]  
(3.144)
\[
= \sqrt{1.5625}(1.57) = 1.96 \]  
(3.145)

The ZFE loss can be shown to be
\[
\gamma_{ZFE} = 10 \cdot \log_{10}(w_0 \cdot \|p\|) = 3.9 \text{ dB} ,
\]  
(3.146)

which is better than the last channel because the frequency spectrum is not as near zero in this complex example as it was earlier on the PAM example. Nevertheless, considerably better performance is also possible on this complex channel, but not with the ZFE.
To compute $P_e$ for 4QAM, the designer calculates

$$P_e \approx Q\left(\sqrt{\text{SNR}_{MFB} - 3.9 \text{ dB}}\right).$$ \hspace{1cm} (3.147)

If $\text{SNR}_{MFB} = 10$ dB, then $P_e \approx 2.2 \times 10^{-2}$. If $\text{SNR}_{MFB} = 17.4$ dB, then $P_e \approx 1.0 \times 10^{-6}$. If $\text{SNR}_{MFB} = 23.4$ dB for 16QAM, then $P_e \approx 2.0 \times 10^{-5}$. 
3.5 Minimum Mean-Square Error Linear Equalization

The Minimum Mean-Square Error Linear Equalizer (MMSE-LE) balances a reduction in ISI with noise enhancement. The MMSE-LE always performs as well as, or better than, the ZFE and is of the same complexity of implementation. Nevertheless, it is slightly more complicated to describe and analyze than is the ZFE. The MMSE-LE uses a linear time-invariant filter $w_k$ for $R$, but the choice of filter impulse response $w_k$ is different than the ZFE.

The MMSE-LE is a linear filter $w_k$ that acts on $y_k$ to form an output sequence $z_k$ that is the best MMSE estimate of $x_k$. That is, the filter $w_k$ minimizes the Mean Square Error (MSE):

**Definition 3.5.1 (Mean Square Error (for the LE))** The LE error signal is given by

$$e_k = x_k - w_k * y_k = x_k - z_k$$ \hspace{1cm} (3.148)

The Minimum Mean Square Error (MMSE) for the linear equalizer is defined by

$$\sigma_{MMSE-LE}^2 \overset{\Delta}{=} \min_{w_k} E[|x_k - z_k|^2]$$ \hspace{1cm} (3.149)

The MSE criteria for filter design does not ignore noise enhancement because the optimization of this filter compromises between eliminating ISI and increasing noise power. Instead, the filter output is as close as possible, in the Minimum MSE sense, to the data symbol $x_k$.

### 3.5.1 Optimization of the Linear Equalizer

Using $D$-transforms,

$$E(D) = X(D) - W(D) \cdot Y(D)$$ \hspace{1cm} (3.150)

By the orthogonality principle in Appendix A, at any time $k$, the error sample $e_k$ must be uncorrelated with any equalizer input signal $y_m$. Succinctly,

$$E\left[E(D)Y^*(D^{-1})\right] = 0$$ \hspace{1cm} (3.151)

Evaluating (3.151), using (3.150), yields

$$0 = \bar{R}_{xy}(D) - W(D) \cdot \bar{R}_{yy}(D)$$ \hspace{1cm} (3.152)

where ($N = 1$ for PAM, $N = 2$ for Quadrature Modulation)$^9$

$$\bar{R}_{xy}(D) = E\left[X(D) \cdot Y^*(D^{-1})\right] / N = \|p\|Q(D)\tilde{e}_x$$

$$\bar{R}_{yy}(D) = E\left[Y(D) \cdot Y^*(D^{-1})\right] / N = \|p\|^2Q^2(D)\tilde{e}_x + \frac{N_0}{2}Q(D) = Q(D)\left(\|p\|^2 \cdot Q(D) \cdot \tilde{e}_x + \frac{N_0}{2}\right)$$

Then the MMSE-LE becomes

$$W(D) = \frac{\bar{R}_{xy}(D)}{\bar{R}_{yy}(D)} = \frac{1}{\|p\| \cdot (Q(D) + 1/\text{SNR}_{MFB})}$$ \hspace{1cm} (3.153)

The MMSE-LE differs from the ZFE only in the additive positive term in the denominator of (3.153). The transfer function for the equalizer $W(e^{-j\omega T})$ is also real and positive for all finite signal-to-noise ratios. This small positive term prevents the denominator from ever becoming zero, and thus makes the MMSE-LE well defined even when the channel (or pulse response) is zero for some frequencies or frequency bands. Also $W(D) = W^*(D^{-1})$.

---

$^9$The expression $R_{xy}(D) \overset{\Delta}{=} E\left[X(D)X^*(D^{-1})\right]$ is used in a symbolic sense, since the terms of $X(D)X^*(D^{-1})$ are of the form $\sum_k x_kx_{k-j}^*$, so that we are implying the additional operation $\lim_{K \to \infty} |1/(2K + 1)|\sum_{-K\leq k \leq K}$ on the sum in such terms. This is permissible for stationary (and ergodic) discrete-time sequences.
Figure 3.28: Example of MMSE-LE versus ZFE.

Figure 3.5.1 repeats Figure 3.4.2 with addition of the MMSE-LE transfer characteristic. The MMSE-LE transfer function has magnitude $\frac{E_x}{\sigma^2}$ at $\omega = \frac{\pi}{T}$, while the ZFE becomes infinite at this same frequency. This MMSE-LE leads to better performance, as the next subsection computes.

3.5.2 Performance of the MMSE-LE

The MMSE is the time-0 coefficient of the error autocorrelation sequence

$$\tilde{R}_{ee}(D) = E [E(D)E^*(D^{-*})] / N$$

$$= \bar{e}_x - W^*(D^{-*})\bar{R}_{yy}(D) - W(D)\bar{R}_{zy}(D^{-*}) + W(D)\bar{R}_{yy}(D)W^*(D^{-*})$$

The third equality follows from

$$W(D)\bar{R}_{yy}(D)W^*(D^{-*}) = W(D)\bar{R}_{yy}(D)\bar{R}_{yy}(D)^{-1}\bar{R}_{zy}(D^{-*})$$

$$= W(D)\bar{R}_{zy}(D^{-*})$$

and that $(W(D)\bar{R}_{yy}(D)W^*(D^{-*}))^* = W(D)\bar{R}_{yy}(D)W^*(D^{-*})$. The MMSE then becomes

$$\sigma_{MMSE-LE}^2 = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \tilde{R}_{ee}(e^{-j\omega T})d\omega = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{\bar{e}_x Q(D) + \frac{N_0}{2}}{||p||^2 (Q(D) + 1/\text{SNR}_{MFB})} d\omega$$

$$= \frac{N_0}{2} / ||p||$$

By recognizing that $W(e^{-j\omega T})$ is multiplied by the constant $\frac{N_0}{2} / ||p||$ in (3.160), then

$$\sigma_{MMSE-LE}^2 = \frac{N_0}{2} / ||p||$$
From comparison of (3.160) and (3.103), that

$$\sigma^2_{MMSE-LE} \leq \sigma^2_{ZFE},$$  \hspace{1cm} (3.162)

with equality if and only if $\text{SNR}_{MFB} \to \infty$. Furthermore, since the ZFE is unbiased, $\text{SNR}_{MMSE-LE} = \text{SNR}_{MMSE-LE,U} + 1$ and $\text{SNR}_{MMSE-LE,U}$ are the maximum-SNR corresponding to unconstrained and unbiased linear equalizers, respectively,

$$\text{SNR}_{ZFE} \leq \text{SNR}_{MMSE-LE,U} = \frac{\bar{\mathbb{E}}_x}{\sigma^2_{MMSE-LE}} - 1 \leq \text{SNR}_{MFB}. \hspace{1cm} (3.163)$$

One confirms that the MMSE-LE is a biased receiver by writing the following expression for the equalizer output

$$Z(D) = W(D)Y(D) \hspace{1cm} (3.164)$$

$$= \frac{1}{\|p\|}(Q(D) + \frac{1}{\text{SNR}_{MFB}}) \left(\frac{Q(D)}{\|p\|}X(D) + N(D)\right) \hspace{1cm} (3.165)$$

$$= X(D) - \frac{1}{\text{SNR}_{MFB}} \left(\frac{Q(D)}{\|p\|}X(D) + \frac{N(D)}{\|p\|}N(D)\right), \hspace{1cm} (3.166)$$

for which the $x_k$-dependent residual ISI term contains a component

$$\text{signal-basis term} = -\frac{1}{\text{SNR}_{MFB}} \cdot \|p\| \cdot x_k \hspace{1cm} (3.167)$$

$$= -\frac{1}{\text{SNR}_{MFB}} \cdot \frac{\sigma^2_{MMSE-LE} \cdot \|p\|^2}{\bar{\mathbb{E}}_x} \cdot x_k \hspace{1cm} (3.168)$$

$$= -\frac{\bar{\mathbb{E}}_x}{\text{SNR}_{MSE-LE}} \cdot x_k \hspace{1cm} (3.169)$$

$$= -\frac{1}{\text{SNR}_{MSE-LE}} \cdot x_k. \hspace{1cm} (3.170)$$

So $z_k = \left(1 - \frac{1}{\text{SNR}_{MMSE-LE}}\right)x_k - e'_k$ where $e'_k$ is the error for unbiased detection and $R$. The optimum unbiased receiver with decision regions scaled by $1 - \frac{1}{\text{SNR}_{MMSE-LE}}$ (see Section 3.2.1) has the signal energy given by

$$\left(1 - \frac{1}{\text{SNR}_{MMSE-LE}}\right)^2 \bar{\mathbb{E}}_x. \hspace{1cm} (3.172)$$

A new error for the scaled decision regions is $e'_k = (1 - 1/\text{SNR}_{MMSE-LE})x_k - z_k = e_k - \frac{1}{\text{SNR}_{MMSE-LE}}x_k$, which is also the old error with the $x_k$ dependent term removed. Since $e'_k$ and $x_k$ are then independent, then

$$\sigma^2_e = \sigma^2_{MMSE-LE} = \sigma^2_e' + \left(\frac{1}{\text{SNR}_{MMSE-LE}}\right)^2 \bar{\mathbb{E}}_x \hspace{1cm} (3.173)$$

leaving

$$\sigma^2_e = \sigma^2_{MMSE-LE} - \left(\frac{1}{\text{SNR}_{MMSE-LE}}\right)^2 \bar{\mathbb{E}}_x = \frac{\text{SNR}^2_{MMSE-LE} - \bar{\mathbb{E}}_x}{\text{SNR}^2_{MMSE-LE}} = \frac{\bar{\mathbb{E}}_x}{\text{SNR}_{MMSE-LE}} (\text{SNR}_{MMSE-LE} - 1) \hspace{1cm} (3.174)$$

The SNR for the unbiased MMSE-LE then becomes (taking the ratio of (3.172) to $\sigma^2_e'$)

$$\text{SNR}_{MMSE-LE,U} = \frac{(\text{SNR}_{MMSE-LE} - 1)^2}{\bar{\mathbb{E}}_x} \bar{\mathbb{E}}_x = \text{SNR}_{MMSE-LE} - 1, \hspace{1cm} (3.175)$$

208
which corroborates the earlier result on the relation of the optimum biased and unbiased SNR’s for any particular receiver structure (at least for the LE structure). The unbiased SNR is

\[
\text{SNR}_{\text{MMSE-LE, U}} = \frac{\bar{E_x}}{\sigma_{\text{MMSE-LE}}^2} - 1 ,
\]

which is the performance level that this text always uses because it corresponds to the best error probability for an SBS detector, as was discussed earlier in Section 3.1. Figure 3.5.2 illustrates the concept with the effect of scaling on a 4QAM signal shown explicitly. Again, MMSE receivers (of which the MMSE-LE is one) reduce noise power at the expense of introducing a bias, the scaling up removes the bias, and ensures the detector achieves the best \( P_e \) it can.

Since the ZFE is also an unbiased receiver, (3.163) must hold. The first inequality in (3.163) tends to equality if the ratio \( \frac{\bar{E_x}}{\sigma^2} \to \infty \) or if the channel is free of ISI (\( Q(D) = 1 \)). The second inequality tends to equality if \( \frac{\bar{E_x}}{\sigma_2} \to 0 \) or if the channel is free of ISI (\( Q(D) = 1 \)).

The unbiased MMSE-LE loss with respect to the MFB is

\[
\gamma_{\text{MMSE-LE}} = \frac{\text{SNR}_{\text{MF}}}{\text{SNR}_{\text{MMSE-LE, U}}} = \left( \frac{\|p\|^2 \sigma_{\text{MMSE-LE}}^2}{\sigma^2} \right) \left( \frac{\bar{E_x}}{\bar{E_x} - \sigma_{\text{MMSE-LE}}^2} \right) .
\]

The \( \frac{\|p\|^2 \sigma_{\text{MMSE-LE}}^2}{\sigma^2} \) in (3.177) is the increase in noise variance of the MMSE-LE, while the term \( \frac{\bar{E_x}}{\bar{E_x} - \sigma_{\text{MMSE-LE}}^2} \) term represents the loss in signal power at the equalizer output that accrues to lower noise enhancement.

The MMSE-LE also requires no additional complexity to implement and should always be used in place of the ZFE when the receiver uses symbol-by-symbol detection on the equalizer output.

The error is not necessarily Gaussian in distribution. Nevertheless, engineers commonly make this assumption in practice, with a good degree of accuracy, despite the potential non-Gaussian residual ISI component in \( \sigma_{\text{MMSE-LE}}^2 \). This text also follows this practice. Thus,

\[
P_e \approx N_e Q \left( \sqrt{\kappa \text{SNR}_{\text{MMSE-LE, U}}} \right) ,
\]

(3.178)
where \( \kappa \) depends on the relation of \( \bar{E}_x \) to \( d_{\min} \) for the particular constellation of interest, for instance \( \kappa = 3/(M-1) \) for Square QAM. The reader may recall that the symbol \( Q \) is used in two separate ways in these notes, for the \( Q \)-function, and for the transform of the matched-filter-pulse-response cascade. The actual meaning should always be obvious in context.)

### 3.5.3 Examples Revisited

This section returns to the earlier ZFE examples to compute the improvement of the MMSE-LE on these same channels.

**EXAMPLE 3.5.1 (PAM - MMSE-LE)** The pulse response of a channel used with binary PAM is again given by

\[
P(\omega) = \begin{cases} \sqrt{T}(1 + 0.9e^{j\omega T}) & |\omega| \leq \frac{\pi}{T} \\ 0 & \text{otherwise} \end{cases},
\]

Let us suppose \( \text{SNR}_{MFB} = 10\text{dB} \) (= \( \bar{E}_x \|p\|^2/N_0 \)) and that \( \bar{E}_x = 1 \). The equalizer is

\[
W(D) = \frac{1}{\|p\|^2 \left( \frac{9}{1.81} D^{-1} + 1.1 + \frac{9}{1.81} D \right)}.
\]

Figure 3.30 shows the frequency response of the equalizer for both the ZFE and MMSE-LE. Clearly the MMSE-LE has a lower magnitude in its response.

![Equalizers for real-channel example](image)

Figure 3.30: Comparison of equalizer frequency-domain responses for MMSE-LE and ZFE on baseband example.
The $\sigma^2_{\text{MMSE-LE}}$ is computed as

$$\sigma^2_{\text{MMSE-LE}} = \frac{T}{2\pi} \int_{-\pi}^{\pi} \frac{\frac{\text{N}_0}{2}}{1.81 + 1.8 \cos(\omega T) + 1.81/10} d\omega$$

(3.181)

$$= \frac{\text{N}_0}{2} \sqrt{1.991^2 - 1.8^2}$$

(3.182)

$$= \frac{\text{N}_0}{2} (1.175)$$

(3.183)

which is considerably smaller than $\sigma^2_{\text{ZFE}}$. The SNR for the MMSE-LE is

$$\text{SNR}_{\text{MMSE-LE},U} = \frac{1 - 1.175(1.181)}{1.175(1.181)} = 3.7 \text{ (5.7dB)}$$

(3.184)

The loss with respect to the MFB is 10dB - 5.7dB = 4.3dB. This is 5.5dB better than the ZFE (9.8dB - 4.3dB = 5.5dB), but still not good for this channel.

Figure 3.31 compares the frequency-domain responses of the equalized channel.

![Frequency-domain comparison of equalized channel responses](image)

Figure 3.31: Comparison of equalized frequency-domain responses for MMSE-LE and ZFE on baseband example.

Figure 3.32 compares the time-domain responses of the equalized channel. The MMSE-LE clearly does not have an ISI-free response, but mean-square error/distortion is minimized and the MMSE-LE has better performance than the ZFE.
The relatively small energy near the Nyquist Frequency in this example is the reason for the poor performance of both linear equalizers (ZFE and MMSE-LE) in this example. To improve performance further, decision-assisted equalization is examined in Section 3.6 and in Chapter 5 (or codes combined with equalization, as discussed in Chapters 10 and 11). Another yet better method appears in Chapter 4.

The complex example also follows in a straightforward manner.

**EXAMPLE 3.5.2 (QAM - MMSE-LE)** Recall that the equivalent baseband pulse response samples were given by

\[
p_k = \frac{1}{\sqrt{T}} \left[ -\frac{1}{2}, \left(1 + \frac{j}{4}\right), -\frac{j}{2} \right].
\]  

Again, \( \text{SNR}_{MFB} = 10 \text{dB} \) \((= \bar{E}_x \|p\|^2 / \left\| \frac{N_0}{2} \right\|^2)\), \( \bar{E}_x = 1 \), and thus

\[
\sigma^2_{\text{MMSE-LE}} = w_0 \frac{N_0}{2} \|p\|^{-1}.
\]  

The equalizer is, noting that \( \frac{N_0}{2} = \frac{\bar{E}_x \|p\|^2}{\text{SNR}_{MFB}} = 1.5625/10 = .15625 \),

\[
W(D) = \frac{1}{(Q(D) + 1/\text{SNR}_{MFB}) \|p\|}
\]  

\[ (3.187) \]
or

\[ \sqrt{1.5625} \]

\[ = -0.25jD^{-2} + 0.625(-1 + j)D^{-1} + 1.5625(1 + j)D^{1} + 0.25jD^{2} \]  \( \cdot \) (3.188)

Figure 3.33 compares the MMSE-LE and ZFE equalizer responses for this same channel. The MMSE-LE high response values are considerably more limited than the ZFE values.

![Figure 3.33: Comparison of MMSE-LE and ZFE for complex channel example.](image)

The roots of \( Q(D) + 1/\text{SNR}_{MFB} \), or of the denominator in (3.188), are

\[ D = 2.217\angle^{-1.632} = -0.1356 - j2.213 \]  \( \cdot \) (3.189)
\[ D = 2.217\angle^{0.612} = 2.213 + j1.356 \]  \( \cdot \) (3.190)
\[ D = .451\angle^{-1.632} = -0.0276 - j.4502 \]  \( \cdot \) (3.191)
\[ D = .451\angle^{0.612} = .4502 + j.0276 \]  \( \cdot \) (3.192)

and (3.188) becomes

\[ W(D) = \frac{[\sqrt{1.5625D^{2}}/(25j)]}{(D - 2.22\angle^{-1.632})(D - 2.22\angle^{0.612})(D - .451\angle^{-1.632})(D - .451\angle^{0.612})} \]  \( \cdot \) (3.193)

or

\[ W(D) = \frac{A}{D - 2.22\angle^{0.612}} + \frac{B}{D - 2.22\angle^{-0.612}} + \ldots \]  \( \cdot \) (3.194)

where the second expression (3.194) has ignored any terms in the partial fraction expansion that will not contribute to \( w_{0} \). Then,

\[ A = \frac{\sqrt{1.5625(-4j)(2.22\angle^{-1.632})^{2}}}{(2.22\angle^{0.612} - 2.22\angle^{0.612})(2.22\angle^{-1.632} - .451\angle^{-1.632})(2.22\angle^{1.632} - .451\angle^{0.612})} \]  \( \cdot \) (3.195)
\[ B = \frac{\sqrt{1.5625(-4j)(2.22\angle^{0.612})^{2}}}{(2.22\angle^{0.612} - 2.22\angle^{-1.632})(2.22\angle^{1.632} - .451\angle^{1.632})(2.22\angle^{-0.612} - .451\angle^{0.612})} \]  \( \cdot \) (3.196)
Then, from (3.194),
\[ w_0 = A(-.451\angle^{1.632}) + B(-.451\angle^{-0.612}) = \sqrt{1.5625(1.125)} \quad (3.197) \]

Then
\[ \sigma^2_{\text{MMSE-LE}} = \frac{N_0}{2} = 1.125(15625) = .1758 \quad , \quad (3.198) \]
\[ \text{SNR}_{\text{MMSE-LE,U}} = \frac{[1 - 1.125(15625)]}{125(15625)} = 4.69 \quad (6.7 \text{ dB}) \quad , \quad (3.199) \]

which is 3.3 dB below the matched filter bound SNR of 10dB, or equivalently,
\[ \gamma_{\text{MMSE-LE}} = 10 \log_{10}(10/4.69) = 10 \log_{10}(2.13) = 3.3 \text{ dB} \quad , \quad (3.200) \]

but .6 dB better than the ZFE. It is also possible to do considerably better on this channel, but structures not yet introduced will be required. Nevertheless, the MMSE-LE is one of the most commonly used equalization structures in practice.

Figure 3.33 compares the equalized channel responses for both the MMSE-LE and ZFE. While the MMSE-LE performs better, there is a significant deviation from the ideal flat response. This deviation is good and gains .6 dB improvement. Also, because of biasing, the MMSE-LE output is everywhere slightly lower than the ZFE output (which is unbiased). This bias can be removed by scaling by $5.69/4.69$.

### 3.5.4 Fractionally Spaced Equalization

To this point in the study of the discrete time equalizer, a matched filter $\varphi_p^*(-t)$ precedes the sampler, as was shown in Figure 3.1.3. While this may be simple from an analytical viewpoint, there are several practical problems with the use of the matched filter. First, $\varphi_p^*(-t)$ is a continuous time filter and may be much more difficult to design accurately than an equivalent digital filter. Secondly, the precise sampling frequency and especially its phase, must be known so that the signal can be sampled when it is at maximum strength. Third, the channel pulse response may not be accurately known at the receiver, and an adaptive equalizer (see Chapter 7), is instead used. It may be difficult to design an adaptive, analog, matched filter.

![Fractionally spaced equalization](image)

For these reasons, sophisticated data transmission systems often replace the matched-filter/sampler/equalizer system of Figure 3.1.3 with the Fractionally Spaced Equalizer (FSE) of Figure 3.5.4. Basically, the sampler and the matched filter have been interchanged with respect to Figure 3.1.3. Also the sampling rate has been increased by some rational number $l$ ($l > 1$). The new sampling rate is chosen sufficiently high so as to be greater than twice the highest frequency of $x_p(t)$. Then, the matched filtering operation and the equalization filtering are performed at rate $l/T$, and the cascade of the two filters is realized as a single filter in practice. An anti-alias or noise-rejection filter always precedes the sampling device.
To derive the settings for the FSE, this section assumes

**3.5.4.0.1 Infinite-Length FSE Settings**

complex to implement (more parameters and higher sampling rate). Frequency, the FSE is avoided, because it provides little performance gain, and is significantly more

the FSE with respect to the symbol-spaced equalizer. In channels with little energy near the Nyquist Frequency in applications that exhibit a significant improvement of

phase. The phase is tacitly corrected to the optimum phase inside the linear filter implementing the

the sampling device need only be locked to the symbol rate, but can otherwise provide any sampling

implementation, if the extra memory and computation can be accommodated. Effectively, with an FSE,

equalizer aliases before it equalizes; the former alternative is often the one of choice in practical system

example of information loss at some frequency; this loss cannot be recovered in a symbol-spaced equalizer

a speed that is too low in the symbol-spaced equalizer without matched filter. This possible notch is an

performed at the symbol rate. Equivalently, information has been lost about the signal by sampling at

symbol-spaced ZFE and/or MMSE-LE can be significantly reduced, because of noise enhancement. The

producing a notch within the critical frequency range (-\(\frac{\pi}{T}\), \(\frac{\pi}{T}\)), and at frequencies just below the Nyquist Frequency) that the two

aliased frequency characteristics \(Q(\omega)e^{-j\omega t_0}\) and \(Q(\omega - \frac{\pi}{T})e^{-j(\omega - \frac{\pi}{T})t_0}\) add to approximately zero, thus

producing a notch within the critical frequency range (-\(\frac{\pi}{T}\), \(\frac{\pi}{T}\)). Then, the best performance of the ensuing symbol-spaced ZFE and/or MMSE-LE can be significantly reduced, because of noise enhancement. The

resulting noise enhancement can be a major problem in practice, and the loss in performance can be several dB for reasonable timing offsets. When the anti-alias filter output is sampled at greater than twice the highest frequency in \(x_p(t)\), then information about the entire signal waveform is retained. Equivalently, the FSE can synthesize, via its transfer characteristic, a phase adjustment (effectively

interpolating to the correct phase) so as to correct the timing offset, \(t_0\), in the sampling device. The symbol-spaced equalizer cannot interpolate to the correct phase, because no interpolation is correctly

performed at the symbol rate. Equivalently, information has been lost about the signal by sampling at a speed that is too low in the symbol-spaced equalizer without matched filter. This possible notch is an example of information loss at some frequency; this loss cannot be recovered in a symbol-spaced equalizer without a matched filter. In effect, the FSE equalizes before it aliases (aliasing does occur at the output of the equalizer where it decimates by \(l\) for symbol-by-symbol detection), whereas the symbol-spaced equalizer aliases before it equalizes; the former alternative is often the one of choice in practical system implementation, if the extra memory and computation can be accommodated. Effectively, with an FSE, the sampling device need only be locked to the symbol rate, but can otherwise provide any sampling phase. The phase is tacitly corrected to the optimum phase inside the linear filter implementing the FSE.

The sensitivity to sampling phase is channel dependent: In particular, there is usually significant channel energy near the Nyquist Frequency in applications that exhibit a significant improvement of the FSE with respect to the symbol-spaced equalizer. In channels with little energy near the Nyquist Frequency, the FSE is avoided, because it provides little performance gain, and is significantly more complex to implement (more parameters and higher sampling rate).

**3.5.4.0.1 Infinite-Length FSE Settings**

To derive the settings for the FSE, this section assumes that \(l\), the oversampling factor, is an integer. The sampled output of the anti-alias filter can be decom-
posed into \( l \) sampled-at-rate-1/\( T \) interleaved sequences with \( D \)-transforms \( Y_0(D), Y_2(D), \ldots, Y_{l-1}(D) \), where \( Y_i(D) \) corresponds to the sample sequence \( y[kT - iT/l] \). Then,

\[
Y_i(D) = P_i(D) \cdot X(D) + N_i(D) \tag{3.203}
\]

where \( P_i(D) \) is the transform of the symbol-rate-spaced-\( i \)-th-phase-of-\( p(t) \) sequence \( p[kT - (i - 1)T/l] \), and similarly \( N_i(D) \) is the transform of a symbol-rate sampled white noise sequence with autocorrelation function \( R_{nn}(D) = l \cdot \frac{N_0}{2} \) per dimension, and these noise sequences are also independent of one another. A column vector transform is

\[
\mathbf{Y}(D) = \begin{bmatrix} Y_0(D) \\ \vdots \\ Y_{l-1}(D) \end{bmatrix} = \mathbf{P}(D)X(D) + \mathbf{N}(D) . \tag{3.204}
\]

Also

\[
\mathbf{P}(D) = \begin{bmatrix} P_0(D) \\ \vdots \\ P_{l-1}(D) \end{bmatrix} \quad \text{and} \quad \mathbf{N}(D) = \begin{bmatrix} N_0(D) \\ \vdots \\ N_{l-1}(D) \end{bmatrix} . \tag{3.205}
\]

By considering the FSE output at sampling rate 1/\( T \), the interleaved coefficients of the FSE can also be written in a row vector \( \mathbf{W}(D) = [W_1(D), \ldots, W_l(D)] \). Thus the FSE output is

\[
\mathbf{Z}(D) = \mathbf{W}(D)\mathbf{Y}(D) . \tag{3.206}
\]

Again, the orthogonality condition says that \( E(D) = X(D) - Z(D) \) should be orthogonal to \( Y(D) \), which can be written in vector form as

\[
E \left[ E(D)Y^*(D^{-*}) \right] = \mathbf{R}_{\mathbf{Z}\mathbf{Y}}(D) - \mathbf{W}(D)\mathbf{R}_{\mathbf{YY}}(D) = 0 , \tag{3.207}
\]

where

\[
\mathbf{R}_{\mathbf{Z}\mathbf{Y}}(D) \triangleq E \left[ X(D)Y^*(D^{-*}) \right] = \mathbf{\hat{e}}_x\mathbf{P}^*(D^{-*}) \tag{3.208}
\]

\[
\mathbf{R}_{\mathbf{YY}}(D) \triangleq E \left[ Y(D)Y^*(D^{-*}) \right] = \mathbf{\hat{e}}_x\mathbf{P}(D)\mathbf{P}^*(D^{-*}) + l \cdot \frac{N_0}{2} \mathbf{I} . \tag{3.209}
\]

MMSE-FSE filter setting is then

\[
\mathbf{W}(D) = \mathbf{R}_{\mathbf{Z}\mathbf{Y}}(D)\mathbf{R}_{\mathbf{YY}}^{-1}(D) = \mathbf{P}^*(D^{-*}) \left[ \mathbf{P}(D)\mathbf{P}^*(D^{-*}) + l / \text{SNR} \right]^{-1} . \tag{3.210}
\]

The corresponding error sequence has autocorrelation function

\[
\hat{R}_{\mathbf{e}\mathbf{e}}(D) = \mathbf{\hat{e}}_x - \mathbf{R}_{\mathbf{Z}\mathbf{Y}}(D)\mathbf{R}_{\mathbf{YY}}^{-1}(D)\mathbf{R}_{\mathbf{Y}\mathbf{e}}(D) = \frac{l \cdot \frac{N_0}{2}}{\mathbf{P}^*(D^{-*}) \mathbf{P}(D) + l / \text{SNR}} . \tag{3.211}
\]

The MMSE is then computed as

\[
\text{MMSE}_{\text{MMSE-FSE}} = \frac{T}{2\pi} \int_{-\pi}^{\pi} \left\| \mathbf{P}(e^{-j\omega T}) \right\|^2 + l / \text{SNR} = \text{MMSE}_{\text{MMSE-LE}} , \tag{3.212}
\]

where \( \left\| \mathbf{P}(e^{-j\omega T}) \right\|^2 = \sum_{i=1}^{l} \left| P_i(e^{-j\omega T}) \right|^2 \). The SNR’s, biased and unbiased, are also then exactly the same as given for the MMSE-LE, as long as the sampling rate exceeds twice the highest frequency of \( P(f) \).

The reader is cautioned against letting \( \frac{N_0}{2} \to 0 \) to get the ZF-FSE. This is because the matrix \( \mathbf{R}_{\mathbf{YY}}(D) \) will often be singular when this occurs. To avoid problems with singularity, an appropriate pseudoinverse, which zeros the FSE characteristic when \( P(\omega) = 0 \) is recommended.
3.5.4.1  **Passband Equalization**

This chapter so far assumed for the complex baseband equalizer that the passband signal has been de-modulated according to the methods described in Chapter 2 before filtering and sampling. An alternative method, sometimes used in practice when $\omega_c$ is not too high (or intermediate-frequency up/down conversion is used), is to commute the demodulation and filtering functions. Sometimes this type of system is also called “direct conversion,” meaning that no carrier demodulation occurs prior to sampling. This commuting can be done with either symbol-spaced or fractionally spaced equalizers. The main reason for the interchange of filtering and demodulation is the recovery of the carrier frequency in practice. Postponing the carrier demodulation to the equalizer output can lead to significant improvement in the tolerance of the system to any error in estimating the carrier frequency, as Chapter 6 investigates. In the present (perfect carrier phase lock) development, passband equalization and baseband equalization are exactly equivalent and the settings for the passband equalizer are identical to those of the corresponding baseband equalizer, other than a translation in frequency by $\omega_c$ radians/sec.

Passband equalization is best suited to CAP implementations (See Section 4 of Chapter 2) where the complex channel is the analytic equivalent. The complex equalizer then acts on the analytic equivalent channel and signals to eliminate ISI in a MMSE sense. When used with CAP, the final rotation shown in Figure 3.5.4.1 is not necessary – such a rotation is only necessary with a QAM transmit signal.

![Diagram of Passband Equalization](image)

**Figure 3.35:** Passband equalization, direct synthesis of analytic-equivalent equalizer.

The passband equalizer is illustrated in Figure 3.5.4.1. The phase splitting is the forming of the analytic equivalent with the use of the Hilbert Transform, as discussed in Chapter 2. The matched filtering and equalization are then performed on the passband signal with digital demodulation to baseband deferred to the filter output. The filter $w_k$ can be a ZFE, a MMSE-LE, or any other desired setting (see Chapter 4 and Sections 3.5 and 3.6). In practice, the equalizer can again be realized as a fractionally spaced equalizer by interchanging the positions of the matched filter and sampler, increasing the sampling rate, and absorbing the matched filtering operation into the filter $w_k$, which would then be identical to the (modulated) baseband-equivalent FSE with output sample rate decimated appropriately to produce outputs only at the symbol sampling instants.

Often in practice, the cross-coupled nature of the complex equalizer $W(D)$ is avoided by sampling the FSE at a rate that exceeds twice the highest frequency of the passband signal $y(t)$ prior to demodulation. (If intermediate frequency (IF) demodulation is used, then twice the highest frequency after the IF.) In this case the phase splitter (contains Hilbert transform in parallel with a unit gain) is also absorbed into the equalizer, and the same sampled sequence is applied independently to two equalizers, one of which estimates the real part, and one the imaginary part, of the analytic representation of the data sequence $x_k e^{j\omega_c kT}$. The two filters can sometimes (depending on the sampling rate) be more cost effective to implement than the single complex filter, especially when one considers that the Hilbert transform is incorporated into the equalizer. This approach is often taken with adaptive implementations of the FSE, as the Hilbert transform is then implemented more exactly adaptively than is possible with fixed design.\(^\text{10}\) This is often called **Nyquist Inband Equalization**. A particularly common variant of this

\(^{10}\)However, this structure also slows convergence of many adaptive implementations.
type of equalization is to sample at rate $2/T$ both of the outputs of a continuous-time phase splitter, one stream of samples staggered by $T/4$ with respect to the other. The corresponding $T/2$ complex equalizer can then be implemented adaptively with four independent adaptive equalizers (rather than the two filters, real and imaginary part, that nominally characterize complex convolution). The adaptive filters will then correct for any imperfections in the phase-splitting process, whereas the two filters with fixed conjugate symmetry could not.
3.6 Decision Feedback Equalization

Decision Feedback Equalization makes use of previous decisions in attempting to estimate the current symbol (with an SBS detector). Any “trailing” intersymbol interference caused by previous symbols is reconstructed and then subtracted. The DFE is inherently a nonlinear receiver. However, it can be analyzed using linear techniques, if one assumes all previous decisions are correct. There is both a MMSE and a zero-forcing version of decision feedback, and as was true with linear equalization in Sections 3.4 and 3.5, the zero-forcing solution will be a special case of the least-squares solution with the SNR \( \rightarrow \infty \). Thus, this section derives the MMSE solution, which subsumes the zero-forcing case when SNR \( \rightarrow \infty \).

The Decision Feedback Equalizer (DFE) is shown in Figure 3.6. The configuration contains a linear feedforward equalizer, \( W(D) \), (the settings for this linear equalizer are not necessarily the same as those for the ZFE or MMSE-LE), augmented by a linear, causal, feedback filter, \( 1 - B(D) \), where \( b_0 = 1 \). The feedback filter accepts as input the decision from the previous symbol period; thus, the name “decision feedback.” The output of the feedforward filter is denoted \( Z(D) \), and the input to the decision element \( Z'(D) \). The feedforward filter will try to shape the channel output signal so that it is a causal signal. The feedback section will then subtract (without noise enhancement) any trailing ISI. Any bias removal in Figure 3.6 is absorbed into the SBS.

This section assumes that previous decisions are correct. In practice, this may not be true, and can be a significant weakness of decision-feedback that cannot be overlooked. Nevertheless, the analysis becomes intractable if it includes errors in the decision feedback section. To date, the most efficient way to specify the effect of feedback errors has often been via measurement. Section 3.7 provides an exact error-propagation analysis for finite-length DFE’s that can (unfortunately) require enormous computation on a digital computer. Section 3.8.1 shows how to eliminate error propagation with precoders.

3.6.1 Minimum-Mean-Square-Error Decision Feedback Equalizer (MMSE-DFE)

The Minimum-Mean-Square Error Decision Feedback Equalizer (MMSE-DFE) jointly optimizes the settings of both the feedforward filter \( w_k \) and the feedback filter \( \delta_k - b_k \) to minimize the MSE:

\[
|e_k| = |x_k - z'_k|.
\]
The MMSE for the MMSE-DFE is

\[
\sigma^2_{\text{MMSE-DFE}} = \min_{w_k, b_k} E \left[ |x_k - z'_k|^2 \right].
\] (3.214)

The error sequence can be written as

\[
E(D) = X(D) - W(D) \cdot Y(D) - [1 - B(D)]X(D) = B(D) \cdot X(D) - W(D) \cdot Y(D).
\] (3.215)

For any fixed \(B(D)\), \(E[D]Y^{*}(D^{-*}) = 0\) to minimize MSE, which leads us to the relation

\[
B(D) \cdot \bar{R}_{xy}(D) - W(D) \cdot \bar{R}_{yy}(D) = 0.
\] (3.216)

Thus,

\[
W(D) = \frac{B(D)}{||p|| \left( Q(D) + \frac{1}{\text{SNR}_{MFB}} \right)} = B(D) \cdot W_{\text{MMSE-LE}(D)}
\] (3.217)

for any \(B(D)\) with \(b_0 = 1\). (Also, \(W(D) = W_{\text{MMSE-LE}(D)} \cdot B(D)\), a consequence of the linearity of the MMSE estimate, so that \(E(D) = B(D) \cdot E_{\text{MMSE-LE}(D)}\))

The autocorrelation function for the error sequence with arbitrary monic \(B(D)\) is

\[
\tilde{R}_{ee}(D) = B(D)\tilde{\ell}_x B^{*}(D^{-*}) - 2\Re \{B(D)\bar{R}_{xy}(D)W^{*}(D^{-*})\} + W(D)\bar{R}_{yy}(D)W^{*}(D^{-*})
\] (3.218)

\[
= B(D)\tilde{\ell}_x B^{*}(D^{-*}) - W(D)\bar{R}_{yy}(D)W^{*}(D^{-*})
\] (3.219)

\[
= B(D)\tilde{\ell}_x B^{*}(D^{-*}) - B(D) \frac{\bar{R}_{xy}(D)}{||p|| \left( Q(D) + \frac{1}{\text{SNR}_{MFB}} \right)} B^{*}(D^{-*})
\] (3.220)

\[
= B(D)R_{\text{MMSE-LE}(D)}B^{*}(D^{-*})
\] (3.221)

where \(R_{\text{MMSE-LE}(D)} = \frac{\gamma_0}{||p||^2} 1/(Q(D) + 1/\text{SNR}_{MFB})\) is the autocorrelation function for the error sequence of a MMSE linear equalizer. The solution for \(B(D)\) is then the forward prediction filter associated with this error sequence as discussed in the Appendix. The linear prediction approach is developed more in what is called the “linear prediction DFE,” in an exercise where the MMSE-DFE is the concatenation of a MMSE-LE and a linear-predictor that whitens the error sequence.

In more detail on \(B(D)\), the (scaled) inverse autocorrelation has spectral factorization:

\[
Q(D) + \frac{1}{\text{SNR}_{MFB}} = \gamma_0 \cdot G(D) \cdot G^{*}(D^{-*}),
\] (3.222)

where \(\gamma_0\) is a positive real number and \(G(D)\) is a canonical filter response. A filter response \(G(D)\) is called canonical if it is causal \((g_k = 0 \text{ for } k < 0)\), monic \((g_0 = 1)\), and minimum-phase (all its poles are outside the unit circle, and all its zeroes are on or outside the unit circle). If \(G(D)\) is canonical, then \(G^{*}(D^{-*})\) is anti-canonical, i.e., anti-causal, monic, and “maximum-phase” (all poles inside the unit circle, and all zeros in or on the unit circle). Using this factorization,

\[
\tilde{R}_{ee}(D) = \frac{B(D) \cdot B^{*}(D^{-*})}{Q(D) + 1/\text{SNR}_{MFB} \cdot ||p||^2} \cdot \frac{\gamma_0}{2}
\] (3.223)

\[
= B(D) \cdot G^{*}(D^{-*}) \cdot \frac{\gamma_0}{2} \cdot \frac{1}{\gamma_0 ||p||^2}
\] (3.224)

\[
r_{ee,0} \geq \frac{\gamma_0}{2} \frac{1}{\gamma_0 ||p||^2},
\] (3.225)
with equality if and only if $B(D) = G(D)$. Thus, the MMSE will then be $\sigma_{\text{MMSE-DFE}}^2 = \frac{N_0^2}{\gamma_0 \|p\|^2}$. The feedforward filter then becomes

$$
W(D) = \frac{G(D)}{\|p\| \cdot \gamma_0 \cdot G(D) \cdot G^*(D^{-*})} = \frac{1}{\|p\| \cdot \gamma_0 \cdot G^*(D^{-*})}.
$$

(3.226)

The last step in (3.225) follows from the observations that

$$
\bar{r}_{ee,0} = \|B\| \frac{N_0^2}{\gamma_0 \cdot \|p\|^2},
$$

(3.227)

the fractional polynomial inside the squared norm is necessary monic and causal, and therefore the squared norm has a minimum value of 1. $B(D)$ and $W(D)$ specify the MMSE-DFE:

### Lemma 3.6.1 (MMSE-DFE)

The MMSE-DFE has feedforward section

$$
W(D) = \frac{1}{\|p\| \cdot \gamma_0 \cdot G^*(D^{-*})}
$$

(3.228)

(realized with delay, as it is strictly noncausal) and feedback section

$$
B(D) = G(D)
$$

(3.229)

where $G(D)$ is the unique canonical factor of the following equation:

$$
Q(D) + \frac{1}{\text{SNR}_{MBF}} = \gamma_0 \cdot G(D) \cdot G^*(D^{-*})
$$

(3.230)

This text also calls the joint matched-filter/sampler/W(D) combination in the forward path of the DFE the “Mean-Square Whitened Matched Filter (MS-WMF)”. These settings for the MMSE-DFE minimize the MSE as was shown above.

### 3.6.2 Performance Analysis of the MMSE-DFE

Again, the autocorrelation function for the error sequence is

$$
\bar{r}_{ee}(D) = -\frac{N_0}{\gamma_0 \cdot \|p\|^2}.
$$

(3.231)

This last result states that the error sequence for the MMSE-DFE is “white” when minimized (since $\bar{r}_{ee}(D)$ is a constant) and has MMSE or average energy (per real dimension) $\frac{N_0}{\|p\|^2 \cdot \gamma_0^{-1}}$. Also,

$$
\frac{T}{2\pi} \int_{-\frac{T}{2}}^{\frac{T}{2}} \ln \left( Q(e^{-j\omega T}) + \frac{1}{\text{SNR}_{MBF}} \right) d\omega = \ln(\gamma_0) + \frac{T}{2\pi} \int_{-\frac{T}{2}}^{\frac{T}{2}} \ln \left( G(e^{-j\omega T}) G^*(e^{-j\omega T}) \right) d\omega
$$

(3.232)

$$
= \ln(\gamma_0).
$$

(3.233)

(The last line follows from $\ln \left( G(e^{-j\omega T}) \right)$ being a periodic function integrated over one period of its fundamental frequency, and similarly for $\ln \left( G^*(e^{-j\omega T}) \right)$.) This last result leads to a famous expression for $\sigma_{\text{MMSE-DFE}}^2$, which was first derived by Salz in 1973,

$$
\sigma_{\text{MMSE-DFE}}^2 = \frac{N_0}{\|p\|^2} \cdot e^{-\frac{T}{2\pi} \int_{-\frac{T}{2}}^{\frac{T}{2}} \ln \left( Q(e^{-j\omega T}) + \frac{1}{\text{SNR}_{MBF}} \right) d\omega}.
$$

(3.234)
The SNR for the MMSE-DFE can now be easily computed as

$$\text{SNR}_{\text{MMSE-DFE}} = \frac{\hat{e}_x}{\sigma^2_{\text{MMSE-DFE}}} = \gamma_0 \cdot \text{SNR}_{\text{MFB}}$$ (3.235)

$$= \text{SNR}_{\text{MFB}} \frac{T}{2\pi} \frac{1}{\pi^2} \ln \left( Q(e^{-\pi^2}) + \frac{1}{\text{SNR}_{\text{MFB}}} \right) d\omega \cdot \text{ (3.236)}$$

From the $k = 0$ term in the defining spectral factorization, $\gamma_0$ can also be written

$$\gamma_0 = \frac{1 + 1/\text{SNR}_{\text{MFB}}}{\|g\|^2} = \frac{1 + 1/\text{SNR}_{\text{MFB}}}{1 + \sum_{i=1}^{\infty} |g_i|^2} \ . \ (3.237)$$

From this expression, if $G(D) = 1$ (no ISI), then $\text{SNR}_{\text{MMSE-DFE}} = \text{SNR}_{\text{MFB}} + 1$, so that the SNR would be higher than the matched filter bound. The reason for this apparent anomaly is the artificial signal power introduced by biased decision regions. This bias is noted by writing

$$Z'(D) = X(D) - E(D) \quad (3.238)$$

$$= X(D) - G(D) \cdot X(D) + \frac{1}{\|p\| \cdot \gamma_0 \cdot G^{\ast}(D^{-\ast})} Y(D) \quad (3.239)$$

$$= X(D) - G(D)X(D) + \frac{Q(D)}{\gamma_0 G^{\ast}(D^{-\ast})} X(D) + N(D) \|p\| \cdot \gamma_0 G^{\ast}(D^{-\ast}) \quad (3.240)$$

$$= X(D) - \frac{1}{\text{SNR}_{\text{MFB}}} \frac{\gamma_0 \cdot G^{\ast}(D^{-\ast})}{\gamma_0 \cdot G^{\ast}(D^{-\ast})} X(D) + N(D) \|p\| \cdot \gamma_0 G^{\ast}(D^{-\ast}) \cdot (3.241)$$

The current sample, $x_k$, corresponds to the time zero sample of $1 - \frac{1}{\text{SNR}_{\text{MFB}}} \gamma_0 \cdot G^{\ast}(D^{-\ast})$, and is equal to $(1 - \frac{1}{\text{SNR}_{\text{MFB}}} \gamma_0 \cdot G^{\ast}(D^{-\ast})) x_k$. Thus $z_k'$ contains a signal component of $x_k$ that is reduced in magnitude. Thus, again using the result from Section 3.2.1, the SNR corresponding to the lowest probability of error corresponds to the same MMSE-DFE receiver with output scaled to remove the bias. This leads to the more informative SNR

$$\text{SNR}_{\text{MMSE-DFE,U}} = \text{SNR}_{\text{MMSE-DFE}} - 1 = \gamma_0 \cdot \text{SNR}_{\text{MFB}} - 1 \ . \ (3.242)$$

index SNR!MMSE!DFE If $G(D) = 1$, then $\text{SNR}_{\text{MMSE-DFE,U}} = \text{SNR}_{\text{MFB}}$. Also,

$$\gamma_{\text{MMSE-DFE}} = \frac{\text{SNR}_{\text{MFB}}}{\text{SNR}_{\text{MMSE-DFE,U}}} = \frac{\text{SNR}_{\text{MFB}}}{\gamma_0 \cdot \text{SNR}_{\text{MFB}} - 1} = \frac{1}{\gamma_0 - 1/\text{SNR}_{\text{MFB}}} \ . \ (3.243)$$

Rather than scale the decision input, the receiver can scale (up) the feedforward output by $\frac{1}{\text{SNR}_{\text{MMSE-DFE}}}$. This will remove the bias, but also increase MSE by the square of the same factor. The SNR will then be $\text{SNR}_{\text{MMSE-DFE,U}}$. This result is verified by writing the MS-WMF output

$$Z(D) = [X(D) \cdot \|p\| \cdot Q(D) + N(D)] \cdot \frac{1}{\|p\| \cdot \gamma_0 \cdot G^{\ast}(D^{-\ast})}$$ (3.244)

where $N(D)$, again, has autocorrelation $R_{nn}(D) = \frac{\pi}{2} Q(D)$. $Z(D)$ expands to

$$Z(D) = [X(D) \cdot \|p\| \cdot Q(D) + N(D)] \cdot \frac{1}{\|p\| \cdot \gamma_0 \cdot G^{\ast}(D^{-\ast})} \quad (3.245)$$

$$= X(D) \gamma_0 \cdot G(D) \cdot G^{\ast}(D^{-\ast}) - 1/\text{SNR}_{\text{MFB}} + N(D) \|p\| \cdot \gamma_0 \cdot G^{\ast}(D^{-\ast}) \quad (3.246)$$

$$= X(D) \cdot G(D) - \frac{X(D)}{\text{SNR}_{\text{MFB}} \cdot \gamma_0 \cdot G^{\ast}(D^{-\ast})} + N'(D) \quad (3.247)$$

$$= X(D) \left[ G(D) - \frac{1}{\text{SNR}_{\text{MMSE-DFE}}} \right] + \frac{1}{\text{SNR}_{\text{MMSE-DFE}}} \left[ 1 - \frac{1}{G^{\ast}(D^{-\ast})} \right] X(D) + N'(D) \quad (3.248)$$

$$= X(D) \left[ G(D) - \frac{1}{\text{SNR}_{\text{MMSE-DFE}}} \right] + E'(D) \ , \quad (3.249)$$

222
where \( N'(D) \) is the filtered noise at the MS-WMF output, and has autocorrelation function

\[
\tilde{R}_{n'n'}(D) = \frac{\hat{N}'(D)}{\|p\|^2 \cdot \gamma_0} \cdot G(D) \cdot G^*(D^{-*}) = \frac{\tilde{\epsilon}_x}{SNR_{MMSE-DFE}} \left[ 1 - \frac{1}{SNR_{MFB}} \frac{Q(D)}{Q(D) + 1/SNR_{MFB}} \right],
\]

and

\[
\epsilon'_k = -\epsilon_k + \frac{1}{SNR_{MMSE-DFE}} x_k. \tag{3.251}
\]

The error \( \epsilon'_k \) is not a white sequence in general. Since \( x_k \) and \( \epsilon'_k \) are uncorrelated, \( \sigma^2_e = \sigma^2_{MMSE-DFE} = \sigma^2_e + \frac{\tilde{\epsilon}_x}{SNR_{MMSE-DFE}} \). If one defines an “unbiased” monic, causal polynomial

\[
G_U(D) = \frac{SNR_{MMSE-DFE}}{SNR_{MMSE-DFE,U}} \left[ G(D) - \frac{1}{SNR_{MMSE-DFE}} \right],
\]

then the MMSE-DFE output is

\[
Z(D) = \frac{SNR_{MMSE-DFE,U}}{SNR_{MMSE-DFE}} X(D) G_U(D) + E'(D). \tag{3.253}
\]

\( Z_U(D) \) removes the scaling

\[
Z_U(D) = Z(D) \frac{SNR_{MMSE-DFE,U}}{SNR_{MMSE-DFE}} = X(D) G_U(D) + E_U(D), \tag{3.254}
\]

where \( E_U(D) = \frac{SNR_{MMSE-DFE}}{SNR_{MMSE-DFE,U}} E'(D) \) and has power

\[
\sigma^2_{e_U} = \left( \sigma^2_{MMSE-DFE} - \frac{\tilde{\epsilon}_x}{SNR_{MMSE-DFE,U}} \right) \left( \frac{SNR_{MMSE-DFE}}{SNR_{MMSE-DFE,U}} \right)^2 \tag{3.255}
\]

\[
= \frac{SNR^2_{MMSE-DFE} \sigma^2_{MMSE-DFE} - \tilde{\epsilon}_x}{SNR^2_{MMSE-DFE}} \cdot \frac{SNR^2_{MMSE-DFE}}{SNR^2_{MMSE-DFE,U}} \tag{3.256}
\]

\[
= \frac{\tilde{\epsilon}_x ([SNR_{MMSE-DFE,U}]^{-1} - 1)}{SNR^2_{MMSE-DFE,U}} \tag{3.257}
\]

\[
= \frac{\tilde{\epsilon}_x}{SNR_{MMSE-DFE,U}} \tag{3.258}
\]

SNR for this scaled and unbiased MMSE-DFE is thus verified to be \( \frac{\tilde{\epsilon}_x}{SNR_{MMSE-DFE,U}} = SNR_{MMSE-DFE,U} \).

The feedback section of this new unbiased MMSE-DFE becomes the polynomial \( 1 - G_U(D) \), if the scaling is performed before the summing junction, but is \( 1 - G(D) \) if scaling is performed after the summing junction. The alternative unbiased MMSE-DFE is shown in Figure 3.37.

All the results on fractional spacing for the LE still apply to the feedforward section of the DFE – thus this text does not reconsider them for the DFE, other than to state that the characteristic of the feedforward section is different than the LE characteristic in all but trivial channels. Again, the realization of any characteristic can be done with a matched filter and symbol-spaced sampling or with an anti-alias filter and fractionally spaced sampling, as in Section 3.5.4.

### 3.6.3 Zero-Forcing DFE

The ZF-DFE feedforward and feedback filters are found by setting \( SNR_{MFB} \to \infty \) in the expressions for the MMSE-DFE. The spectral factorization (assuming \( \ln(Q(e^{j\omega T})) \) is integrable over \((-\frac{\pi}{T}, \frac{\pi}{T})\), see Appendix A)

\[
Q(D) = \eta_0 \cdot P_c(D) \cdot P'_c(D^{-*}) \tag{3.260}
\]
Figure 3.37: Decision feedback equalization with unbiased feedback filter.

then determines the ZF-DFE filters as

\[ B(D) = P_c(D) \quad , \]

\[ W(D) = \frac{1}{\eta_0 \cdot ||p|| \cdot P_c^*(D^{-1})} \quad . \]

\( P_c \) is sometimes called a canonical pulse response for the channel. Since \( q_0 = 1 = \eta_0 \cdot ||P_c||^2 \), then

\[ \eta_0 = \frac{1}{1 + \sum_{i=1}^{\infty} |p_{c,i}|^2} \quad . \]  

Equation (3.263) shows that there is a signal power loss of the ratio of the squared first tap magnitude in the canonical pulse response to the squared norm of all the taps. This loss ratio is minimized for a minimum phase polynomial, and \( P_c(D) \) is the minimum-phase equivalent of the channel pulse response.

The noise at the output of the feedforward filter has (white) autocorrelation function

\[ N_0^2 / (||p||^2 \cdot \eta_0) \], so

\[ \sigma_{ZF-DFE}^2 = \eta_0 \cdot \operatorname{SNR}_{MFB} = \frac{1}{1 + \sum_{i=1}^{\infty} |p_{c,i}|^2} \cdot \operatorname{SNR}_{MFB} \quad . \]  

3.6.4 Examples Revisited

**EXAMPLE 3.6.1 (PAM - DFE)** The pulse response of a channel used with binary PAM is again given by

\[ P(\omega) = \begin{cases} \sqrt{T} (1 + .9 e^{j\omega T}) & |\omega| \leq \frac{T}{2} \\ 0 & |\omega| > \frac{T}{2} \end{cases} \quad . \]  

Again, the \( \operatorname{SNR}_{MFB} \) is 10dB and \( \varepsilon_\infty = 1 \).

Then \( \tilde{Q}(D) \triangleq Q(D) + 1/\operatorname{SNR}_{MFB} \) is computed as

\[ \tilde{Q}(e^{-j\omega T}) = \frac{1.81 + 1.8 \cos(\omega T) + 1.81/10}{1.81} \quad . \]  

\[ \tilde{Q}(D) = \frac{1}{1.81} (1 + .9D)(1 + .9D^{-1}) + .1 \quad . \]  

\[ = \frac{1}{1.81} (.9D + 1.991 + .9D^{-1}) \quad . \]  

The roots of \( \tilde{Q}(D) \) are \(-.633\) and \(-1.58\). The function \( \tilde{Q}(D) \) has canonical factorization

\[ \tilde{Q}(D) = .785(1 + .633D)(1 + .633D^{-1}) \quad . \]
Then, $\gamma_0 = .785$. The feedforward section is

$$W(D) = \frac{1}{\sqrt{1.81(.785)(1 + .633D^{-1})}} = \frac{.9469}{1 + .633D^{-1}} ,$$

and the feedforward filter transfer function appears in Figure 3.38. The feedback section is

$$B(D) = G(D) = 1 + .633D ,$$

with magnitude in Figure 3.39. The MS-WMF filter transfer function appears in Figure 3.40, illustrating the near all-pass character of this filter.

The MMSE is

$$\sigma^2_{MMSE-DFE} = \frac{N_0}{2} \frac{1}{\|p\|^2 \gamma_0} = \frac{.181}{1.81 \cdot .785} = .1274$$

and

$$\text{SNR}_{MMSE-DFE,U} = \frac{1 - .1274}{.1274} = 6.85 \ (8.4\text{dB}) .$$

Thus, the MMSE-DFE is only 1.6dB below the MFB on this channel. However, the ZF-DFE for this same channel would produce $\eta_0 = 1/\|p_c\|^2 = .5525$. In this case $\sigma^2_{ZFDFE} = .181/(.5525)1.81 = .181$ and the loss with respect to the MFB would be $\eta_0 = 2.6 \text{ dB}$, 1 dB lower than the MMSE-DFE for this channel.
Figure 3.39: Feedback filter transfer function for real-channel example.
Figure 3.40: MS-WMF transfer function for real-channel example.
It will in fact be possible to do yet better on this channel, using sequence detection, as discussed in Chapter 9, and/or by codes (Chapters 10 and 11). However, what originally appeared as a a stunning loss of 9.8 dB with the ZFE has now been reduced to a much smaller 1.6 dB.

The QAM example is also revisited here:

**EXAMPLE 3.6.2 (QAM - DFE)** Recall that the equivalent baseband pulse response samples were given by

\[ p_k = \frac{1}{\sqrt{T}} \left[ -\frac{1}{2} \left( 1 + \frac{j}{4} \right) - \frac{j}{2} \right] . \]  

(3.275)

The SNR\(_{MFB}\) is again 10 dB. Then,

\[ \hat{Q} = \begin{cases} -0.25D^{-2} + 0.625(-1 + j)D^{-1} + 1.5625(1 + .1) - 0.625(1 + j)D + 0.25jD^2 \\ 1.5625 \end{cases} . \]  

(3.276)

or

\[ \hat{Q} = \begin{cases} (1 - 0.451\angle^{1.632}D)(1 - 0.451\angle^{-0.612}D)(1 - 0.451\angle^{-1.632}D^{-1})(1 - 0.451\angle^{0.612}D^{-1}) \cdot \\ -j \\ 1.5625 \cdot 4 \cdot (0.451\angle^{-1.632})(0.451\angle^{0.612}) \end{cases} \]  

(3.277)

from which one extracts \( \gamma_0 = 0.7866 \) and \( G(D) = 1 - 0.4226(1 + j)D + 0.2034jD^2 \). The feedforward and feedback sections can be computed in a straightforward manner, as

\[ B(D) = G(D) = 1 - 0.4226(1 + j)D + 0.2034jD^2 \]  

(3.278)

and

\[ W(D) = 1.017 \cdot 1 - 0.226(1 - j)D^{-1} - 0.2034jD^{-2} \]  

(3.279)

The MSE is

\[ \sigma^2_{MMSE-DFE} = (0.15625) \cdot \frac{1}{0.7866(1.5625)} = 0.1271 \]  

(3.280)

and the corresponding SNR is

\[ \text{SNR}_{MMSE-DFE,U} = \frac{1}{0.1271} - 1 = 8.4 \text{dB} \]  

(3.281)

The loss is (also coincidentally) 1.6 dB with respect to the MFB and is 1.7 dB better than the MMSE-LE on this channel.

For the ZF-DFE, 

\[ Q(D) = (1 - 0.5\angle^{\pi/2}D)(1 - 0.5D)(1 - 0.5\angle^{-\pi/2}D^{-1})(1 - 0.5D^{-1}) \cdot \]  

(3.282)

\[ \frac{-j}{1.5625 \cdot 4 \cdot (0.5\angle^{-\pi/2})(5)} \]  

(3.283)

and thus \( \eta_0 = 0.6400 \) and \( P_c(D) = 1 - 0.5(1 + j)D + 0.25jD^2 \). The feedforward and feedback sections can be computed in a straightforward manner, as

\[ B(D) = P_c(D) = 1 - 0.5(1 + j)D + 0.25jD^2 \]  

(3.284)

and

\[ W(D) = 1.25 \]  

(3.285)

The output noise variance is

\[ \sigma^2_{ZFDFE} = (0.15625) \cdot \frac{1}{0.6400(1.5625)} = 0.1563 \]  

(3.286)

and the corresponding SNR is

\[ \text{SNR}_{ZFDFE} = \frac{1}{0.1563} = 6.4 = 8.0 \text{dB} \]  

(3.287)

The loss is 2.0 dB with respect to the MFB and is .4 dB worse than the MMSE-DFE.
3.7 Finite Length Equalizers

The previous developments of discrete-time equalization structures presumed that the equalizer could exist over an infinite interval in time. Equalization filters, $w_k$, are almost exclusively realized as finite-impulse-response (FIR) filters in practice. Usually these structures have better numerical properties than IIR structures. Even more importantly, adaptive equalizers (see Section 3.8 and Chapter 13) are often implemented with FIR structures for $w_k$ (and also for $b_k$, in the case of the adaptive DFE). This section studies the design of and the performance analysis of FIR equalizers.

Both the LE and the DFE cases are examined for both zero-forcing and least-squares criteria. This section describes design for the MMSE situation and then lets $\text{SNR} \to \infty$ to get the zero-forcing special case. Because of the finite length, the FIR zero-forcing equalizer cannot completely eliminate ISI in general.

3.7.1 FIR MMSE-LE

Returning to the FSE in Figure 3.5.4, the matched filtering operation will be performed digitally (at sampling rate $l/T$) and the FIR MMSE-LE will then incorporate both matched filtering and equalization. Perfect anti-alias filtering with gain $\sqrt{T}$ precedes the sampler and the combined pulse-response/anti-alias filter is $\tilde{p}(t)$. The assumption that $l$ is an integer simplifies the specification of matrices.

One way to view the oversampled channel output is as a set of $l$ parallel $T$-spaced subchannels whose pulse responses are offset by $T/l$ from each other as in Figure 3.7.1.

Each subchannel produces one of the $l$ phases per symbol period of the output set of samples at sampling rate $l/T$. Mathematically, it is convenient to represent such a system with vectors as shown below.

The channel output $y(t)$ is

$$y(t) = \sum_m x_m \cdot \tilde{p}(t - mT) + n(t),$$

(3.288)

which, if sampled at time instants $t = kT - \frac{iT}{l}$, $i = 0, ..., l - 1$, becomes

$$y(kT - \frac{iT}{l}) = \sum_{m=-\infty}^{\infty} x_m \cdot \tilde{p}(kT - \frac{iT}{l} - mT) + n(kT - \frac{iT}{l}).$$

(3.289)

The (per-dimensional) variance of the noise samples is $\frac{N_0}{T} \cdot l$ because the gain of the anti-alias filter, $\sqrt{T}$,
is absorbed into $p(t)$. The $t$ phases per symbol period of the oversampled $y(t)$ are contained in

$$y_k = \begin{bmatrix} y(kT) \\ y(kT - \frac{T}{T}) \\ \vdots \\ y(kT - \frac{l-1}{T}) \end{bmatrix}. \quad (3.290)$$

The vector $y_k$ can be written as

$$y_k = \sum_{m=-\infty}^{\infty} x_m \cdot p_{k-m} + n_k = \sum_{m=-\infty}^{\infty} x_{k-m} \cdot p_m + n_k, \quad (3.291)$$

where

$$p_k = \begin{bmatrix} \tilde{p}(kT) \\ \tilde{p}(kT - \frac{T}{T}) \\ \vdots \\ \tilde{p}(kT - \frac{l-1}{T}) \end{bmatrix} \quad \text{and} \quad n_k = \begin{bmatrix} n(kT) \\ n(kT - \frac{T}{T}) \\ \vdots \\ n(kT - \frac{l-1}{T}) \end{bmatrix}. \quad (3.292)$$

The response $\tilde{p}(t)$ is assumed to extend only over a finite time interval. In practice, this assumption requires any nonzero component of $p(t)$ outside of this time interval to be negligible. This time interval is $0 \leq t \leq \nu T$. Thus, $p_k = 0$ for $k < 0$, and for $k > \nu$. The sum in (3.289) becomes

$$y_k = \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-N_f+1} \end{bmatrix} + n_k. \quad (3.293)$$

Each row in (3.293) corresponds to a sample at the output of one of the filters in Figure 3.7.1. More generally, for $N_f$ successive $l$-tuples of samples of $y(t)$,

$$Y_k \triangleq \begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-N_f+1} \end{bmatrix} \quad (3.294)$$

$$= \begin{bmatrix} p_0 & p_1 & \cdots & p_{\nu} & 0 & 0 & \cdots & 0 \\ 0 & p_0 & p_1 & \cdots & p_{\nu} & 0 & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & p_0 & p_1 & \cdots & p_{\nu} & \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-N_f-\nu+1} \end{bmatrix} + \begin{bmatrix} n_k \\ n_{k-1} \\ \vdots \\ n_{k-N_f+1} \end{bmatrix}. \quad (3.295)$$

This text uses $P$ to denote the large $(N_f \cdot l) \times (N_f + \nu)$ matrix in (3.295), while $X_k$ denotes the data vector, and $N$ denotes the noise vector. Then, the oversampled vector representation of the channel is

$$Y_k = PX_k + N_k. \quad (3.296)$$

230
When \( l = \frac{n}{m} \) (a rational fraction), then (3.296) still holds with \( P \) an \([N_f \cdot \frac{m}{n}] \times (N_f + \nu)\) matrix that does not follow the form in (3.295), with each row possibly having a set of coefficients unrelated to the other rows.\(^{11}\)

An equalizer is applied to the sampled channel output vector \( Y_k \) by taking the inner product of an \( N_f \)-dimensional row vector of equalizer coefficients, \( w \), and \( Y_k \), so \( Z(D) = W(D)Y(D) \) can be written

\[
z_k = wY_k .
\]

(3.297)

For causality, the designer picks a channel-equalizer system delay \( \Delta \cdot T \) symbol periods,

\[
\Delta \approx \frac{\nu + N_f}{2} ,
\]

(3.298)

with the exact value being of little consequence unless the equalizer length \( N_f \) is very short. The delay is important in finite-length design because a non-causal filter cannot be implemented, and the delay allows time for the transmit data symbol to reach the receiver. With infinite-length filters, the need for such a delay does not enter the mathematics because the infinite-length filters are not realizable in any case, so that infinite-length analysis simply provides best bounds on performance. \( \Delta \) is approximately the sum of the channel and equalizer delays in symbol periods. The equalizer output error is then

\[
e_k = x_{k-\Delta} - z_k ,
\]

(3.299)

and the corresponding MSE is

\[
\sigma_{\text{MMSE-LE}}^2 = E \{ |e_k|^2 \} = E \{ e_k e_k^* \} = E \{ (x_{k-\Delta} - z_k) (x_{k-\Delta} - z_k)^* \} .
\]

(3.300)

Using the Orthogonality Principle of Appendix A.2, the MSE in (3.300) is minimized when

\[
E \{ e_k Y_k^* \} = 0 .
\]

(3.301)

In other words, the equalizer error signal will be minimized (in the mean square sense) when this error is uncorrelated with any channel output sample in the delay line of the FIR MMSE-LE. Thus

\[
E \{ x_{k-\Delta} Y_k^* \} - wE \{ Y_k Y_k^* \} = 0 .
\]

(3.302)

The two statistical correlation quantities in (3.302) have a very special significance (both \( y(t) \) and \( x_k \) are stationary):

**Definition 3.7.1 (Autocorrelation and Cross-Correlation Matrices)** The FIR MMSE autocorrelation matrix is

\[
R_{YY}^\Delta = E \{ Y_k Y_k^* \} / N ,
\]

(3.303)

while the FIR MMSE-LE cross correlation vector is defined by

\[
R_{Yx}^\Delta = E \{ Y_k x_{k-\Delta}^* \} / N ,
\]

(3.304)

where \( N = 1 \) for real and \( N = 2 \) for complex. Note \( R_{xY} = R_{Yx}^* \), and \( R_{YY} \) is not a function of \( \Delta \).

With the above definitions in (3.302), the designer obtains the MMSE-LE:

**Definition 3.7.2 (FIR MMSE-LE)** The FIR MMSE-LE for sampling rate \( l/T \), delay \( \Delta \), and of length \( N_f \) symbol periods has coefficients

\[
w = R_{Yx} R_{YY}^{-1} = R_{xY} R_{YY}^{-1} ,
\]

(3.305)

or equivalently,

\[
w^* = R_{YY}^{-1} R_{Yx} .
\]

(3.306)\(^{11}\)In this case, (3.288) is used to compute each row at the appropriate sampling instants. \( N_f \cdot \frac{m}{n} \) should also be an integer so that \( P \) is constant. Otherwise, \( P \) becomes a “time-varying” matrix \( P_k \).
In general, it may be of interest to derive more specific expressions for $\mathbf{R}_{YY}$ and $\mathbf{R}_{xY}$.

$$
\mathbf{R}_{xY} = E\{x_k - \Delta \mathbf{x}_k^*\} = E\{x_k - \Delta \mathbf{X}_k^*\} \mathbf{P}^* + E\{x_k - \Delta \mathbf{N}_k\} = \tilde{\mathbf{e}}_x \mathbf{P}^* + 0 \quad (3.307)
$$

$$
\mathbf{R}_{YY} = E\{\mathbf{Y}_k \mathbf{Y}_k^*\} = \mathbf{P} E\{\mathbf{X}_k \mathbf{X}_k^*\} \mathbf{P}^* + E\{\mathbf{N}_k \mathbf{N}_k^*\}

\quad = \tilde{\mathbf{e}}_x \mathbf{P} \mathbf{P}^* + \frac{N_0}{2} R_{nn} \quad (3.308)
$$

where the $(l \cdot \frac{N_0}{2})$-normalized) noise autocorrelation matrix $R_{nn}$ is equal to $I$ when the noise is white.

A convenient expression is

$$
\begin{bmatrix}
0 & \cdots & 0 \\
0 & \cdots & 0 \\
0 & \cdots & 0
\end{bmatrix} = \mathbf{1}_\Delta \mathbf{P}^* \quad (3.309)
$$

so that $\mathbf{1}_\Delta$ is an $N_f + \nu$-vector of 0’s and a 1 in the $(\Delta + 1)^{th}$ position. Then the FIR MMSE-LE becomes

$$
\mathbf{w} = \mathbf{1}_\Delta \mathbf{P}^* \left[ \mathbf{P} \mathbf{P}^* + \frac{1}{\text{SNR}} R_{nn} \right]^{-1} = \mathbf{1}_\Delta \left[ \mathbf{P} R_{nn}^{-1} \mathbf{P}^* + \frac{l}{\text{SNR}} \cdot I \right]^{-1} \mathbf{P} R_{nn}^{-1} \quad (3.310)
$$

The matrix inversion lemma of Appendix ?? is used in deriving the above equation.

The position of the nonzero elements in (3.309) depend on the choice of $\Delta$. The MMSE, $\sigma_{\text{MMSE-LE}}^2$, can be found by evaluating (3.300) with (3.305):

$$
\sigma_{\text{MMSE-LE}}^2 = E\left\{ x_k - \Delta x_k^* \mathbf{e}_k - x_k - \Delta \mathbf{X}_k^* \mathbf{w}^* - \mathbf{w} \mathbf{Y}_k \mathbf{X}_k^* \mathbf{w}^* + \mathbf{w} \mathbf{Y}_k \mathbf{Y}_k^* \mathbf{w}^* \right\} = \tilde{\mathbf{e}}_x - \mathbf{R}_{xY} \mathbf{w}^* - \mathbf{w} \mathbf{R}_{YY} \mathbf{w}^* \quad (3.311)
$$

$$
\quad = \tilde{\mathbf{e}}_x - w \mathbf{R}_{YY} \mathbf{w}^* \quad (3.312)
$$

$$
\quad = \tilde{\mathbf{e}}_x - \mathbf{w} \mathbf{R}_{YY} \mathbf{R}_{Yx} \quad (3.313)
$$

$$
\quad = \tilde{\mathbf{e}}_x - \mathbf{w} \mathbf{R}_{Yx} \quad ,
$$

With algebra, (3.318) is the same as

$$
\sigma_{\text{MMSE-LE}}^2 = \frac{l \cdot \frac{N_0}{2}}{l \cdot \frac{N_0}{2}} \tilde{\mathbf{e}}_x \mathbf{1}_\Delta^* \left[ \mathbf{P} \mathbf{R}_{nn}^{-1} \mathbf{P}^* + \frac{l}{\text{SNR}} \cdot I \right]^{-1} \mathbf{1}_\Delta \mathbf{Q} - \mathbf{1}_\Delta \quad (3.319)
$$

so that the best value of $\Delta$ (the position of the 1 in the vectors above) corresponds to the smallest diagonal element of the inverted (“Q-tilde”) matrix in (3.319) – this means that only one matrix need be inverted to obtain the correct $\Delta$ value as well as compute the equalizer settings. A homework problem develops some interesting relationships for $\mathbf{w}$ in terms of $\mathbf{P}$ and SNR. Thus, the SNR for the FIR MMSE-LE is

$$
\text{SNR}_{\text{MMSE-LE}} = \frac{\tilde{\mathbf{e}}_x}{\sigma_{\text{MMSE-LE}}^2} \quad ,
$$

and the corresponding unbiased SNR is

$$
\text{SNR}_{\text{MMSE-LE,U}} = \text{SNR}_{\text{MMSE-LE}} - 1 \quad (3.320)
$$

The loss with respect to the MFB is, again,

$$
\gamma_{\text{MMSE-LE}} = \frac{\text{SNR}_{\text{MFB}}}{\text{SNR}_{\text{MMSE-LE,U}}} \quad .
$$

(3.321)
3.7.2 FIR ZFE

One obtains the FIR ZFE by letting the SNR $\rightarrow \infty$ in the FIR MMSE, which alters $R_{YY}$ to

$$R_{YY} = PP^* \hat{\xi}_x$$ \hspace{1cm} (3.323)

and then $w$ remains as

$$w = R_xY R_{YY}^{-1}.$$ \hspace{1cm} (3.324)

Because the FIR equalizer may not be sufficiently long to cancel all ISI, the FIR ZFE may still have nonzero residual ISI. This ISI power is given by

$$\sigma_{\text{MMSE-ZFE}}^2 = \bar{\xi}_x - wR_{Yx}.$$ \hspace{1cm} (3.325)

However, (3.325) still ignores the enhanced noise at the FIR ZFE output.\(^{12}\) The power of this noise is easily found to be

$$\text{FIR ZFE noise power} = N_0^2 \cdot l \cdot \|w\|_2 R_{nn},$$ \hspace{1cm} (3.326)

making the SNR at the ZFE output

$$\text{SNR}_{ZFE} = \frac{\bar{\xi}_x - wR_{Yx} + N_0^2 l \|w\|_2 R_{nn}}{\xi_x - wR_{Yx}}.$$ \hspace{1cm} (3.327)

The ZFE is still unbiased in the finite-length case.\(^{13}\)

Equation (3.331) is true if $\Delta$ is a practical value and the finite-length ZFE has enough taps. The loss is the ratio of the SNR to SNR\(_{ZFE}\),

$$\gamma_{ZFE} = \frac{\text{SNR}_{MFB}}{\text{SNR}_{ZFE}}.$$ \hspace{1cm} (3.332)

3.7.3 example

For the earlier PAM example, one notes that sampling with $l = 1$ is sufficient to represent all signals. First, choose $N_f = 3$ and note that $\nu = 1$. Then

$$P = \begin{bmatrix} .9 & 1 & 0 & 0 \\ 0 & .9 & 1 & 0 \\ 0 & 0 & .9 & 1 \end{bmatrix}.$$ \hspace{1cm} (3.333)

With a choice of $\Delta = 2$, then

$$R_{Yx} = \hat{\xi}_x P \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

\hspace{1cm} (3.334)

\hspace{1cm} (3.335)

\(^{12}\)The notation $\|w\|_2 R_{nn}$ means $wR_{nn}w^*$.

\(^{13}\)The matrix $P^*(PP^*)^{-1}P$ is a “projection matrix” and $PP^*$ is full rank; therefore the entry of $x_{k-\Delta}$ passes directly (or is zeroed, in which case $\Delta$ needs to be changed).
and

\[ R_{YY} = \hat{\xi}_x \left( PP^* + \frac{1}{\text{SNR}} I \right) \quad (3.336) \]

\[ = \begin{bmatrix} 1.991 & .9 & 0 \\ .9 & 1.991 & .9 \\ 0 & .9 & 1.991 \end{bmatrix}, \quad (3.337) \]

The FIR MMSE is

\[ w^* = R_{YY}^{-1} R_Y x = \begin{bmatrix} .676 & -.384 & .174 \\ -.384 & .849 & -.384 \\ .174 & -.384 & .676 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ .9 \end{bmatrix} = \begin{bmatrix} -.23 \\ .51 \\ .22 \end{bmatrix} \quad (3.338) \]

Then

\[ \sigma_{MMSE-LE}^2 = \left( 1 - [-.23 \quad .51 \quad .22] \begin{bmatrix} 0 \\ 1 \\ .9 \end{bmatrix} \right) = .294 \quad (3.339) \]

The SNR is computed to be

\[ \text{SNR}_{MMSE-LE,U} = \frac{1}{.294} - 1 = 2.4 (3.8 \text{ dB}) \quad (3.340) \]

which is 6.2dB below MFB performance and 1.9dB worse than the infinite length MMSE-LE for this channel.

With sufficiently large number of taps for this channel, the infinite length performance level can be attained. This performance is plotted as versus the number of equalizer taps in Figure 234.
Figure 3.42: FIR equalizer performance for $1 + 0.9D^{-1}$ versus number of equalizer taps.

3.42. Clearly, 15 taps are sufficient for infinite-length performance.

**EXAMPLE 3.7.1** For the earlier complex QAM example, sampling with $l = 1$ is sufficient to represent all signals. First, choose $N_f = 4$ and note that $\nu = 2$. Then

$$P = \begin{bmatrix}
-0.5 & 1 + j/4 & -j/2 & 0 & 0 & 0 \\
0 & -0.5 & 1 + j/4 & -j/2 & 0 & 0 \\
0 & 0 & -0.5 & 1 + j/4 & -j/2 & 0 \\
0 & 0 & 0 & -0.5 & 1 + j/4 & -j/2 \\
\end{bmatrix}.$$  (3.341)

The matrix $(P^*P + 0.15625 \cdot I)^{-1}$ has the same smallest element 1.3578 for both $\Delta = 2, 3$, so choosing $\Delta = 2$ will not unnecessarily increase system delay. With a choice of $\Delta = 2$, then

$$R_{Yx} = \bar{\xi}_x P = \begin{bmatrix}
0 \\
0 \\
1 \\
0 \\
0 \\
\end{bmatrix}.$$

$$= \begin{bmatrix}
-0.5j \\
1 + j/4 \\
-0.5 \\
\end{bmatrix},$$  (3.343)
The FIR MMSE is

\[ w^* = R_{Y \hat{X}}^{-1} R_{Y \hat{x}} = \begin{bmatrix} 0.2570 - 0.0422j \\ 0.7313 - 0.0948j \\ -0.1182 - 0.2982j \\ -0.1376 + 0.0409j \end{bmatrix} \]

(3.345)

Then

\[ \sigma_{MMSE-LE}^2 = (1 - w R_{Y \hat{x}}) \]

(3.346)

\[ = .2121 \]  

(3.347)

The SNR is computed to be

\[ SNR_{MMSE-LE,U} = \frac{1}{.2121} - 1 = 3.714 \text{ (5.7 dB)} \]

(3.348)

which is 4.3 dB below MFB performance and 2.1 dB worse than the infinite-length MMSE-LE for this channel. With sufficiently large number of taps for this channel, the infinite-length performance level can be attained.

### 3.7.4 FIR MMSE-DFE

The FIR DFE case is similar to the feed-forward equalizers just discussed, except that we now augment the fractionally spaced feed-forward section with a symbol-spaced feedback section. The MSE for the DFE case, as long as \( w \) and \( b \) are sufficiently long, is

\[ MSE = E \left\{ |x_{k-\Delta} - w Y_k + b x_{k-\Delta-1}|^2 \right\} , \]

(3.349)

where \( b \) is the vector of coefficients for the feedback FIR filter

\[ b \overset{\Delta}{=} \begin{bmatrix} b_1 \ b_2 \ldots \ b_{N_b} \end{bmatrix} , \]

(3.350)

and \( x_{k-\Delta-1} \) is the vector of data symbols in the feedback path. It is mathematically convenient to define an augmented response vector for the DFE as

\[ \tilde{w} \overset{\Delta}{=} \begin{bmatrix} w^\top - b \end{bmatrix} , \]

(3.351)

and a corresponding augmented DFE input vector as

\[ \tilde{Y}_k \overset{\Delta}{=} \begin{bmatrix} Y_k \\ x_{k-\Delta-1} \end{bmatrix} . \]

(3.352)

Then, (3.349) becomes

\[ MSE = E \left\{ |x_{k-\Delta} - \tilde{w} \tilde{Y}_k|^2 \right\} . \]

(3.353)

The solution, paralleling (3.304) - (3.306), is determined using the following two definitions:
Definition 3.7.3 (FIR MMSE-DFE Autocorrelation and Cross-Correlation Matrices)

The FIR MMSE-DFE autocorrelation matrix is defined as

\[
R_{\hat{Y}\hat{Y}} = \frac{1}{N} \left\{ \hat{Y}_k \hat{Y}_k^* \right\} / N \tag{3.354}
\]

\[
= \begin{bmatrix}
R_{Y\bar{Y}} \\
E[x_k - \Delta - 1 Y_k^*]
\end{bmatrix}
\tag{3.355}
\]

\[
= \begin{bmatrix}
\tilde{\bar{E}}_x \left( PP^* + l \frac{R_{nn}}{SNR} \right) \\
\tilde{\bar{E}}_x P J_{\Delta} \nabla
\end{bmatrix}
\tag{3.356}
\]

where \( J_{\Delta} \) is an \((N_f + \nu) \times N_b \) matrix of 0’s and 1’s, which has the upper \( \Delta + 1 \) rows zeroed and an identity matrix of dimension \( \min(N_b, N_f + \nu - \Delta - 1) \) with zeros to the right (when \( N_f + \nu - \Delta - 1 < N_b \)), zeros below (when \( N_f + \nu - \Delta - 1 > N_b \)), or no zeros to the right or below exactly fitting in the bottom of \( J_{\Delta} \) (when \( N_f + \nu - \Delta - 1 = N_b \)).

The corresponding FIR MMSE-DFE cross-correlation vector is

\[
R_{\hat{Y}x} \triangleq \frac{1}{N} \left\{ \hat{Y}_k x_{k-\Delta} \right\} \tag{3.357}
\]

\[
= \begin{bmatrix}
R_{Yx} \\
0
\end{bmatrix}
\tag{3.358}
\]

\[
= \begin{bmatrix}
\tilde{\bar{E}}_x P 1_{\Delta} \\
0
\end{bmatrix}
\tag{3.359}
\]

where, again, \( N = 1 \) for real signals and \( N = 2 \) for complex signals.

The FIR MMSE-DFE for sampling rate \( l/T \), delay \( \Delta \), and of length \( N_f \) and \( N_b \) has coefficients

\[
\hat{w} = R_{x\hat{Y}} R_{\hat{Y}\hat{Y}}^{-1} \tag{3.360}
\]

Equation (3.360) can be rewritten in detail as

\[
[\hat{w} \cdot - b] \cdot \tilde{\bar{E}}_x \cdot \begin{bmatrix}
PP^* + \frac{l}{SNR} R_{nn} \\
J_{\Delta}^* P^* \\
J_{\Delta}^* P^* I_{N_b} \\
1 \cdot 1_{\Delta} P^*:0
\end{bmatrix} = \tilde{\bar{E}}_x \cdot 1_{\Delta} P^*:0
\tag{3.361}
\]

which reduces to the pair of equations

\[
w \left( PP^* + \frac{l}{SNR} R_{nn} \right) - b J_{\Delta}^* P^* = 1_{\Delta} P^* \tag{3.362}
\]

\[
w P J_{\Delta} - b = 0 \tag{3.363}
\]

Then

\[
b = w P J_{\Delta} \tag{3.364}
\]

and thus

\[
w \left( PP^* - P J_{\Delta} J_{\Delta}^* P^* + \frac{l}{SNR} R_{nn} \right) = 1_{\Delta} P^* \tag{3.365}
\]

or

\[
w = 1_{\Delta} P^* \left( PP^* - P J_{\Delta} J_{\Delta}^* P^* + \frac{l}{SNR} R_{nn} \right)^{-1} \tag{3.366}
\]

Then

\[
b = 1_{\Delta} P^* \left( PP^* - P J_{\Delta} J_{\Delta}^* P^* + \frac{l}{SNR} R_{nn} \right)^{-1} P J_{\Delta} \tag{3.367}
\]
The MMSE is then
\[ \sigma^2_{MMSE-DFE} = \tilde{\varepsilon}_x - \hat{\omega} R_{Y_x} \]
\[ = \tilde{\varepsilon}_x - w R_{Y_x} \]
\[ = \tilde{\varepsilon}_x \left( 1 - \Delta^* P^* \left( P P^* - P J \Delta^* P^* + \frac{l}{SNR_{Rnn}} \right)^{-1} \right) P \Delta \), (3.368)\]
which is a function to be minimized over \( \Delta \). Thus, the SNR for the unbiased FIR MMSE-DFE is
\[ SNR_{MMSE-DFE,U} = \frac{\tilde{\varepsilon}_x}{\sigma^2_{MMSE-DFE}} - 1 = \frac{\hat{\omega} R_{Y_x}}{\tilde{\varepsilon}_x - \hat{\omega} R_{Y_x}} , (3.371)\]
and the loss is again
\[ \gamma_{MMSE-DFE} = \frac{SNR_{MFB}}{SNR_{MMSE-DFE,U}} . \] (3.372)

**EXAMPLE 3.7.2 (MMSE-DFEs, PAM and QAM)** For the earlier example with \( l = 1 \), \( N_f = 2 \), and \( N_b = 1 \) and \( \nu = 1 \), we will also choose \( \Delta = 1 \). Then
\[ P = \begin{bmatrix} .9 & 1 & 0 \\ 0 & .9 & 1 \end{bmatrix} , \] (3.373)
\[ J \Delta = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} , \] (3.374)
\[ 1 \Delta = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} , \] (3.375)
and
\[ R_{Y Y} = \begin{bmatrix} 1.991 & .9 \\ .9 & 1.991 \end{bmatrix} . \] (3.376)

Then
\[ w = [0 \ 1 \ 0] \left[ \begin{bmatrix} .9 & 0 \\ 1 & .9 \\ 0 & 1 \end{bmatrix} \right] \left\{ \begin{bmatrix} 1.991 & .9 \\ .9 & 1.991 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}^{-1} \] (3.377)
\[ = [1.9] \left[ \begin{bmatrix} 1.991 & .9 \\ .9 & 0.991 \end{bmatrix} \right]^{-1} \] (3.378)
\[ = [.1556 .7668] \] (3.379)
\[ b = w P J \Delta \] (3.380)
\[ = [.1556 .7668] \begin{bmatrix} .9 & 1 & 0 \\ 0 & .9 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \] (3.381)
\[ = .7668 \] (3.382)

Then
\[ \sigma^2_{MMSE-DFE} = 1 - [.16 .76] \begin{bmatrix} 1 \\ .9 \end{bmatrix} = .157 \) (3.383)

The SNR is computed to be
\[ SNR_{MMSE-DFE,U} = \frac{1 - .157}{.157} = 5.4 \ (7.3 \text{dB}) , \] (3.384)
which is 2.7 dB below MFB performance, but 2.8 dB better than the **infinite-length** MMSE-LE for this channel! The loss with respect to the infinite-length MMSE-DFE is 1.1 dB. A picture of the FIR MMSE-DFE is shown in Figure 3.43.

![FIR MMSE-DFE Example.](image)

Figure 3.43: FIR MMSE-DFE Example.

With sufficiently large number of taps for this channel, the infinite-length performance level can be attained. This performance is plotted versus the number of feed-forward taps (only one feedback tap is necessary for infinite-length performance) in Figure 3.44. In this case, 7 feed-forward and 1 feedback taps are necessary for infinite-length performance. Thus, the finite-length DFE not only outperforms the finite or infinite-length LE, it uses less taps (less complexity) also. For the QAM example \( l = 1 \), \( N_f = 2 \), \( N_b = 2 \), and \( \nu = 2 \), we also choose \( \Delta = 1 \). Actually this channel will need more taps to do well with the DFE structure, but we can still choose these values and proceed. Then

\[
P = \begin{bmatrix}
-0.5 & 1 + \frac{j}{4} & -\frac{j}{2} & 0 \\
0 & -0.5 & 1 + \frac{j}{4} & -\frac{j}{2}
\end{bmatrix},
\]  
(3.385)

\[
J_\Delta = \begin{bmatrix}
0 & 0 \\
0 & 0 \\
1 & 0 \\
0 & 1
\end{bmatrix},
\]  
(3.386)

\[
1_\Delta = \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix},
\]  
(3.387)

and

\[
R_{YY} = \begin{bmatrix}
1.7188 & -0.6250 - 0.6250j \\
-0.6250 - 0.6250j & 1.7188
\end{bmatrix},
\]  
(3.388)

Then

\[
w = \begin{bmatrix}
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
-0.5 & 0 \\
1 - \frac{j}{4} & -0.5 \\
\frac{j}{2} & 1 - \frac{j}{4} \\
0 & \frac{j}{2}
\end{bmatrix},
\]  
(3.389)

\[
\left\{ \begin{bmatrix}
1.7188 & -0.6250 - 0.6250j \\
-0.6250 - 0.6250j & 1.7188
\end{bmatrix} - \begin{bmatrix}
\frac{1}{4} & -\frac{1}{8} - \frac{j}{2} \\
-\frac{1}{8} + \frac{j}{2} & 1.3125
\end{bmatrix} \right\}^{-1}
\]  

239
\[ \begin{align*}
&= \left[ 1 - \frac{j}{2} - 0.5 \right] \begin{bmatrix}
0.2277 & 1.5103 + j0.3776 \\
1.5103 - j0.3776 & 4.4366
\end{bmatrix} \\
&= \begin{bmatrix}
0.4720 - j1.1180 & -0.6136 \\
-0.5103 & -j0.3776
\end{bmatrix}
\end{align*} \tag{3.390}

\[ b = wP \Delta \tag{3.391} \]

\[ \begin{align*}
&= \begin{bmatrix}
0.4720 - j1.1180 & -0.6136 \\
-0.5103 & -j0.3776
\end{bmatrix}
\begin{bmatrix}
-0.5 & 1 + j/4 & -j/2 & 0 \\
0 & -0.5 & 1 + j/4 & -j/2
\end{bmatrix}
\begin{bmatrix}
0 & 0 \\
1 & 0 \\
0 & 1
\end{bmatrix} \\
&= \begin{bmatrix}
-0.6726 - j3.3894 & 0.3608j
\end{bmatrix}
\end{align*} \tag{3.392}

Then
\[ \sigma^2_{\text{MMSE-DFE}} = 0.1917 . \tag{3.395} \]

The SNR is computed to be
\[ \text{SNR}_{\text{MMSE-DFE,U}} = 1 - \frac{0.1917}{0.1917} = 6.23 \text{dB} , \tag{3.396} \]

which is 3.77 dB below MFB performance, and not very good on this channel! The reader may evaluate various filter lengths and delays to find a best use of 3, 4, 5, and 6 parameters on this channel.

Figure 3.44: FIR MMSE-DFE performance for $1 + 0.9D^{-1}$ versus number of feed-forward equalizer taps.
3.7.4.0.1 FIR ZF-DFE

One obtains the FIR ZF-DFE by letting the SNR $\to \infty$ in the FIR MMSE-DFE, which alters $R_{\tilde{Y}\tilde{Y}}$ to

$$R_{\tilde{Y}\tilde{Y}} = \begin{bmatrix} \tilde{\xi}_x \tilde{P}^{*} & \tilde{\xi}_x \tilde{P}_{J\Delta} \\ \tilde{\xi}_x \tilde{J}_{\Delta} P^{*} & \tilde{\xi}_x \cdot I_{N_b} \end{bmatrix}$$

(3.397)

and then $\tilde{w}$ remains as

$$\tilde{w} = R_{x\tilde{Y}} R_{\tilde{Y}\tilde{Y}}^{-1}.$$  

(3.398)

Because the FIR equalizer may not be sufficiently long to cancel all ISI, the FIR ZF-DFE may still have nonzero residual ISI. This ISI power is given by

$$\sigma_{MMSE-DFE}^2 = \bar{E}_x - \tilde{w} R_{\tilde{Y}x}.$$  

(3.399)

However, (3.399) still ignores the enhanced noise at the $\tilde{w}_k$ filter output. The power of this noise is easily found to be

FIR ZFDFE noise variance $= \frac{N_0}{2} \cdot l \cdot \|w\|^2,$

(3.400)

making the SNR at the FIR ZF-DFE output

$$\text{SNR}_{ZF-DFE} = \frac{\bar{E}_x}{\bar{E}_x - \tilde{w} R_{\tilde{Y}x} + \frac{\bar{E}_x}{\text{SNR}} l \cdot \|w\|^2}.$$  

(3.401)

The filter is $w$ in (3.400) and (3.401), not $\tilde{w}$ because only the feed-forward section filters noise. The loss is:

$$\gamma_{ZF-DFE} = \frac{\text{SNR}_{MFB}}{\text{SNR}_{ZF-DFE}}.$$  

(3.402)

3.7.5 An Alternative Approach to the DFE

While the above approach directly computes the settings for the FIR DFE, it yields less insight into the internal structure of the DFE than did the infinite-length structures investigated in Section 3.6, particularly the spectral (canonical) factorization into causal and anti-causal ISI components is not explicit. This subsection provides such an alternative caused by the finite-length filters. This is because finite-length filters inherently correspond to non-stationary processing.\(^\text{14}\)

The vector of inputs at time $k$ to the feed-forward filter is again $Y_k$, and the corresponding vector of filter coefficients is $w$, so the feed-forward filter output is $z_k = w Y_k$. Continuing in the same fashion, the DFE error signal is then described by

$$e_k = b X_k(\Delta) - w Y_k,$$

(3.403)

where $x^*_{k-N_b} = [x^*_k \ldots x^*_{k-N_b}]$, and $b$ is now slightly altered to be monic and causal

$$b \overset{\Delta}{=} \begin{bmatrix} 1 & b_2 & \ldots & b_{N_b} \end{bmatrix}.$$  

(3.404)

The mean-square of this error is to be minimized over $b$ and $w$. Signals will again be presumed to be complex, but all developments here simplify to the one-dimensional real case directly, although it is important to remember to divide any complex signal’s variance by 2 to get the energy per real dimension. The SNR equals $\bar{E}_x / \frac{N_0}{2} = \bar{E}_x / N_0$ in either case.

\(^\text{14}\)A more perfect analogy between finite-length and infinite-length DFE’s occurs in Chapter 5, where the best finite-length DFE’s are actually periodic over a packet period corresponding to the length of the feed-forward filter.
3.7.5.0.1 Optimizing the feed-forward and feedback filters  

For any fixed $b$ in (3.403), the cross-correlation between the error and the vector of channel outputs $Y_k$ should be zero to minimize MSE,

$$E[e_k Y_k^*] = 0.$$  \hspace{1cm} (3.405)

So,

$$w R_{YY} = b R_{XY}(\Delta),$$  \hspace{1cm} (3.406)

and

$$R_{XY}(\Delta) = E \left\{ X_k(\Delta) \left[ x_k^* x_{k-1}^* \ldots x_{k-N_f-\nu+1}^* \right] P^* \right\}$$  \hspace{1cm} (3.407)

where the matrix $J_\Delta$ has the first $\Delta$ columns all zeros, then an (up to) $N_b \times N_b$ identity matrix at the top of the up to $N_b$ columns, and zeros in any row entries below that identity, and possibly zeroed columns following the identity if $N_f + \nu - 1 > N_b$. The following matlab commands produce $J_\Delta$, using for instance $N_f = 8$, $N_b = 5$, $\Delta = 3$, and $\nu = 5$,

```
>> size=min(Delta,Nb);
>> Jdelta=[ zeros(size,size) eye(size) zeros(size, max(Nf+nu-2*size,0))
zeros(max(Nb-Delta,0),Nf+nu)]
```

$$Jdelta =$$

```
0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 0 0 1
```

or for $N_f = 8$, $N_b = 2$, $\Delta = 3$, and $\nu = 5$,

```
>> >> size=min(Delta,Nb)
size = 2
>> Jdelta=[ zeros(size,size) eye(size) zeros(size, max(Nf+nu-2*size,0))
zeros(max(Nb-Delta,0),Nf+nu)]
```

$$Jdelta =$$

```
0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0 0
```

The MSE for this fixed value of $b$ then becomes

$$\sigma^2(\Delta) = b^* \left( \bar{E} x_I - R_{XY}(\Delta) R_{YY}^{-1} R_{YY} X(\Delta) \right) b$$  \hspace{1cm} (3.409)

$$= b^* R_{XY}^\perp Y(\Delta) b$$  \hspace{1cm} (3.410)

$$= b^* \left( \bar{E} x_I N_b - J_\Delta^* P^* \left( PP^* + \frac{l}{SNR} R_{NN} \right)^{-1} P J_\Delta(\Delta) \right) b$$  \hspace{1cm} (3.411)

$$= l \cdot \frac{N_0}{2} \left\{ b^* J_\Delta^* \left( P^* R_{NN}^{-1} P + \frac{l}{SNR} I \right)^{-1} P J_\Delta(\Delta) J_\Delta b \right\}$$  \hspace{1cm} (3.412)

and $R_{XY}^\perp(\Delta) \triangleq \bar{E} x_I - R_{XY}(\Delta) R_{YY}^{-1} R_{YY} X(\Delta)$, the autocorrelation matrix corresponding to the MMSE vector in estimating $X(\Delta)$ from $Y_k$. This expression is the equivalent of (3.221). That is $R_{XY}^\perp(\Delta)$ is the autocorrelation matrix for the error sequence of length $N_b$ that is associated with a
“matrix” MMSE-LE. The solution then requires factorization of the inner matrix into canonical factors, which is executed with Cholesky factorization for finite matrices.

By defining
\[
\tilde{Q}(\Delta) = \left\{ \tilde{J}_\Delta \left( P^*P + \frac{l}{\text{SNR}}I \right)^{-1} \tilde{J}_\Delta \right\}^{-1},
\]
(3.413)
this matrix is equivalent to (3.222), except for the “annoying” \( \tilde{J}_\Delta \) matrices that become identities as \( N_f \) and \( N_b \) become infinite, then leaving \( \tilde{Q} \) as essentially the autocorrelation of the channel output. The \( \tilde{J}_\Delta \) factors, however, cannot be ignored in the finite-length case. Canonical factorization of \( \tilde{Q} \) proceeds according to
\[
\sigma^2(\Delta) = bG_\Delta^{-1}S_\Delta G_\Delta^{-*}b^*,
\]
(3.414)
which is minimized when
\[
b = g(\Delta),
\]
(3.415)
the top row of the upper-triangular matrix \( G_\Delta \). The MMSE is thus obtained by computing Cholesky factorizations of \( \tilde{Q} \) for all reasonable values of \( \Delta \) and then setting \( b = g(\Delta) \). Then
\[
\sigma^2(\Delta) = S_0(\Delta).
\]
(3.416)

From previous developments, as the lengths of filters go to infinity, any value of \( \Delta \) works and also \( S_0 \rightarrow \gamma_0 \frac{N_0}{T}/\bar{E}_x \) to ensure the infinite-length MMSE-DFE solution of Section 3.6.

The feed-forward filter then becomes
\[
w = bR_{XY}(\Delta)R^{-1}_{YY}
\]
(3.417)
\[
= g(\Delta)\tilde{J}_\Delta P^* \left( PP^* + \frac{l}{\text{SNR}}I \right)^{-1}
\]
(3.418)
\[
= g(\Delta)\tilde{J}_\Delta \left( PP^* + \frac{l}{\text{SNR}}I \right)^{-1}
\]
(3.419)
\[
\text{matched filter}
\]
\[
\text{feedforward filter}
\]
which can be interpreted as a matched filter followed by a feed-forward filter that becomes \( 1/G^*(D^{-*}) \) as its length goes to infinity. However, the result that the feed-forward filter is an anti-causal factor of the canonical factorization does not follow for finite length. Chapter 5 will find a situation that is an exact match and for which the feed-forward filter is an inverse of a canonical factor, but this requires the DFE filters to become periodic in a period equal to the number of taps of the feed-forward filter (plus an excess bandwidth factor).

The SNR is as always
\[
\text{SNR}_{\text{MMSE-DFE, U}} = \frac{\bar{E}_x}{S_0(\Delta)} - 1.
\]
(3.420)

Bias can be removed by scaling the decision-element input by the ratio of \( \text{SNR}_{\text{MMSE-DFE, U}}/\text{SNR}_{\text{MMSE-DFE, U}} \), thus increasing its variance by \( (\text{SNR}_{\text{MMSE-DFE, U}}/\text{SNR}_{\text{MMSE-DFE, U}})^2 \).

### 3.7.5.1 Finite-length Noise-Predictive DFE

Problem ?? introduces the “noise-predictive” form of the DFE, which is repeated here in Figure 3.7.5.1.
In this figure, the error sequence is feedback instead of decisions. The filter essentially tries to predict the noise in the feedforward filter output and then cancel this noise, whence the name “noise-predictive” DFE. Correct solution to the infinite-length MMSE filter Problem 3.8 will produce that the filter \( B(D) \) remains equal to the same \( G(D) \) found for the infinite-length MMSE-DFE. The filter \( U(D) \) becomes the MMSE-LE so that \( Z(D) \) has no ISI, but a strongly correlated (and enhanced) noise. The predictor then reduces the noise to a white error sequence. If \( W(D) = \frac{1}{\gamma_0 \| p \|} G_{-1}(D) \) of the normal MMSE-DFE, then it can be shown also (see Problem 3.8) that

\[
U(D) = \frac{W(D)}{G(D)}.
\]

The MMSE, all SNR’s, and biasing/unbiasing remain the same.

An analogous situation occurs for the finite-length case, and it will be convenient notationally to say that the number of taps in the feedforward filter \( u \) is \((N_f - \nu)\). Clearly if \( \nu \) is fixed as always, then any number of taps (greater than \( \nu \) can be investigated without loss of generality with this early notational abuse. Clearly by abusing \( N_f \) (which is not the number of taps), ANY positive number of taps in \( u \) can be constructed without loss of generality for any value of \( \nu \geq 0 \). In this case, the error signal can be written as

\[
e_k = b_k X_k(\Delta) - b Z
\]

where

\[
Z_k = \begin{bmatrix} z_k \\ \vdots \\ z_{k-\nu} \end{bmatrix}.
\]

Then,

\[
Z_k = U Y_k
\]

where

\[
U = \begin{bmatrix} u & 0 & \ldots & 0 \\ 0 & u & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & u \end{bmatrix}
\]

and

\[
Y_k = \begin{bmatrix} y_k \\ \vdots \\ y_{k-N_f+1} \end{bmatrix}.
\]
By defining the \((N_f)l\)-tap filter
\[
    w \triangleq bU,
\]
then the error becomes
\[
e_k = b_k X_k(\Delta) - wY,
\]
which is the same error as in the alternate viewpoint of the finite-length DFE earlier in this section. Thus, solving for the \(b\) and \(w\) of the conventional finite-length MMSE-DFE can be followed by the step of solving Equation (3.427) for \(U\) when \(b\) and \(w\) are known. However, it follows directly then from Equation (3.406) that
\[
u = R_{XY}(\Delta)R_{YY}^{-1},
\]
so following the infinite-case form, \(u\) is the finite-length MMSE-LE with \((N_f - \nu)l\) taps. The only difference from the infinite-length case is the change in length.

### 3.7.6 The Stanford DFE Program

A matlab program has been created, used, and refined by the students of EE379A over a period of many years. The program has the following call command:

```matlab
function (SNR, wt) = dfe (l,p,nff, nbb, delay, Ex, noise);
```

- \(l\) = oversampling factor
- \(p\) = pulse response, oversampled at \(l\) (size)
- \(nff\) = number of feed-forward taps
- \(nbb\) = number of feedback taps
- \(delay\) = delay of system \(N_{ff} + \text{length of } p - nbb\); if \(delay = -1\), then choose the best delay
- \(Ex\) = average energy of signals
- \(noise\) = autocorrelation vector (size \(l\times nff\)) (NOTE: noise is assumed to be stationary). For white noise, this vector is simply \([\sigma^2 0 ... 0]\).
- \(SNR\) = equalizer SNR, unbiased and in dB
- \(wt\) = equalizer coefficients

The student may use this program to a substantial advantage in avoiding tedious matrix calculations. The program has come to be used throughout the industry to compute/project equalizer performance (setting \(nbb=0\) also provides a linear equalizer). The reader is cautioned against the use of a number of rules of thumb (like “DFE SNR is the SNR(f) at the middle of the band”) used by various who call themselves experts and often over-estimate the DFE performance using such formulas. Difficult transmission channels may require large numbers of taps and considerable experimentation to find the best settings.

```matlab
function [dfseSNR, w_t, opt_delay]=dfsecolorsnr(l,p,nff,nbb,delay,Ex,noise);
```

The student may use this program to a substantial advantage in avoiding tedious matrix calculations. The program has come to be used throughout the industry to compute/project equalizer performance (setting \(nbb=0\) also provides a linear equalizer). The reader is cautioned against the use of a number of rules of thumb (like “DFE SNR is the SNR(f) at the middle of the band”) used by various who call themselves experts and often over-estimate the DFE performance using such formulas. Difficult transmission channels may require large numbers of taps and considerable experimentation to find the best settings.
% NOTE: noise is assumed to be stationary
%
% outputs:
% dfseSNR = equalizer SNR, unbiased in dB
% w_t = equalizer coefficients [w -b]
% opt_delay = optimal delay found if delay =-1 option used.
% otherwise, returns delay value passed to function
% created 4/96;
% ---------------------------------------------------------------

size = length(p);
u = ceil(size/l)-1;
p = [p zeros(1,(nu+1)*l-size)];

% error check
if nff<=0
    error('number of feed-forward taps > 0');
end
if delay > (nff+nu-1-nbb)
    error('delay must be <= (nff+(length of p)-2-nbb)');
elseif delay < -1
    error('delay must be >= -1');
elseif delay == -1
    delay = 0:1:nff+nu-1-nbb;
end

ptmp = [p_0 p_1 ... p_nu] where p_i=[p(i*l) p(i*l-1)... p((i-1)*l+1)
ptmp(1:l,1) = [p(1); zeros(l-1,1)];
for i=1:nu
    ptmp(1:l,i+1) = conj((p(i*l+1:-1:(i-1)*l+2))');
end

%form matrix P, vector channel matrix
P = zeros(nff*l+nbb,nff+nu);
for i=1:nff,
P(((i-1)*l+1):(i*l),i:(i+nu)) = ptmp;
end

%precompute Rn matrix - constant for all delays
Rn = zeros(nff*l+nbb);
Rn(1:nff+1,1:nff+1) = toeplitz(noise);
dfseSNR = -100;
P_init = P;
%loop over all possible delays
for d = delay,
P = P_init;
P(nff*l+1:nff*l+nbb,d+2:d+1+nbb) = eye(nbb);
%P
    temp= zeros(1,nff+nu);
    temp(d+1)=1;
%construct matrices
Ry = Ex*P*P' + Rn;
Rxy = Ex*temp*P';
new_w_t = Rxy*inv(Ry);
sigma_dfse = Ex - real(new_w_t*Rxy');
new_dfseSNR = 10*log10(Ex/sigma_dfse - 1);
% save setting of this delay if best performance so far
if new_dfseSNR >= dfseSNR
    w_t = new_w_t;
    dfseSNR = new_dfseSNR;
    opt_delay = d;
end
end

3.7.7 Error Propagation in the DFE

To this point, this chapter has ignored the effect of decision errors in the feedback section of the DFE. At low error rates, say $10^{-5}$ or below, this is reasonable. There is, however, an accurate way to compute the effect of decision feedback errors, although enormous amounts of computational power may be necessary (much more than that required to simulate the DFE and measure error rate increase). At high error rates of $10^{-3}$ and above, error propagation can lead to several dB of loss. In coded systems (see Chapters 7 and 8), the inner channel (DFE) may have an error rate that is unacceptably high that is later reduced in a decoder for the applied code. Thus, it is common for low decision-error rates to occur in a coded DFE system. The Precoders of Section 3.8 can be used to eliminate the error propagation problem, but this requires that the channel be known in the transmitter – which is not always possible, especially in transmission systems with significant channel variation, i.e., digital mobile telephony. An analysis for the infinite-length equalizer has not yet been invented, thus the analysis here applies only to FIR DFE’s with finite $N_b$.

By again denoting the equalized pulse response as $v_k$ (after removal of any bias), and assuming that any error signal, $e_k$, in the DFE output can is uncorrelated with the symbol of interest, $x_k$, this error signal can be decomposed into 4 constituent signals

1. precursor ISI - $e_{pr,k} = \sum_{i=-\infty}^{-1} v_i x_{k-i}$
2. postcursor ISI - $e_{po,k} = \sum_{i=0}^{N_b+1} v_i x_{k-i}$
3. filtered noise - $e_{n,k} = \sum_{i=-\infty}^{\infty} w_i n_{k-i}$
4. feedback errors - $e_{f,k} = \sum_{i=1}^{N_b} v_i (x_{k-i} - \hat{x}_{k-i})$

The sum of the first 3 signals constitute the MMSE with power $\sigma_{MMSE-DFE}^2$, which can be computed according to previous results. The last signal is often written

$$e_{f,k} = \sum_{i=1}^{N_b} v_i \cdot \epsilon_{k-i}$$ (3.430)

where $\epsilon_k \triangleq x_k - \hat{x}_k$. This last distortion component has a discrete distribution with $(2^M - 1)^{N_b}$ possible points.

The error event vector, $\epsilon_k$, is

$$\epsilon_k \triangleq [\epsilon_{k-N_b+1} \; \epsilon_{k-N_b+2} \; \ldots \; \epsilon_k]$$ (3.431)

Given $\epsilon_{k-1}$, there are only $2^M - 1$ possible values for $\epsilon_k$. Equivalently, the evolution of an error event, can be described by a finite-state machine with $(2^M - 1)^{N_b}$ states, each corresponding to one of the $(2^M - 1)^{N_b}$ possible length-$N_b$ error events. Such a finite state machine is shown in
Figure 3.46: Finite State Machine for $N_b = 2$ and $M = 2$ in evaluating DFE Error Propagation

Figure 3.7.7. The probability that $\epsilon_k$ takes on a specific value, or that the state transition diagram goes to the corresponding state at time $k$, is (denoting the corresponding new entry in $\epsilon_k$ as $\epsilon_k$)

$$P_{\epsilon_k/\epsilon_{k-1}} = Q \left[ \frac{d_{\min} - |\epsilon_{f,k}(\epsilon_{k-1})|}{2\sigma_{MMSE-DFE}} \right].$$

(3.432)

There are $2M - 1$ such values for each of the $(2M - 1)^{N_b} N_b$ states. $\Upsilon$ denotes a square $(2M - 1)^{N_b} \times (2M - 1)^{N_b}$ matrix of transition probabilities where the $i,j$th element is the probability of entering state $i$, given the DFE is in state $j$. The column sums are thus all unity. For a matrix of non-negative entries, there is a famous “Peron-Frobenious” lemma from linear algebra that states that there is a unique eigenvector $\rho$ of all nonnegative entries that satisfies the equation

$$\rho = \Upsilon \rho.$$

(3.433)

The solution $\rho$ is called the stationary state distribution or Markov distribution for the state transition diagram. The $i$th entry, $\rho_i$, is the steady-state probability of being in state $i$. Of course, $\sum_{i=1}^{(2M-1)^{N_b}} \rho_i = 1$. By denoting the set of states for which $\epsilon_k \neq 0$ as $\mathcal{E}$, one determines the probability of error as

$$P_e = \sum_{i \in \mathcal{E}} \rho_i.$$

(3.434)

Thus, the DFE will have an accurate estimate of the error probability in the presence of error propagation. A larger state-transition diagram, corresponding to the explicit consideration of residual ISI as a discrete probability mass function, would yield a yet more accurate estimate of the error probability for the equalized channel - however, the relatively large magnitude of the error propagation samples usually makes their explicit consideration more important than the (usually) much smaller residual ISI.

The number of states can be very large for reasonable values of $M$ and $N_b$, so that the calculation of the stationary distribution $\rho$ could exceed the computation required for a direct measurement of SNR with a DFE simulation. There are a number of methods that can be used to reduce the number of states in the finite-state machine, most of which will reduce the accuracy of the probability of error estimate.
In the usual case, the constellation for $x_k$ is symmetric with respect to the origin, and there is essentially no difference between $\epsilon_k$ and $-\epsilon_k$, so that the analysis may merge the two corresponding states and only consider one of the error vectors. This can be done for almost half\textsuperscript{15} the states in the state transition diagram, leading to a new state transition diagram with $M^{N_b}$ states. Further analysis then proceeds as above, finding the stationary distribution and adding over states in $\mathcal{E}$. There is essentially no difference in this $P_e$ estimate with respect to the one estimated using all $(2M-1)^{N_b}$ states; however, the number of states can remain unacceptably large.

At a loss in $P_e$ accuracy, we may ignore error magnitudes and signs, and compute error statistics for binary error event vectors, which we now denote $\epsilon = [\epsilon_1, \ldots, \epsilon_{N_b}]$, of the type (for $N_b = 3$)

\[
[0 0 0], [0 1 0], [1 0 0], [1 0 1].
\]

(3.435)

This reduces the number of states to $2^{N_b}$, but unfortunately the state transition probabilities no longer depend only on the previous state. Thus, we must try to find an upper bound on these probabilities that depends only on the previous states. In so doing, the sum of the stationary probabilities corresponding to states in $\mathcal{E}$ will also upper bound the probability of error. $\mathcal{E}$ corresponds to those states with a nonzero entry in the first position (the “odd” states, $\epsilon_1 = 1$). To get an upperbound on each of the transition probabilities we write

\[
P_{\epsilon_{1,k}=1/\epsilon_{k-1}} = \sum_{i \mid \epsilon_{1,k}(i)\text{is allowed transition from }\epsilon_{k-1}} P_{\epsilon_{1,k}=1/\epsilon_{k-1}, \epsilon_{1,k}(i)} P_{\epsilon_{1,k}(i)} \leq \sum_{i \mid \epsilon_{1,k}(i)\text{is allowed transition from }\epsilon_{k-1}} \max_{\epsilon_{1,k}(i)} P_{\epsilon_{1,k}=1/\epsilon_{k-1}, \epsilon_{1,k}(i)} P_{\epsilon_{1,k}(i)}\]

(3.436)

\[
= \max_{\epsilon_{1,k}(i)} P_{\epsilon_{1,k}=1/\epsilon_{k-1}, \epsilon_{1,k}(i)} P_{\epsilon_{1,k}(i)}.
\]

(3.437)

(3.438)

Explicit computation of the maximum probability in (3.438) occurs by noting that this error probability corresponds to a worst-case signal offset of

\[
\delta_{\max}(\epsilon) = (M-1)d \sum_{i=1}^{N_b} |v_i|\epsilon_{1,k-i},
\]

(3.439)

which the reader will note is analogous to peak distortion (the distortion is understood to be the worst of the two QAM dimensions, which are assumed to be rotated so that $d_{\min}$ lies along either or both of the dimensions ). As long as this quantity is less than the minimum distance between constellation points, the corresponding error probability is then upper bounded as

\[
P_{\epsilon/\epsilon_{k-1}} \leq Q \left[ \frac{d_{\min} - \delta_{\max}(\epsilon)}{\sqrt{\sigma_{\text{MSE-DFE}}}} \right].
\]

(3.440)

Now, with the desired state-dependent (only) transition probabilities, the upper bound for $P_e$ with error propagation is

\[
P_{e} \leq \sum_{\epsilon} Q \left[ \frac{d_{\min} - \delta_{\max}(\epsilon)}{\sqrt{\sigma_{\text{MSE-DFE}}}} \right] P_{\epsilon}.
\]

(3.441)

Even in this case, the number of states $2^{N_b}$ can be too large.

A further reduction to $N_b+1$ states is possible, by grouping the $2^{N_b}$ states into groups that are classified only by the number of leading in $\epsilon_{k-1}$; thus, state $i = 1$ corresponds to any state of the form $[0 \epsilon]$, while state $i = 2$ corresponds to $[0 0 \epsilon]$, etc. The upperbound on probability of error for transitions into any state, $i$, then uses a $\delta_{\max}(i)$ given by

\[
\delta_{\max}(i) = (M-1)d \sum_{i=i+1}^{N_b} |v_i|.
\]

(3.442)

\textsuperscript{15}There is no reduction for zero entries.
and $\delta_{\text{max}}(N_b) = 0$.

Finally, a trivial bound that corresponds to noting that after an error is made, we have $M^{N_b} - 1$ possible following error event vectors that can correspond to error propagation (only the all zeros error event vector corresponds to no additional errors within the time-span of the feedback path). The probability of occurrence of these error event vectors is each no greater than the initial error probability, so they can all be considered as nearest neighbors. Thus adding these to the original probability of the first error,

$$P_e(\text{errorprop}) \leq M^{N_b} P_e(\text{first}).$$

(3.443)

It should be obvious that this bound gives useful results only if $N_b$ is small (that is a probability of error bound of $.5$ for i.i.d. input data may be lower than this bound even for reasonable values of $M$ and $N_b$). That is suppose, the first probability of error is $10^{-5}$, and $M = 8$ and $N_b = 8$, then this last (easily computed) bound gives $P_e \leq 100$ !

### 3.7.8 Look-Ahead

![Diagram](image)

Figure 3.47: Look-Ahead mitigation of error propagation in DFEs.

Figure 3.7.8 illustrates a “look-ahead” mechanism for reducing error propagation. Instead of using the decision, $M^\nu$ possible decision vectors are retained. The vector is of dimension $\nu$ and can be viewed as an address $A_{k-\Delta-1}$ to the memory. The possible output for each of the $M^\nu$ is computed and subtracted from $z_{U,k-\Delta}$. $M^\nu$ decisions of the symbol-by-symbol detector can then be computed and compared in terms of the distance from a potential symbol value, namely smallest $|\hat{E}_{U,k}(A_k^*)|$. The smallest such distance is used to select the decision for $\hat{x}_{k-\Delta}$. This method is called “look-ahead” decoding basically because all possible previous decisions' ISI are precomputed and stored, in some sense looking ahead in the calculation. If $M^\nu$ calculations (or memory locations) is too complex, then the largest $\nu' < \nu$ taps can be used (or typically the most recent $\nu'$ and the rest subtracted in typical DFE fashion for whatever the decisions previous to the $\nu'$ tap interval in a linear filter. Look-ahead leads to the Maximum Likelihood Sequence detection (MLSD) methods of Chapter 9 that are no longer symbol-by-symbol based. Look-ahead methods can never exceed \text{SNR}_{M\text{MSE}-DFE,U}$ in terms of performance, but can come very close since error propagation can be very small. (MLSD methods can exceed \text{SNR}_{M\text{MSE}-DFE,U}.)
3.8 Precoding

This section discusses solutions to the error-propagation problem of DFE’s. The first is precoding, which essentially moves the feedback section of the DFE to the transmitter with a minimal (but nonzero) transmit-power-increase penalty, but with no reduction in DFE SNR. The second approach or partial response channels (which have trivial precoders) has no transmit power penalty, but may have an SNR loss in the DFE because the feedback section is fixed to a desirable preset value for $B(D)$ rather than optimized value. This preset value (usually with all integer coefficients) leads to simplification of the ZF-DFE structure.

3.8.1 The Tomlinson Precoder

Error propagation in the DFE can be a major concern in practical application of this receiver structure, especially if constellation-expanding codes, or convolutional codes (see Chapter 10), are used in concatenation with the DFE (because the error rate on the inner DFE is lower (worse) prior to the decoder). Error propagation is the result of an incorrect decision in the feedback section of the DFE that produces additional errors that would not have occurred if that first decision had been correct.

The Tomlinson Precoder (TPC), more recently known as a Tomlinson-Harashima Precoder, is a device used to

![Diagram of the Tomlinson Precoder](image)

Figure 3.48: The Tomlinson Precoder.
eliminate error propagation and is shown in Figure 3.8.1. Figure 3.8.1(a) illustrates the case for real one-dimensional signals, while Figure 3.8.1(b) illustrates a generalization for complex signals. In the second complex case, the two real sums and two one-dimensional modulo operators can be generalized to a two-dimensional modulo where arithmetic is modulo a two-dimensional region that tesselates two-dimensional space (for example, a hexagon, or a square).

The Tomlinson precoder appears in the transmitter as a preprocessor to the modulator. The Tomlinson Precoder maps the data symbol $x_k$ into another data symbol $x'_k$, which is in turn applied to the modulator (not shown in Figure 3.8.1). The basic idea is to move the DFE feedback section to the transmitter where decision errors are impossible. However, straightforward moving of the filter $1/B(D)$ to the transmitter could result in significant transmit-power increase. To prevent most of this power increase, modulo arithmetic is employed to bound the value of $x'_k$:

**Definition 3.8.1 (Modulo Operator)** The modulo operator $\Gamma_M(x)$ is a nonlinear function, defined on an $M$-ary (PAM or QAM square) input constellation with uniform spacing $d$, such that

$$\Gamma_M(x) = x - Md\left\lfloor \frac{x + Md}{Md} \right\rfloor$$

(3.444)

where $\lfloor y \rfloor$ means the largest integer that is less than or equal to $y$. $\Gamma_M(x)$ need not be an integer. This text denotes modulo $M$ addition and subtraction by $\oplus_M$ and $\ominus_M$ respectively. That is

$$x \oplus_M y \triangleq \Gamma_M[x + y]$$

(3.445)

and

$$x \ominus_M y \triangleq \Gamma_M[x - y]$$

(3.446)

For complex QAM, each dimension is treated modulo $\sqrt{M}$ independently.

Figure 3.8.1 illustrates modulo arithmetic for $M = 4$ PAM signals with $d = 2$.

![Figure 3.49: Illustration of modulo arithmetic operator.](image)

The following lemma notes that the modulo operation distributes over addition:

$$x_k, \quad \Gamma(x_k)$$

$$+3, +3$$

$$+3$$

$$\quad -3$$

$$\quad -3$$

$$\quad +3.5$$

$$\quad -3$$

$$\quad -6$$

$$\quad +2$$

$$\quad +6.1$$

$$\quad -1.9$$
Lemma 3.8.1 (Distribution of $\Gamma_M(x)$ over addition) The modulo operator can be distributed over a sum in the following manner:

\[
\begin{align*}
\Gamma_M[x + y] &= \Gamma_M(x) \oplus_M \Gamma_M(y) \quad (3.447) \\
\Gamma_M[x - y] &= \Gamma_M(x) \ominus_M \Gamma_M(y) \quad (3.448)
\end{align*}
\]

The proof is trivial.

The Tomlinson Precoder generates an internal signal

\[
\tilde{x}_k = x_k - \sum_{i=1}^{\infty} b_i x'_{k-i} \quad (3.449)
\]

where

\[
x'_k = \Gamma_M[\tilde{x}_k] = \Gamma_M\left[x_k - \sum_{i=1}^{\infty} b_i x'_{k-i}\right] \quad (3.450)
\]

The scaled-by-SNR\textsubscript{MMSE-DFE}/SNR\textsubscript{MMSE-DFE,U} output of the MS-WMF in the receiver is an optimal unbiased MMSE approximation to $X(D) \cdot G_U(D)$. That is

\[
E\left[\frac{\text{SNR}_{\text{MMSE-DFE}}}{\text{SNR}_{\text{MMSE-DFE,U}}} z_k/x_k, x_{k-1}, \ldots\right] = \sum_{i=0}^{\infty} g_{U,i} x_{k-1} \quad (3.451)
\]

Thus, $B(D) = G_U(D)$. From Equation (3.254) with $x'(D)$ as the new input, the scaled feedforward filter output with the Tomlinson precoder is

\[
z_{U,k} = \left(x'_k + \sum_{i=1}^{\infty} g_{U,i} x'_{k-i}\right) + e_{U,k} \quad (3.452)
\]

$\Gamma_M[z_{U,k}]$ is determined as

\[
\begin{align*}
\Gamma_M[z_{U,k}] &= \Gamma_M[\Gamma_M(x_k - \sum_{i=1}^{\infty} g_{U,i} x'_{k-i}) + \sum_{i=1}^{\infty} g_{U,i} x'_{k-i} + e_{U,k}] \\
&= \Gamma_M[x_k - \sum_{i=1}^{\infty} g_{U,i} x'_{k-i}] + \sum_{i=1}^{\infty} g_{U,i} x'_{k-i} + e_{U,k}] + \sum_{i=1}^{\infty} g_{U,i} x'_{k-i} + e_{U,k}] \\
&= \Gamma_M[x_k + e_{U,k}] \quad (3.455) \\
&= x_k \oplus_M \Gamma_M[e_{U,k}] \quad (3.456) \\
&= x_k + e'_{U,k} \quad (3.457)
\end{align*}
\]

Since the probability that the error $e_{U,k}$ being larger than $Md/2$ in magnitude is small in a well-designed system, one can assume that the error sequence is of the same distribution and correlation properties after the modulo operation. Thus, the Tomlinson Precoder has allowed reproduction of the input sequence at the (scaled) MS-WMF output, without ISI. The original Tomlinson work was done for the ZF-DFE, which is a special case of the theory here, with $G_U(D) = P_e(D)$. The receiver corresponding to Tomlinson precoding.

253
is shown in Figure 3.8.1.

The noise power at the feed-forward output is thus almost exactly the same as that of the corresponding MMSE-DFE and with no error propagation because there is no longer any need for the feedback section of the DFE. As stated here without proof, there is only a small price to pay in increased transmitter power when the TPC is used.

**Theorem 3.8.1 (Tomlinson Precoder Output)** The Tomlinson Precoder output, when the input is an i.i.d. sequence, is also approximately i.i.d., and furthermore the output sequence is approximately uniform in distribution over the interval $[-Md/2, Md/2)$.

There is no explicit proof of this theorem for finite $M$, although it can be proved exactly as $M \to \infty$. This proof notes that the unbiased and biased receivers are identical as $M \to \infty$, because the SNR must also be infinite. Then, the modulo element is not really necessary, and the sequence $\hat{x}'_k$ can be shown to be equivalent to a prediction-error or “innovations” sequence, which is known in the estimation literature to be i.i.d. The i.i.d. part of the theorem appears to be valid for almost any $M$. The distribution and autocorrelation properties of the TPC, in closed form, remain an unsolved problem at present.

Using the uniform distribution assumption, the increase in transmit power for the TPC is from the nominal value of $\frac{(M^2-1)d^2}{12}$ to the value for a continuous uniform random variable over the output interval $[-Md/2, Md/2)$, which is $\frac{M^2d^2}{12}$, leading to an input power increase of

$$\frac{M^2}{M^2 - 1}$$

for PAM and correspondingly

$$\frac{M}{M - 1}$$

for QAM. When the input is not a square constellation, the Tomlinson Precoder Power loss is usually larger, but never more than a few dB. The number of nearest neighbors also increases to $\bar{N}_e = 2$ for all constellations. These losses can be eliminated (and actually a gain is possible) using techniques called Trellis Precoding and/or shell mapping (See Section 10.6).

### 3.8.1 Laroia or Flexible Precoder

Figure 3.8.1.1 shows the Laroia precoder, which is a variation on Tomlinson precoding introduced by Rajiv Laroia, mainly to reduce the transmit-power loss of the Tomlinson precoder. The Laroia precoder
largely preserves the shape of the transmitted constellation. The equivalent circuit is also shown in Figure 3.8.1.1 where the input is considered to be the difference between the actual input symbol value $x_k$ and the “decision” output $\lambda_k$. The decision device finds the closest point in the infinite extension of the constellation\footnote{The maximum range of such an infinite extension is actually the sum $\sum_{i=1}^\nu |\Re \arg g_{V,i}| \cdot |R_{\Re}(D)|$ or $\Im x_{\text{max}}$.}. The extension of the constellation is the set of points that continue to be spaced by $d_{\text{min}}$ from the points on the edges of the original constellation and along the same dimensions as the original constellation. $m_k$ is therefore a small error signal that is uniform in distribution over $(-d/2, d/2)$, thus having variance $d^2/12 \ll \varepsilon_x$.

The energy increase is therefore

$$\frac{\bar{E}_x + d^2}{\bar{E}_x}. \quad (3.460)$$

For PAM, this increase would be

$$\frac{M^2 - 1}{3} \frac{d^2}{4} + \frac{d^2}{4} = \frac{M^2 - 1}{M^2 - 2}, \quad (3.461)$$

while similarly it would be for SQ QAM:

$$\frac{M - 1}{M - 2}, \quad (3.462)$$

Cross constellations take a little more effort but follow the same basic idea above.

Because of the combination of channel and feedforward equalizer filters, the feedforward filter output is

$$Z_U(D) = [X(D) + M(D)] \cdot G_U(D) + E_U(D) = X(D) - \lambda(D) + E_U(D). \quad (3.463)$$

Processing $Z_U(D)$ by a decision operation leaves $X(D) - \lambda(D)$, which essentially is the decision. However, to recover the input sequence $X(D)$, the receiver forms

$$Z'(D) = \frac{X(D) - \lambda(D)}{G_U(D)} = X(D) + M(D). \quad (3.464)$$

Since $M(D)$ is uniform and has magnitude always less than $d/2$, then $M(D)$ is removed by a second truncation (which is not really a decision, but operates using essentially the same logic as the first decision).
Figure 3.51: Flexible Precoding.
EXAMPLE 3.8.1 (1 + .9D⁻¹ system with flexible precoding) The flexible precoder and corresponding receiver for the 1 + .9D⁻¹ example appear in Figure 3.8.1.2. The bias removal factor has been absorbed into all filters (7.85/6.85 multiples .633 to get .725 in feedback section, and multiplies .9469 to get 1.085 in feedforward section). The decision device is binary in this case since binary antipodal transmission is used. Note the IIR filter in the receiver. If an error is made in the first SBS in the receiver, then the magnitude of the error is 2 = | + 1 − (−1)| = | − 1 − (+1)|. Such an error is like an impulse of magnitude 2 added to the input of the IIR filter, which has impulse response (−.725)ᵏ · uₖ, producing a contribution to the correct sequence of 2 · (−.725)ᵏ · uₖ. This produces an additional 1 error half the time and an additional 2 errors 1/4 the time. Longer strings of errors will not occur. Thus, the bit and symbol error rate in this case increase by a factor of 1.5. (less than .1 dB loss).

A minor point is that the first decision in the receiver has an increased nearest neighbor coefficient of Nᵣ = 1.5.

Generally speaking, when an error is made in the receiver for the flexible precoder,

\[(x_k - \lambda_k) - \text{decision}(x_k - \lambda_k) = \epsilon \cdot \delta_k\]  \hspace{1cm} (3.465)

where \(\epsilon\) could be complex for QAM. Then the designer needs to compute for each type of such error, its probability of occurrence and then investigate the string (with \((1/b)ₖ\) being the impulse response of the post-first-decision filter in the receiver)

\[|\epsilon \cdot (1/b)ₖ|^\leq d_{\min}/2\]  \hspace{1cm} (3.466)
For some $k$, this relation will hold and that is the maximum burst length possible for the particular type of error. Given the filter $b$ is monic and minimum phase, this string should not be too long as long as the roots are not too close to the unit circle. The single error may cause additional errors, but none of these additional errors in turn cause yet more errors (unlike a DFE where second and subsequent errors can lead to some small probability of infinite-length bursts), thus the probability of an infinitely long burst is zero for the flexible precoder situation (or indeed any burst longer than the $k$ that solves the above equation.

### 3.8.2 Partial Response Channel Models

Partial-response methods are a special case of precoding design where the ISI is forced to some known well-defined pattern. Receiver detectors are then designed for such partial-response channels directly, exploiting the known nature of the ISI rather than attempting to eliminate this ISI.

Classic uses of partial response abound in data transmission, some of which found very early use. For instance, the earliest use of telephone wires for data transmission inevitably found large intersymbol interference because of a transformer that was used to isolate DC currents at one end of a phone line from those at the other (see Prob 3.35). The transformer did not pass low frequencies (i.e., DC), thus inevitably leading to non-Nyquist\textsuperscript{17} pulse shapes even if the phone line otherwise introduced no significant intersymbol interference. Equalization may be too complex or otherwise undesirable as a solution, so a receiver can use a detector design that instead presumes the presence of a known fixed ISI. Another example is magnetic recording (see Problem 3.36) where only flux changes on a recording surface can be sensed by a read-head, and thus D.C. will not pass through the “read channel,” again inevitably leading to ISI. Straightforward equalization is often too expensive at the very high speeds of magnetic-disk recording systems.

Partial-Response (PR) channels have non-Nyquist pulse responses – that is, PR channels allow intersymbol interference – over a few (and finite number of) sampling periods. A unit-valued sample of the response occurs at time zero and the remaining non-zero response samples at subsequent sampling times, thus the name “partial response.” The study of PR channels is facilitated by mathematically presuming the tacit existence of a whitened matched filter, as will be described shortly. Then, a number of common PR channels can be easily addressed.

\textsuperscript{17}That is, they do not satisfy Nyquist’s condition for nonzero ISI - See Chapter 3, Section 3.
Figure 3.53: Minimum Phase Equivalent Channel.

Figure 3.8.2(a) illustrates the whitened-matched-filter. The minimum-phase equivalent of Figure 3.8.2(b) exists if $Q(D) = \eta_0 \cdot H(D) \cdot H^*(D^{-*})$ is factorizable.\(^{18}\) This chapter focuses on the discrete-time channel and presumes the WMF’s presence without explicitly showing or considering it. The signal output of the discrete-time channel is

$$y_k = \sum_{m=\infty}^{\infty} h_m \cdot x_{k-m} + n_k,$$  \(3.467\)

or

$$Y(D) = H(D) \cdot X(D) + N(D),$$  \(3.468\)

where $n_k$ is sampled Gaussian noise with autocorrelation $\tilde{r}_{nn,k} = \|p\|^{-2} \cdot \eta_0^{-1} \cdot \frac{N_0}{2} \cdot \delta_k$ or $\tilde{R}_{nn}(D) = \|p\|^{-2} \cdot \eta_0^{-1} \cdot \frac{N_0}{2}$ when the whitened-matched filter is used, and just denoted $\sigma_{pr}^2$ otherwise. In both cases, this noise is exactly AWGN with mean-square sample value $\sigma_{pr}^2$, which is conveniently abbreviated $\sigma^2$ for the duration of this chapter.

**Definition 3.8.2 (Partial-Response Channel)** A **partial-response (PR)** channel is any discrete-time channel with input/output relation described in (3.467) or (3.468) that also satisfies the following properties:

1. $h_k$ is finite-length and causal; $h_k = 0 \ \forall \ k < 0$ or $k > \nu$, where $\nu < \infty$ and is often called the **constraint length** of the partial-response channel,

---

\(^{18}\) $H(D) = P_c(D)$ in Section 3.6.3 on ZF-DFE.
2. $h_k$ is monic; $h_0 = 1$,
3. $h_k$ is minimum phase; $H(D)$ has all $\nu$ roots on or outside the unit circle,
4. $h_k$ has all integer coefficients; $h_k \in \mathbb{Z} \forall k \neq 0$ (where $\mathbb{Z}$ denotes the set of integers).

More generally, a discrete finite-length channel satisfying all properties above except the last restriction to all integer coefficients is known as a controlled intersymbol-interference channel.

Controlled ISI and PR channels are a special subcase of all ISI channels, for which $h_k = 0 \forall k > \nu$. $\nu$ is the **constraint length** of the controlled ISI channel. Thus, effectively, the channel can be modeled as an FIR filter, and $h_k$ is the minimum-phase equivalent of the sampled pulse response of that FIR filter. The constraint length defines the span in time $\nu T$ of the non-zero samples of this FIR channel model.

The controlled ISI polynomial ($D$-transform)\(^{19}\) for the channel simplifies to

$$H(D) \triangleq \sum_{m=0}^{\nu} h_m D^m ,$$

(3.469)

where $H(D)$ is always monic, causal, minimum-phase, and an all-zero (FIR) polynomial. If the receiver processes the channel output of an ISI-channel with the same whitened-matched filter that occurs in the ZF-DFE of Section 3.6, and if $P_c(D)$ (the resulting discrete-time minimum-phase equivalent channel polynomial when it exists) is of finite degree $\nu$, then the channel is a controlled intersymbol interference channel with $H(D) = P_c(D)$ and $\sigma^2 = \frac{N_0}{2} \cdot \|p\|^{-2} \cdot \eta_0^{-1}$. Any controlled intersymbol-interference channel is in the form that the Tomlinson Precoder of Section 3.8.1 could be used to implement symbol-by-symbol detection on the channel output. As noted in Section 3.8.1, a (usually small) transmit symbol energy increase occurs when the Tomlinson Precoder is used. Section 3.8.4 shows that this loss can be avoided for the special class of polynomials $H(D)$ that are partial-response channels.

### 3.8.2.1 Equalizers for the Partial-response channel.

This small subsection serves to simplify and review equalization structure on the partial-response channel.

The combination of integer coefficients and minimum-phase constraints of partial-response channels allows a very simple implementation of the ZF-DFE as in Figure 3.8.2.1. Since the PR channel is already discrete time, no sampling is necessary and the minimum-phase characteristic of $H(D)$ causes the WMF of the ZF-DFE to simplify to a combined transfer function of 1 (that is, there is no feedforward section because the channel is already minimum phase). The ZF-DFE is then just the feedback section shown, which easily consists of the feedback coefficients $-h_m$ for the delayed decisions corresponding to $\hat{x}_{k-m}, m = 1, \ldots, \nu$. The loss with respect to the matched filter bound is trivially $1/\|H\|^2$, which is easily computed as 3 dB for $H(D) = 1 \pm D^\nu$ (any finite $\nu$) with simple ZF-DFE operation $z_k = y_k - \hat{x}_{k-\nu}$ and 6 dB for $H(D) = 1 + D - D^2 - D^3$ with simple ZF-DFE operation $z_k = y_k - \hat{x}_{k-1} + \hat{x}_{k-2} + \hat{x}_{k-3}$.

---

\(^{19}\)The $D$-Transform of FIR channels ($\nu < \infty$) is often called the “channel polynomial,” rather than its “$D$-Transform” in partial-response theory. This text uses these two terms interchangeably.
The ZF-LE does not exist if the channel has a zero on the unit circle, which all partial-response channels do. A MMSE-LE could be expected to have significant noise enhancement while existing. The MMSE-DFE will perform slightly better than the ZF-DFE, but is not so easy to compute as the simple DFE. Tomlinson or Flexible precoding could be applied to eliminate error propagation for a small increase in transmit power ($\frac{M^2}{M^2 - 1}$).

### 3.8.3 Classes of Partial Response

A particularly important and widely used class of partial response channels are those with $H(D)$ given by

$$H(D) = (1 + D)^l (1 - D)^n,$$

(3.470)

where $l$ and $n$ are nonnegative integers.

For illustration, Let $l = 1$ and $n = 0$ in (3.470), then

$$H(D) = 1 + D,$$

(3.471)

which is sometimes called a “duobinary” channel (introduced by Lender in 1960). The Fourier transform of the duobinary channel is

$$H(e^{-j\omega T}) = H(D)|_{D = e^{-j\omega T}} = 1 + e^{-j\omega T} = 2e^{-j\omega T/2} \cdot \cos \left( \frac{\omega T}{2} \right).$$

(3.472)

The transfer function in (3.472) has a notch at the Nyquist Frequency and is generally “lowpass” in shape, as is shown in Figure 3.55.
A discrete-time ZFE operating on this channel would produce infinite noise enhancement. If $\text{SNR}_{MFB} = 16 \text{ dB}$ for this channel, then

$$Q(e^{-j\omega T}) + \frac{1}{\text{SNR}_{MFB}} = (1 + 1/40) + \cos \omega T.$$ (3.473)

The MMSE-LE will have performance (observe that $N_0^2 = 0.05$)

$$\sigma^2_{\text{MMSE-LE}} = \frac{N_0}{2} \frac{T}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\|p\|^2(1.025 + \cos \omega T)} d\omega = \frac{N_0}{2} \frac{1/2}{\sqrt{1.025^2 - 1^2}} = 2.22 \frac{N_0}{2},$$ (3.474)

The SNR$_{\text{MMSE-LE}}$ is easily computed to be 8 (9dB), so the equalizer loss is 7 dB in this case. For the MMSE-DFE, $Q(D) + \frac{1}{\text{SNR}_{MFB}} = \frac{1.025}{1.025^2 - 1^2}$, so that $\gamma_0 = 1.025/1.64 = 0.625,$ and thus $\gamma_{\text{MMSE-DFE}} = 2.2$dB. For the ZF-DFE, $\gamma_0 = \frac{1}{2}$, and thus $\gamma_{\text{ZF-DFE}} = 3$dB.

It is possible to achieve MFB performance on this channel with complexity far less than any equalizer studied earlier in this chapter, as will be shown in Chapter 9. It is also possible to use a precoder with no transmit power increase to eliminate the error-propagation-prone feedback section of the ZF-DFE.

There are several specific channels that are used in practice for partial-response detection:

**EXAMPLE 3.8.2 (Duobinary 1 + D)** The duobinary channel (as we have already seen) has

$$H(D) = 1 + D.$$ (3.475)

The frequency response was already plotted in Figure 3.55. This response goes to zero at the Nyquist Frequency, thus modeling a lowpass-like channel. For a binary input of $x_k = \pm 1,$ the channel output (with zero noise) takes on values $\pm 2$ with probability $1/4$ each and 0 with probability $1/2.$ In general, for $M$-level inputs ($\pm 1 \pm 3 \pm 5 \ldots \pm (M-1)$), there are $2M - 1$ possible output levels, $-2M + 2, \ldots, 0, \ldots, 2M - 2.$ These output values are all possible sums of pairs of input symbols.
EXAMPLE 3.8.3 (DC Notch $1 - D$) The DC Notch channel has

$$H(D) = 1 - D \quad ,$$

so that $l = 0$ and $n = 1$ in (3.470). The frequency response is

$$H(e^{-j\omega T}) = 1 - e^{-j\omega T} = 2je^{-j\omega T} \cdot \sin \frac{\omega T}{2} \quad .$$

The response goes to zero at the DC ($\omega = 0$), thus modeling a highpass-like channel. For a binary input of $x_k = \pm 1$, the channel output (with zero noise) takes on values $\pm 2$ with probability $1/4$ each and 0 with probability $1/2$. In general for $M$-level inputs ($\pm 1 \pm 3 \pm 5 ... \pm (M - 1)$), there are $2M - 1$ possible output levels, $-2M + 2, ..., 0, ..., 2M - 2$.

When the $1 - D$ shaping is imposed in the modulator itself, rather than by a channel, the corresponding modulation is known as AMI (Alternate Mark Inversion) if a differential encoder is also used as shown later in this section. AMI modulation prevents “charge” (DC) from accumulating and is sometimes also called “bipolar coding,” although the use of the latter term is often confusing because bipolar transmission may have other meanings for some communications engineers. AMI coding, and closely related methods are used in multiplexed T1 (1.544 Mbps DS1 or “ANSI T1.403”) and E1 (2.048 Mbps or “ITU-T G.703”) speed digital data transmission on twisted pairs or coaxial links. These signals were once prevalent in telephone-company non-fiber central-office transmission of data between switch elements.

EXAMPLE 3.8.4 (Modified Duobinary $1 - D^2$) The modified duobinary channel has

$$H(D) = 1 - D^2 = (1 + D)(1 - D) \quad ,$$

so $l = n = 1$ in (3.470). Modified Duobinary is sometimes also called “Partial Response Class IV” or PR4 or PRIV in the literature. The frequency response is

$$H(e^{-j\omega T}) = 1 - e^{-j\omega T} = 2je^{-j\omega T} \cdot \sin(\omega T) \quad .$$

The response goes to zero at the DC ($\omega = 0$) and at the Nyquist frequency ($\omega = \pi/T$), thus modeling a bandpass-like channel. For a binary input of $x_k = \pm 1$, the channel output (with zero noise) takes on values $\pm 2$ with probability $1/4$ each and 0 with probability $1/2$. In general, for $M$-level inputs ($\pm 1 \pm 3 \pm 5 ... \pm (M - 1)$), there are $2M - 1$ possible output levels. Modified duobinary is equivalent to two interleaved $1 - D$ channels, each independently acting on the inputs corresponding to even (odd) time samples, respectively. Many commercial disk drives use PR4.
EXAMPLE 3.8.5 (Extended Partial Response 4 and 6 $(1 + D)^l(1 - D)^n$) The EPR4 channel has $l = 2$ and $n = 1$ or

$$H(D) = (1 + D)^2(1 - D) = 1 + D - D^2 - D^3.$$ \hspace{1cm} (3.480)

This channel is called EPR4 because it has 4 non-zero samples (Thapar). The frequency response is

$$H(e^{-j\omega T}) = (1 + e^{-j\omega T})^2(1 - e^{-j\omega T}).$$ \hspace{1cm} (3.481)

The EPR6 channel has $l = 4$ and $n = 1$ (6 nonzero samples)

$$H(D) = (1 + D)^4(1 - D) = 1 + 3D + 2D^2 - 2D^3 - 3D^4 - D^5.$$ \hspace{1cm} (3.482)

The frequency response is

$$H(e^{-j\omega T}) = (1 + e^{-j\omega T})^4(1 - e^{-j\omega T}).$$ \hspace{1cm} (3.483)

These 2 channels, along with PR4, are often used to model disk storage channels in magnetic disk or tape recording. The response goes to zero at DC ($\omega = 0$) and at the Nyquist frequency in both EPR4 and EPR6, thus modeling bandpass-like channels. The magnitude of these two frequency characteristics are shown in Figure 3.56. These are increasingly used in commercial disk storage read detectors.

The higher the $l$, the more “lowpass” in nature that the EPR channel becomes, and the more appropriate as bit density increases on any given disk.

For partial-response channels, the use of Tomlinson Precoding permits symbol-by-symbol detection, but also incurs an $M^2/(M^2 - 1)$ signal energy loss for PAM (and $M/(M - 1)$ for QAM). A simpler method for PR channels, that also has no transmit energy penalty, is known simply as a “precoder.”

3.8.4 Simple Precoding

The simplest form of precoding for the duobinary, DC-notch, and modified duobinary partial-response channels is the so-called “differential encoder.” The message at time $k$, $m_k$, is assumed to take on values $m = 0, 1, ..., M - 1$. The differential encoder for the case of $M = 2$ is most simply described as
the device that observes the input bit stream, and changes its output if the input is 1 and repeats the last output if the input is 0. Thus, the differential encoder input, \( m_k \), represents the difference (or sum) between adjacent differential encoder output (\( \bar{m}_k \) and \( \bar{m}_{k-1} \)) messages.\(^{20}\)

**Definition 3.8.3 (Differential Encoder)** Differential encoders for PAM or QAM modulation obey one of the two following relationships:

\[
\bar{m}_k = m_k \oplus \bar{m}_{k-1}, \\
\bar{m}_k = m_k \ominus \bar{m}_{k-1},
\]

where \( \ominus \) represents subtraction modulo \( M \) (and \( \oplus \) represents addition modulo \( M \)). For SQ QAM, the modulo addition and subtraction are performed independently on each of the two dimensions, with \( \sqrt{M} \) replacing \( M \). (A differential phase encoder is also often used for QAM and is discussed later in this section.)

A differential encoder is shown on the left side of Figure 3.8.4. As an example if \( M = 4 \) the corresponding inputs and outputs are given in the following table:

<table>
<thead>
<tr>
<th>( k )</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{m}_k )</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>( m_k )</td>
<td>-3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

With either PAM or QAM constellations, the dimensions of the encoder output \( \bar{m}_k \) are converted into channel input symbols \( (\pm \frac{d}{2}, \pm \frac{3d}{2}, \ldots) \) according to

\[
x_k = [2\bar{m}_k - (M - 1)]\frac{d}{2}.
\]

\(^{20}\)This operation is also very useful even on channels without ISI, as an unknown inversion in the channel (for instance, an odd number of amplifiers) will cause all bits to be in error if (differential) precoding is not used.
outputs is that of differential encoding as defined in (3.484). The noiseless minimum-phase-equivalent channel output \( \tilde{y}_k \) is
\[
\tilde{y}_k = x_k + x_{k-1} = 2(\bar{m}_k + m_{k-1})\frac{d}{2} - 2(M - 1)\frac{d}{2}
\] (3.487)
so
\[
\frac{\tilde{y}_k}{d} + (M - 1) = \bar{m}_k + m_{k-1}
\] (3.488)
\[
\left( \frac{\tilde{y}_k}{d} + (M - 1) \right)_{M} = m_k \oplus m_{k-1}
\] (3.489)
\[
\left( \frac{\tilde{y}_k}{d} + (M - 1) \right)_{M} = m_k
\] (3.490)
where the last relation (3.490) follows from the definition in (3.484). All operations are integer mod-\( M \). Equation (3.490) shows that a decision on \( \tilde{y}_k \) about \( m_k \) can be made without concern for preceding or succeeding \( y_k \), because of the action of the precoder. The decision boundaries are simply the obvious regions symmetrically placed around each point for a memoryless ML detector of inputs and the decoding rules for \( M = 2 \) and \( M = 4 \) are shown in Figure 3.8.4. In practice, the decoder observes \( y_k \), not \( \tilde{y}_k \), so that \( y_k \) must first be quantized to the closest noise-free level of \( y_k \). That is, the decoder first quantizes \( y_k \) to one of the values of \(-2(2M + 2)(d/2), \ldots, (2M - 2)(d/2)\). Then, the minimum distance between outputs is \( d_{\text{min}} = d \), so that \( d_{\text{min}} \frac{\sigma^2_{\tilde{p}}}{\sigma^2_p} = \frac{d}{2\sigma^2_p} \), which is 3 dB below the \( \sqrt{MFB} = \frac{\sqrt{2}}{\sigma^2_p}(d/2) \) for the \( 1 + D \) channel because \( \|p\|^2 = 2 \). (Again, \( \sigma^2_{\tilde{p}} = \|p\|^2 \cdot \eta_0^{-1} \cdot \frac{\sqrt{2}}{\sigma^2_p} \) with a WMF and otherwise is just \( \sigma^2_{p} \).) This loss is identical to the loss in the ZF-DFE for this channel. One may consider the feedback section of the ZF-DFE as having been pushed through the linear channel back to the transmitter, where it becomes the precoder. With some algebra, one can show that the TPC, while also effective, would produce a 4-level output with 1.3 dB higher average transmit symbol energy for binary inputs. The output levels are assumed equally likely in determining the decision boundaries (half way between the levels) even though these levels are not equally likely.

The precoded partial-response system eliminates error propagation, and thus has lower \( P_e \) than the ZF-DFE. This elimination of error propagation can be understood by investigating the nearest neighbor coefficient for the precoded situation in general. For the \( 1 + D \) channel, the (noiseless) channel output levels are \(-2M - 2 - (2M - 4) \ldots 0 \ldots (2M - 4) (2M - 2)\) with probabilities of occurrence \( \frac{1}{M^2} \cdot \frac{2}{M^2} \ldots \frac{1}{M^2} \), assuming a uniform channel-input distribution. Only the two outer-most levels have one nearest neighbor, all the rest have 2 nearest neighbors. Thus,
\[
\bar{N}_e = \frac{2}{M^2} (1) + \frac{M^2 - 2}{M^2} (2) = 2 \left( 1 - \frac{1}{M^2} \right)
\] (3.491)
For the ZF-DFE, the input to the decision device \( \hat{z}_k \) as
\[
\hat{z}_k = x_k + x_{k-1} + n_k - \hat{x}_{k-1}
\] (3.492)
which can be rewritten
\[
\hat{z}_k = x_k + (x_{k-1} - \hat{x}_{k-1}) + n_k
\] (3.493)
Equation 3.493 becomes \( \hat{z}_k = x_k + n_k \) if the previous decision was correct on the previous symbol. However, if the previous decision was incorrect, say +1 was decided (binary case) instead of the correct -1, then
\[
\hat{z}_k = x_k - 2 + n_k
\] (3.494)
which will lead to a next-symbol error immediately following the first almost surely if \( x_k = 1 \) (and no error almost surely if \( x_k = -1 \)). The possibility of \( \hat{z}_k = x_k + 2 + n_k \) is just as likely to occur and follows

---

\(^{21}(\bullet)_M \) means the quantity is computed in \( M \)-level arithmetic, for instance, \((5)_4 = 1 \). Also note that \( \Gamma_M(x) \neq (x)_M \), and therefore \( \oplus \) is different from the \( \oplus_M \) of Section 3.5. The functions \((x)_M \) and \( \oplus \) have strictly integer inputs and have possible outputs 0,...,\( M - 1 \) only.
an identical analysis with signs reversed. For either case, the probability of a second error propagating is 1/2. The other half of the time, only 1 error occurs. Half the times that 2 errors occur, a third error also occurs, and so on, effectively increasing the error coefficient from \( N_e = 1 \) to

\[
N_e = 2 = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + 4\left(\frac{1}{16}\right) + \ldots = \sum_{k=1}^{\infty} k \cdot 0.5^k = \frac{5}{(1 - 0.5)^2} .
\]  
(3.495)

(In general, the formula \( \sum_{k=1}^{\infty} k \cdot r^k = \frac{r^2}{(1-r)^2}, r < 1 \) may be useful in error propagation analysis.) The error propagation can be worse in multilevel PAM transmission, when the probability of a second error is \( (M-1)/M \), leading to

\[
\frac{N_e(\text{error prop})}{N_e(\text{no error prop})} = \frac{1 \cdot \frac{1}{M} + 2 \cdot \frac{M-1}{M} \cdot \frac{1}{M} + 3 \cdot \left(\frac{M-1}{M}\right)^2 \cdot \frac{1}{M} + \ldots}{M}. 
\]  
(3.496)

Precoding eliminates this type of error propagation, although \( \tilde{N}_e \) increases by a factor\(^\text{22}\) of \((1 + 1/M)\) with respect to the case where no error propagation occurred.

For the ZF-DFE system, \( \tilde{P}_e = 2(M-1)Q\left(\frac{d}{2}\sigma_{pr}\right) \), while for the precoded partial-response system, \( \tilde{P}_e = 2 \left(1 - \frac{1}{M^2}\right)Q\left(\frac{d}{2\sigma_{pr}}\right) \). For \( M \geq 2 \), the precoded system always has the same or fewer nearest neighbors, and the advantage becomes particularly pronounced for large \( M \). Using a rule-of-thumb that a factor of 2 increase in nearest neighbors is equivalent to an SNR loss of .2 dB (which holds at reasonable error rates in the \( 10^{-5} \) to \( 10^{-6} \) range), the advantage of precoding is almost .2 dB for \( M = 4 \). For \( M = 8 \), the advantage is about .6 dB, and for \( M = 64 \), almost 1.2 dB. Precoding can be simpler to implement than a ZF-DFE because the integer partial-response channel coefficients translate readily into easily realized finite-field operations in the precoder, while they represent full-precision add (and shift) operations in feedback section of the ZF-DFE.

### 3.8.4.1 Precoding the DC Notch or Modified Duobinary Channels

The \( \tilde{m}_k = m_k \ominus \tilde{m}_{k-1} \) differential encoder works for the \( 1 + D \) channel. For the \( 1 - D \) channel, the equivalent precoder is

\[
\tilde{m}_k = m_k \ominus \tilde{m}_{k-1} ,
\]  
(3.499)

which is sometimes also called NRZI (non-return-to-zero inverted) precoding, especially by storage-channel engineers. In the \( 1 - D \) case, the channel output is

\[
\tilde{y}_k = x_k - x_{k-1} = 2(\tilde{m}_k - \tilde{m}_{k-1})\frac{d}{2} \]  
(3.500)

\[
\tilde{y}_k \quad \text{mod} \quad M = \tilde{m}_k \ominus \tilde{m}_{k-1} \]  
(3.502)

\[
\tilde{y}_k \quad \text{mod} \quad M = m_k \]  
(3.503)

The minimum distance and number of nearest neighbors are otherwise identical to the \( 1 + D \) case just studied, as is the improvement over the ZF-DFE. The \( 1 - D^2 \) case is identical to the \( 1 - D \) case, on two interleaved \( 1 - D \) channels at half the rate. The overall precoder for this situation is

\[
\tilde{m}_k = m_k \ominus \tilde{m}_{k-2} ,
\]  
(3.504)

\(^\text{22}\)The ratio of \( 2(1 - 1/M^2) \) with precoding to \( 2(1 - 1/M) \) for \( M \)-ary PAM with no error propagation effects included
and the decision rule is
\[
\left( \frac{\tilde{y}_k}{d} \right)_M = m_k .
\] (3.505)

The combination of precoding with the \(1-D\) channel is often called “alternate mark inversion (AMI)” because each successive transmitted “1” bit value causes a nonzero channel output amplitude of polarity opposite to the last nonzero channel output amplitude, while a “0” bit always produces a 0 level at the channel output.

### 3.8.4.2 Precoding EPR4

An example of precoding for the extended Partial Response class is EPR4, which has \(l = 2\) and \(n = 1\) in (3.480), or EPR4. Then,
\[
\tilde{y}_k = x_k + x_{k-1} - x_{k-2} - x_{k-3}
\] (3.506)
\[
\tilde{y}_k = d(m_k + \tilde{m}_{k-1} - \tilde{m}_{k-2} - \tilde{m}_{k-3})
\] (3.507)
\[
\frac{\tilde{y}_k}{d} = \tilde{m}_k + \tilde{m}_{k-1} - \tilde{m}_{k-2} - \tilde{m}_{k-3}
\] (3.508)
\[
\left( \frac{\tilde{y}_k}{d} \right)_M = \tilde{m}_k \oplus \tilde{m}_{k-1} \oplus \tilde{m}_{k-2} \oplus \tilde{m}_{k-3}
\] (3.509)
\[
\left( \frac{\tilde{y}_k}{d} \right)_M = m_k
\] (3.510)

where the precoder, from (3.510) and (3.509), is
\[
\tilde{m}_k = m_k \oplus \tilde{m}_{k-1} \oplus \tilde{m}_{k-2} \oplus \tilde{m}_{k-3} .
\] (3.511)

The minimum distance at the channel output is still \(d\) in this case, so \(P_e \leq N_e \cdot Q(d/2\sigma_{pr})\), but the MFB = \(d^2\sigma_{pr}^2\), which is 6dB higher. The same 6dB loss that would occur with an error-propagation-free ZF-DFE on this channel.

For the \(1 + D - D^2 - D^3\) channel, the (noiseless) channel output levels are \(-(4M-4) - (4M-6)... 0 ... (4M-6) (4M-4)\). Only the two outer-most levels have one nearest neighbor, all the rest have 2 nearest neighbors. Thus,
\[
N_e = \frac{2}{M^4} (1) + \frac{M^4 - 2}{M^4} (2) = 2 \left( 1 - \frac{1}{M^4} \right) .
\] (3.512)

Thus the probability of error for the precoded system is \(P_e = 2 \left( 1 - \frac{1}{M^4} \right) Q(\frac{1}{\sigma_{pr}})\). The number of nearest neighbors for the ZF-DFE, due to error propagation, is difficult to compute, but clearly will be worse.

### 3.8.5 General Precoding

The general partial response precoder can be extrapolated from previous results:

\textbf{Definition 3.8.4 (The Partial-Response Precoder)} The partial-response precoder for a channel with partial-response polynomial \(H(D)\) is defined by
\[
\tilde{m}_k = m_k \bigoplus_{i=1}^{\nu} (-h_i) \cdot \tilde{m}_{k-i} .
\] (3.513)

The notation \(\bigoplus\) means a mod-\(M\) summation, and the multiplication can be performed without ambiguity because the \(h_i\) and \(\tilde{m}_{k-i}\) are always integers.
The corresponding memoryless decision at the channel output is

\[
\hat{m}_k = \left( \frac{\hat{y}_k}{d} + \sum_{i=0}^{\nu} h_i \left( \frac{M-1}{2} \right) \right)_M .
\]

(3.514)

The reader should be aware that while the relationship in (3.513) is general for partial-response channels, the relationship can often simplify in specific instances, for instance the precoder for “EPR5,” \( H(D) = (1 + D)^3(1 - D) \) simplifies to \( \hat{m}_k = m_k \oplus \bar{m}_{k-4} \) when \( M = 2 \).

In a slight abuse of notation, engineers often simplify the representation of the precoder by simply writing it as

\[
P(D) = \frac{1}{H(D)}
\]

(3.515)

where \( P(D) \) is a polynomial in \( D \) that is used to describe the “modulo-\( M \)” filtering in (3.513). Of course, this notation is symbolic. Furthermore, one should recognize that the \( D \) means unit delay in a finite field, and is therefore a delay operator only – one cannot compute the Fourier transform by inserting \( D = e^{-j\omega T} \) into \( P(D) \); Nevertheless, engineers commonly refer to the NRZI precoder as a \( 1/(1 \oplus D) \) precoder.

**Lemma 3.8.2 (Memoryless Decisions for the Partial-Response Precoder)**

The partial-response precoder, often abbreviated by \( P(D) = 1/H(D) \), enables symbol-by-symbol decoding on the partial-response channel \( H(D) \). An upper bound on the performance of such symbol-by-symbol decoding is

\[
P_e \leq 2 \left( 1 - \frac{1}{M^{\nu+1}} \right) Q \left[ \frac{d}{2\sigma_{pr}} \right],
\]

(3.516)

where \( d \) is the minimum distance of the constellation that is output to the channel.

**Proof:** The proof follows by simply inserting (3.513) into the expression for \( \hat{y}_k \) and simplifying to cancel all terms with \( h_i \), \( i > 0 \). The nearest-neighbor coefficient and \( d/2\sigma_{pr} \) follow trivially from inspection of the output. Adjacent levels can be no closer than \( d \) on a partial-response channel, and if all such adjacent levels are assumed to occur for an upper bound on probability of error, then the \( P_e \) bound in (3.516) holds. QED.

### 3.8.6 Quadrature PR

Differential encoding of the type specified earlier is not often used with QAM systems, because QAM constellations usually exhibit 90° symmetry. Thus a 90° offset in carrier recovery would make the constellation appear exactly as the original constellation. To eliminate the ambiguity, the bit assignment for QAM constellations usually uses a precoder that uses the four possibilities of the most-significant bits that represent each symbol to specify a phase rotation of 0°, 90°, 180°, or 270° with respect to the last symbol transmitted. For instance, the sequence 01, 11, 10, 00 would produce (assuming an initial phase of 0°, the sequence of subsequent phases 90°, 0°, 180°, and 180°. By comparing adjacent decisions and their phase difference, these two bits can be resolved without ambiguity even in the presence of unknown phase shifts of multiples of 90°. This type of encoding is known as **differential phase encoding** and the remaining bits are assigned to points in large \( M \) QAM constellations so that they are the same for points that are just 90° rotations of one another. (Similar methods could easily be derived using 3 or more bits for constellations with even greater symmetry, like 8PSK.)

Thus the simple precoder for the \( 1 + D \) and \( 1 - D \) (or \( 1 - D^2 \)) given earlier really only is practical for PAM systems. The following type of precoder is more practical for these channels:
A Quadrature Partial Response (QPR) situation is specifically illustrated in Figure 3.8.6 for \( M = 4 \) or \( b = 1 \) bit per dimension. The previous differential encoder could be applied individually to both dimensions of this channel, and it could be decoded without error. All previous analysis is correct, individually, for each dimension. There is, however, one practical problem with this approach: If the channel were to somehow rotate the phase by \( \pm 90^\circ \) (that is, the carrier recovery system locked on the wrong phase because of the symmetry in the output), then there would be an ambiguity as to which part was real and which was imaginary. Figure 3.8.6 illustrates the ambiguity: the two messages \((0,1) = 1\) and \((1,0) = 2\) commute if the channel has an unknown phase shift of \( \pm 90^\circ \). No ambiguity exists for either the message \((0,0) = 0\) or the message \((1,1) = 3\). To eliminate the ambiguity, the precoder encodes the 1 and 2 signals into a difference between the last 1 (or 2) that was transmitted. This precoder thus specifies that an input of 1 (or 2) maps to no change with respect to the last input of 1 (or 2), while an input of 2 maps to a change with respect to the last input of 1 or 2.

A precoding rule that will eliminate the ambiguity is then

**Rule 3.8.1 (Complex Precoding for the 1+D Channel)** if \( m_k = (m_{i,k}, m_{q,k}) = (0,0) \) or \((1,1)\) then

\[
\tilde{m}_{i,k} = m_{i,k} \oplus \tilde{m}_{i,k-1} \\
\tilde{m}_{q,k} = m_{q,k} \oplus \tilde{m}_{q,k-1}
\]

else if \( m_k = (m_{i,k}, m_{q,k}) = (0,1) \) or \((1,0)\), check the last \( \tilde{m} = (0,1) \) or \((1,0)\) transmitted, call it \( \tilde{m}_{90} \), (that is, was \( \tilde{m}_{90} = 1 \) or 2 ?). If \( \tilde{m}_{90} = 1 \), the precoder leaves \( m_k \)
unchanged prior to differential encoding according to (3.517) and (3.518). The operations \(\oplus\) and \(\ominus\) are the same in binary arithmetic.

If \(\vec{m}_{90} = 2\), then the precoder changes \(m_k\) from 1 to 2 (or from 2 to 1) prior to encoding according to (3.517) and (3.518).

The corresponding decoding rule is (keeping a similar state \(\hat{y}_{90} = 1, 2\) at the decoder)

\[
\hat{m}_k = \begin{cases} 
(0, 0) & \hat{y}_k = [\pm 2 \pm 2] \\
(1, 1) & \hat{y}_k = [0 0] \\
(0, 1) & (\hat{y}_k = [0 \pm 2] \text{ or } \hat{y}_k = [\pm 2 0]) \text{ and } \angle{\hat{y}}_k - \angle{\hat{y}}_{90} = 0 \\
(1, 0) & (\hat{y}_k = [0 \pm 2] \text{ or } \hat{y}_k = [\pm 2 0]) \text{ and } \angle{\hat{y}}_k - \angle{\hat{y}}_{90} \neq 0
\end{cases}
\]

(3.519)

The probability of error and minimum distance are the same as was demonstrated earlier for this type of precoder, which only resolves the 90° ambiguity, but is otherwise equivalent to a differential encoder. There will however be limited error propagation in that one detection error on the 1 or 2 message points leads to two decoded symbol errors on the decoder output.
3.9 Diversity Equalization

Diversity in transmission occurs when there are multiple channels from a single message source to several receivers, as illustrated in Figure 3.9.1, and as detailed in Subsection 3.9.1. Optimally, the principles of Chapter 1 apply directly where the channel’s conditional probability distribution $p_{y|x}$ typically has a larger channel-output dimensionality for $y$ than for the input, $x$, $N_y > N_x$. This diversity often leads to a lower probability of error for the same message, mainly because a greater channel-output minimum distance between possible (noiseless) output data symbols can be achieved with a larger number of channel output dimensions. However, intersymbol interference between successive transmissions, along with interference between the diversity dimensions, again can lead to a potentially complex optimum receiver and detector. Thus, equalization again allows productive use of suboptimal SBS detectors with diversity. The diversity equalizers become matrix equivalents of those studied in Sections 3.5 - 3.7.

3.9.1 Multiple Received Signals and the RAKE

Figure 3.59: Basic diversity channel model.

Figure 3.9.1 illustrates the basic diversity channel. Channel outputs caused by the same channel input have labels $y_l(t)$, $l = 0, ..., L - 1$. These channel outputs can be created intentionally by retransmission of the same data symbols at different times and/or (center) frequencies. Spatial diversity often occurs in wireless transmission where $L$ spatially separated antennas may all receive the same transmitted signal, but possibly with different filtering and with noises that are at least partially independent.
With each channel output following the model
\[ y_{p,l}(t) = \sum_k x_k \cdot p_l(t - kT) + n_l(t) \quad , \] (3.520)
a corresponding \( L \times 1 \) vector channel description is
\[ y_p(t) = \sum_k x_k \cdot p(t - kT) + n_p(t) \quad , \] (3.521)
where
\[ y_p(t) \triangleq \begin{bmatrix} y_0(t) \\ y_1(t) \\ \vdots \\ y_{L-1}(t) \end{bmatrix}, \quad p(t) \triangleq \begin{bmatrix} p_0(t) \\ p_1(t) \\ \vdots \\ p_{L-1}(t) \end{bmatrix}, \quad \text{and} \quad n_p(t) \triangleq \begin{bmatrix} n_0(t) \\ n_1(t) \\ \vdots \\ n_{L-1}(t) \end{bmatrix}. \] (3.522)
From Section 1.1.5, generalization of an inner product
\[ \langle x(t), y(t) \rangle = \sum_i \int_{-\infty}^{\infty} x_i^*(t) y_i(t) \, dt \quad . \] (3.523)

Without loss of generality, the noise can be considered to be white on each of the \( L \) diversity channels, independent of the other diversity channels, and with equal power spectral densities \( \frac{N_0}{2} \).\(^{23}\)

A single transmission of \( x_0 \) corresponds to a vector signal
\[ y_p(t) = x_0 \cdot p(t) + n(t) = x_0 \cdot \|p\| \cdot \varphi_p(t) + n(t) \quad . \] (3.524)
This situation generalizes slightly that considered in Chapter 1, where matched-filter demodulators there combined all time instants through integration, a generalization of the inner product’s usual sum of products. A matched filter in general simply combines the signal components from all dimensions that have independent (pre-whitened if necessary) noise. Here, the inner product includes also the components corresponding to each of the diversity channels so that all signal contributions are summed to create maximum signal-to-noise ratio. The relative weighting of the different diversity channels is thus maintained through \( L \) unnormalized parallel matched filters each corresponding to one of the diversity channels. When several copies are combined across several diversity channels or new dimensions (whether created in frequency, long delays in time, or space), the combination is known as the RAKE matched filter of Figure 3.9.1:

\[ \textbf{Definition 3.9.1 (RAKE matched filter) A RAKE matched filter is a set of parallel matched filters each operating on one of the diversity channels in a diversity transmission system that is followed by a summing device as shown in Figure 3.9.1. Mathematically, the operation is denoted by} \]
\[ y_p(t) = \sum_{l=0}^{L-1} p_l^*(-t) \ast y_l(t) \quad . \] (3.525)

\(^{23}\)In practice, the noises may be correlated with each other on different subchannels and not white with covariance matrix \( R_n(t) \) and power spectral density matrix \( R_n(f) = \frac{N_0}{2} \cdot R_n^{1/2}(f) R_n^{1/2}(f) \). By prefiltering the vector channel output by the matrix filter \( R_n^{-1/2}(f) \), the noise will be whitened and the noise equivalent matrix channel becomes \( P(f) \rightarrow R_n^{-1/2}(f) P(f) \). Analysis with the equivalent channel can then proceed as if the noise were white, independent on the diversity channels, and of the same variance \( \frac{N_0}{2} \) on all.
The RAKE was originally so named by Green and Price in 1958 because of the analogy of the various matched filters being the “fingers” of a garden rake and the sum corresponding to the collection of the fingers at the rake’s pole handle. The RAKE is sometimes also called a **diversity combiner**, although the latter term also applies to other lower-performance suboptimal combining methods that do not maximize overall signal-to-noise strength through matched filter. One structure, often called **maximal combining**, applies a matched filter only to the strongest of the $L$ diversity paths to save complexity. The equivalent channel for this situation is then the channel corresponding to this maximum-strength individual path. The original RAKE concept was conceived in connection with a spread-spectrum transmission method that achieves diversity essentially in frequency (but more precisely in a code-division dimension to be discussed in the appendix of Chapter 5), but the matched filtering implied is easily generalized. Some of those who later studied diversity combining were not aware of the connection to the RAKE and thus the multiple names for the same structure, although diversity combining is a more accurate name for the method.

This text also defines $r(t) = p(t) * p^*(t)$ and

$$r(t) = \sum_{l=0}^{L-1} r_l(t)$$  \hspace{1cm} (3.526)

for an equivalent RAKE-output equivalent channel and the norm

$$\|p\|^2 = \sum_{l=0}^{L-1} \|p_l\|^2$$  \hspace{1cm} (3.527)

Then, the normalized equivalent channel $q(t)$ is defined through

$$r(t) = \|p\|^2 \cdot q(t)$$  \hspace{1cm} (3.528)

The sampled RAKE output has $D$-transform

$$Y(D) = X(D) \cdot \|p\|^2 \cdot Q(D) + N(D)$$  \hspace{1cm} (3.529)

which is essentially the same as the early channel models used without diversity except for the additional scale factor of $\|p\|^2$, which also occurs in the noise autocorrelation, which is

$$R_{nn}(D) = \frac{N_0}{2} \cdot \|p\|^2 \cdot Q(D)$$  \hspace{1cm} (3.530)

An $\text{SNR}_{MFB} = \frac{\hat{\xi}_e \|p\|^2}{\sqrt{N_0}}$ and all other detector and receiver principles previously developed in this text now apply directly.
If there were no ISI, or only one-shot transmission, the RAKE plus symbol-by-symbol detection would be an optimum ML/MAP detector. However, from Subsection 3.1.3, the set of samples created by this set of matched filters is sufficient. Thus, a single equalizer can be applied to the sum of the matched filter outputs without loss of generality as in Subsection 3.9.2. However, if the matched filters are absorbed into a fractionally-spaced and/or FIR equalizer, then the equalizers become distinct set of coefficients for each rail before being summed and input to a symbol-by-symbol detector as in Subsection 3.9.3.

3.9.2 Infinite-length MMSE Equalization Structures

The sampled RAKE output can be scaled by the factor $\|p\|^{-1}$ to obtain a model identical to that found earlier in Section 3.1 in Equation (3.26) with $Q(D)$ and $\|p\|$ as defined in Subsection 3.9.1. Thus, the MMSE-DFE, MMSE-LE, and ZF-LE/DFE all follow exactly as in Sections 3.5-3.7. The matched-filter bound SNR and each of the equalization structures tend to work better with diversity because $\|p\|^2$ is typically larger on the equivalent channel created by the RAKE. Indeed the RAKE will work better than any of the individual channels, or any subset of the diversity channels, with each of the equalizer structures.

Often while one diversity channel has severe characteristics, like an inband notch or poor transmission characteristic, a second channel is better. Thus, diversity systems tend to be more robust.

**EXAMPLE 3.9.1 (Two ISI channels in parallel)** Figure 3.9.2 illustrates two diversity channels with the same input and different intersymbol interference. The first upper channel has a sampled time equivalent of $1 + .9D^{-1}$ with noise variance per sample of .181 (and thus could be the channel consistently examined throughout this Chapter so $\mathcal{E}_X = 1$). This channel is in effect anti-causal (or in reality, the .9 comes first). A second channel has causal response $1 + .8D$ with noise variance .164 per sample and is independent of the noise in the first channel. The ISI effectively spans 3 symbol periods among the two channels at a common receiver that will decide whether $x_k = \pm 1$ has been transmitted.

The SNR$_{MF,B}$ for this channel remains $\text{SNR}_{MF,B} = \frac{\mathcal{E}_X \|P\|^2}{\Delta}$, but it remains to compute this quantity correctly. First the noise needs to be whitened. While the two noises are
independent, they do not have the same variance per sample, so a pre-whitening matrix is
\[
\begin{bmatrix}
1 & 0 \\
0 & \sqrt{\frac{.181}{.164}}
\end{bmatrix}
\]
and so then the energy quantified by \(\|p\|^2\) is
\[
\|p\|^2 = \|p_1\|^2 + \|\tilde{p}_2\|^2 = 1.81 + \left(\frac{.181}{.164}\right)1.64 = 2(1.81)
\]
Then
\[
\text{SNR}_{MFB} = \frac{1 \cdot 2(1.81)}{.181} = 13 \text{ dB}
\]
Because of diversity, this channel has a higher potential performance than the single channel alone. Clearly having a second look at the input through another channel can’t hurt (even if there is more ISI now). The ISI is characterized as always by \(Q(D)\), which in this case is
\[
Q(D) = \frac{1}{2(1.81)} \left[ (1 + .9D^{-1})(1 + .9D) + \frac{.181}{.164} (1 + .8D)(1 + .8D^{-1}) \right]
\]
\[
= .492D + 1 + .492D^{-1}
\]
\[
= .589 \cdot (1 + .835D) \cdot (1 + .835D^{-1})
\]
\[
\tilde{Q}(D) = .492D + (1 + 1/20) + .492D^{-1}
\]
\[
= .7082 \cdot (1 + .695D) \cdot (1 + .695D^{-1})
\]
Thus, the SNR of a MMSE-DFE would be
\[
\text{SNR}_{MMSE-DFE,U} = .7082(20) - 1 = 13.16 \text{ (11.15 dB)}
\]
The improvement of diversity with respect to a single channel is about 2.8 dB in this case. The receiver is a MMSE-DFE essentially designed for the 1+.839D ISI channel after adding the matched-filter outputs. The loss with respect to the MFB is about 1.8 dB.

![Figure 3.61: Two channel example.](image)

### 3.9.3 Finite-length Multidimensional Equalizers

The case of finite-length diversity equalization becomes more complex because the matched filters are implemented within the (possibly fractionally spaced) equalizers associated with each of the diversity subchannels. There may be thus many coefficients in such a diversity equalizer.\(^{24}\)

\(^{24}\)Maximal combiners that select only one (the best) of the diversity channels for equalization are popular because they reduce the equalization complexity by at least a factor of \(L\) – and perhaps more when the best subchannel needs less equalization.
In the case of finite-length equalizers shown in figure 3.9.3, each of the matched filters of the RAKE
is replaced by a lowpass filter of wider bandwidth (usually \( l \) times wider as in Section 3.7), a sampling
device at rate \( l/T \), and a fractionally spaced equalizer prior to the summing device. With vectors \( p_k \) in
Equation (3.293) now becoming \( l \cdot L \)-tuples,

\[
p_k = p(kT),
\]

as do the channel output vectors \( y_k \) and the noise vector \( n_k \), the channel input/output relationship (in
Eq (3.295)) again becomes

\[
Y_k \triangleq \begin{bmatrix}
y_k \\
y_{k-1} \\
\vdots \\
y_{k-N_f+1}
\end{bmatrix} = \begin{bmatrix}
p_0 & p_1 & \ldots & p_\nu & 0 & 0 & \ldots & 0 \\
0 & p_0 & p_1 & \ldots & p_\nu & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & p_0 & p_1 & \ldots & p_\nu & \ldots & 0 \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
\end{bmatrix} \begin{bmatrix}
x_k \\
x_{k-1} \\
\vdots \\
x_{k-N_f-\nu+1}
\end{bmatrix} + \begin{bmatrix}
n_k \\
n_{k-1} \\
\vdots \\
n_{k-N_f+1}
\end{bmatrix},
\]

(3.542)

The rest of Section 3.7 then directly applies with the matrix \( P \) changing to include the larger \( l \cdot L \)-tuples
corresponding to \( L \) diversity channels, and the corresponding equalizer \( W \) having its \( 1 \times L \) coefficients
corresponding to \( w_0 \ldots w_{N_f} \). Each coefficient thus contains \( l \) values for each of the \( L \) equalizers.

The astute reader will note that the diversity equalizer is the same in principle as a fractionally
spaced equalizer except that the oversampling that creates diversity in the FSE generalizes to simply
any type of additional dimensions per symbol in the diversity equalizer. The dfecolor program can be
used where the oversampling factor is simply \( l \cdot L \) and the vector of the impulse response is appropriately
organized to have \( l \cdot L \) phases per entry. Often, \( l = 1 \), so there are just \( L \) antennas or lines of samples
per symbol period entry in that input vector.

\[\text{Figure 3.62: Fractionally spaced RAKE MMSE-DFE.}\]
3.9.4 DFE RAKE Program

A DFE RAKE program similar to the DFE RAKE matlab program has again been written by the students (and instructor debugging!) over the past several years. It is listed here and is somewhat self-explanatory if the reader is already using the DFECOLOR program.

```
% DFE design program for RAKE receiver
% Prepared by: Debarag Banerjee, edited by Olutsein OLATUNBOSUN
% and Yun-Hsuan Sung to add colored and spatially correlated noise
% Significant Corrections by J. Cioffi and above to get correct results --
% March 2005
% function [dfseSNR,W,b]=dfsecolorsnr(l,p,nff,nbb,delay,Ex,noise);
% ****************************
%**** only computes SNR ****
% l = oversampling factor
% L = No. of fingers in RAKE
% p = pulse response matrix, oversampled at l (size), each row corresponding to a diversity path
% nff = number of feedforward taps for each RAKE finger
% nbb = number of feedback taps
% delay = delay of system <= nff+length of p - 2 - nbb
% Ex = average energy of signals
% noise = noise autocorrelation vector (size L x l*nff)
% NOTE: noise is assumed to be stationary, but may be spatially correlated
% outputs:
% dfseSNR = equalizer SNR, unbiased in dB
% ****************************

siz = size(p,2);
L=size(p,1);
nu = ceil(siz/l)-1;
p = [p zeros(L,(nu+1)*l-siz)];

% error check
if nff<=0
    error('number of feedforward taps must be > 0');
end
if delay > (nff+nu-1-nbb)
    error('delay must be <= (nff+(length of p)-2-nbb)');
end
if delay < -1
    error('delay must be >= 0');
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%if length(noise)~=L*l*nff
%error('Length of noise autocorrelation vector must be L*l*nff');
%end
if size(noise,2)~=l*nff | size(noise,1)~=L
    error('Size of noise autocorrelation matrix must be L x l*nff');
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%form ptmp = [p_0 p_1 ... p_nu] where p_i=[p(i*l) p(i*l-1)... p((i-1)*l+1)
```

278
for m=1:L
    ptmp((m-1)*l+1:m*l,1) = [p(m,1); zeros((l-1),1)];
end
for k=1:nu
    for m=1:L
        ptmp((m-1)*l+1:m*l,k+1) = conj((p(m,k*l+1:-1:(k-1)*l+2))');
    end
end
ptmp;

% form matrix P, vector channel matrix
P = zeros(nff*l*L+nbb,nff+nu);

% First construct the P matrix as in MMSE-LE
for k=1:nff,
P(((k-1)*l*L+1):(k*l*L),k:(k+nu)) = ptmp;
end

% Add in part needed for the feedback
P(nff*l*L+1:nff*l*L+nbb,delay+2:delay+1+nbb) = eye(nbb);
temp= zeros(1,nff+nu);
temp(delay+1)=1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Rn = zeros(nff*l*L+nbb);
for i = 1:L
    n_t = toeplitz(noise(i,:));
    for j = 1:l:l*nff
        for k = 1:l:l*nff
            Rn((i-1)*l+(j-1)*L+1:i*l+(j-1)*L, (i-1)*l+(k-1)*L+1:i*l+(k-1)*L) = n_t(j:j+l-1,k:k+l-1);
        end
    end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Ex*P*P';
Ry = Ex*P*P' + Rn;
Rxy = Ex*temp*P';
IRy=inv(Ry);
 w_t = Rxy*IRy;

%%%Reshape the w_t matrix into the RAKE filter bank and feedback matrices
ww=reshape(w_t(1:nff*l*L),l*L,nff);
for m=1:L
    W(m,:)=reshape(ww((m-1)*l+1:m*l,1:nff),1,l*nff);
end
b=-w_t(nff*l*L+1:nff*l*L+nbb);
sigma_dfse = Ex - w_t*Rxy';
dfseSNR = 10*log10(Ex/sigma_dfse - 1);

For the previous example, some command strings that work are (with whitened-noise equivalent channel first):

279
Multichannel Transmission

Multichannel transmission in Chapters 4 and 5 is the logical extension of diversity when many inputs may share a transmission channel. In the multichannel case, each of many inputs may affect each of many outputs, logically extending the concept of intersymbol interference to other types of overlap than simple ISI. Interference from other transmissions is more generally called crosstalk. Different inputs may for instance occupy different frequency bands that may or may not overlap, or they may be transmitted from different antennas in a wireless system and so thus have different channels to some common receiver or set of receivers. Code division systems (See Chapter 5 Appendix) use different codes for the different sources. A diversity equalizer can be designed for some set of channel outputs for each and every of the input sequences, leading to multichannel transmission. In effect the set of equalizers attempts to “diagonalize” the channels so that no input symbol from any source interferes with any other source at the output of each of the equalizers. From an equalizer perspective, the situation is simply multiple instances of the diversity equalizer already discussed in this section. However, when the transmit signals can be optimized also, there can be considerable improvement in the performance of the set of equalizers. Chapters 4 and 5 (EE379C) develop these concepts.

However, the reader may note that a system that divides the transmission band into several different frequency bands may indeed benefit from a reduced need for equalization within each band. Ideally, if each band is sufficiently narrow to be viewed as a “flat” channel, no equalizer is necessary. The SNR of each of these “sub-channels” relates (via the “gap” approximation) how many bits can be transferred with QAM on each. By allocating energy intelligently, such a system can be simpler (avoiding equalization complexity) and actually perform better than equalized wider-band QAM systems. While this chapter has developed in depth the concept of equalization because there are many wide-band QAM and PAM systems that are in use and thus benefit from equalization, an intuition that for the longest time transmission engineers might have been better off never needing the equalizer and simply transmitting in separate disjoint bands is well-founded. Chapters 4 and 5 support this intuition. Progress of understanding in the field of transmission should ultimately make equalization methods obsolete. As in many technical fields, this has become a line of confrontation between those resisting change and those with vision who see a better way.
3.10 MIMO/Matrix Equalizers

This section extends basic infinite-length equalization designs to single-user MIMO systems. Chapter 2’s codes concatenate sets of these \( N = \tilde{N} = 1 \)-dimensional-real and/or \( N = \tilde{N} = 2 \)-dimensional-complex subsymbols to form “codewords” of \( \tilde{N} \geq 1 \) subsymbols/codeword or subsymbols/symbol, all selected from the same subsymbol constellation \( C \) with size \(|C|\). Effectively in this chapter’s equalization theory, a symbol is a subsymbol. Thus, the theory developed here in Chapter 3 remains applicable also to coded systems with the constellation points viewed as equally likely, so a presumption essentially that \( M = |C| \), allowing separation of Chapter 2’s coding and Chapter 3’s equalization.

Chapter 1’s general MIMO channel is \((L_y N) \times (L_x N)\), so even with \( N = 1 \) there is a multidimensional symbol with MIMO. This chapter continues to address time-domain transmission as symbol sequences of \( N = \tilde{N} = 1 \)-dimensional real or \( N = \tilde{N} = 2 \)-dimensional-complex symbols, with \( \tilde{N} = 1 \). However, the transmitted symbol will become \( L_x \) dimensional. The index for these spatial dimensions will be \( t \), while time indices will be \( k \). The index \( n \) will return in Chapter 4 and beyond as a frequency index, and continues to be used as a codeword subsymbol index with codes. That index does not appear in this section.

3.10.1 The MIMO Channel Model

The space-time MIMO channel (without loss of generality) can be viewed as being \( L_y \times L_x \) for each symbol. The input is a sequence of \( L_x \times 1 \) complex-vector-symbol inputs that then modulate the basis-function column vectors of \( L_x \times L_x \) matrix \( \Phi(t) \):

\[
\Phi(t) = \begin{bmatrix}
\varphi_{L_x, L_x}(t) & \varphi_{L_x, L_x-1}(t) & \ldots & \varphi_{L_x, 1} \\
\varphi_{L_x-1, L_x}(t) & \varphi_{L_x-1, L_x-1}(t) & \ldots & \varphi_{L_x-1, 1} \\
\vdots & \vdots & \ddots & \vdots \\
\varphi_{1, L_x} & \varphi_{1, L_x-1} & \ldots & \varphi_{1, 1}(t)
\end{bmatrix}
\tag{3.543}
\]

\[
x(t) = \Phi(t) \cdot x \tag{3.544}
\]

\( \Phi(t) \)'s columns in (3.544), \( \varphi_i(t) \) for \( l = 1, \ldots, L_x \), are \( L_x \times 1 \) basis-vector functions. There is one such basis function for each of \( x \)’s \( L_x \) components. Column basis-vector “orthonormality” holds when the channel’s spatial dimensions have no crosstalk, the spatial equivalent of no intersymbol interference – but more accurately no inter-dimensional interference. Equivalently stated, the column vectors need not be “orthogonal” to one another in time-frequency if they are non overlapping in “space.” Some vector-modulation design methods may lead to such crosstalk-free dimensions. Often, there will be nonzero crosstalk and this subsection’s MIMO equalization methods address both reduction of intersymbol interference and of crosstalk.

However, some channels may introduce crosstalk (which is really “intrasyms” interference among the spatial dimensions that at least entered the channel free of such crosstalk), which is the spatial equivalent of intersymbol interference. There can still also be ISI between the successive overlapping vector symbols transmitted. Each transmit column vector \( \varphi_i(t) \) has unit norm. A simple model of such spatially orthogonal basis-function vectors has diagonal \( \Phi(t) \), probably with the same unit-norm basis function \( \varphi(t) \) along the constant diagonal \( \Phi(t) = \varphi(t) \cdot I \). However, as later chapters show, it can be helpful to be non-diagonal if the ensuing channel has crosstalk.

With \( \Phi(t) \) having \( L_x \) normalized column basis-vector functions, the vector continuous-time transmit signal is

\[
x(t) = \sum_k \Phi(t - kT) \cdot x_k \tag{3.546}
\]

consistent with the scalar successive-transmission development earlier in Section 3.1 of this chapter\(^{25}\). The \( L_x \)-dimensional modulated signal vector simply has a component for each channel-input dimension

\(^{25}\)The use of a “.” in (3.546) simply denotes multiplication. This dot notation is used for appearance of results to
Instead of being a complex scalar ($L_x = 1$). The $L_y$-dimensional channel output is

\[ y_p(t) = H(t) * x(t) + n_p(t) , \]  

(3.547)

an $L_y \times 1$ random-process vector of noise\(^{26}\). Thus, Section 3.9’s diversity is a special case with $L_x = 1$ but $L_y > 1$. MIMO is essentially $L_x$ independent diversity channels each of size $L_y \times 1$. In effect, Section 3.9’s diversity applies to each spatial dimension independently. However, this section’s treatment is more compact and allows some insights into these diversity channels’ data-rate sum as a direct calculation and as a direct interpretation of the MIMO-channel performance for a single user. This MIMO system then has matrix pulse response

\[ P(t) = H(t) * \Phi(t) , \]  

(3.548)

and so the vector channel output (see Figure 3.64) is

\[ y_p(t) = \sum_k P(t - kT) \cdot x_k + n(t) . \]  

(3.549)

A MIMO “one-shot” system sends only the single vector symbol $x_0$, so $y_p(t) = P(t) * x_0 + n_p(t)$. However, an $L_y$-dimensional channel output consequentially occurs with possible crosstalk between the output dimensions. The corresponding ML “MIMO symbol-by-symbol” detector is not a simple slicer on each dimension unless the channel is crosstalk free (or the inputs have been carefully designed, see Chapters 4 and 5). Nonetheless, a constant (operating within the symbol only on crosstalk) MMSE equalizer of $L_x \times L_y$ coefficients could be designed for this one-shot system.

The MIMO stationary noise vector has $L_y \times L_y$ dimensional autocorrelation matrix function $R_n(t)$. Ideally, $R_{nn}(t) = \frac{N_0}{2} \cdot I \cdot \delta(t)$ so that the noise is white in all time dimensions and also uncorrelated between the different spatial dimensions, with the same variance $\frac{N_0}{2}$ on all dimensions. Such a noise autocorrelation matrix function will have a matrix Fourier transform $S_n(f) = \frac{N_0}{2} \cdot I_{L_y}$ (see Appendix A.4). When $S_n(f)$ satisfies the MIMO Paley-Wiener criterion of Appendix A.4 (and all practical noises will),

\[ \int_{-\infty}^{\infty} \frac{|\ln|S_n(f)||}{1 + f^2} \cdot df < \infty , \]  

(3.550)

then a factorization of $S_n(f)$ into a causal and causally invertible filter matrix exists (through matrix spectral factorization exists of Appendix A.4):

\[ S_n(f) = \frac{N_0}{2} \cdot S_n^{1/2}(f) \cdot S_n^{1/2}(-f) . \]  

(3.551)

$S_n^{-1/2}(f)$ is the causal and causally invertible filter matrix. This $S_n^{-1/2}(f)$ is the causal, MIMO, noise-whitening matrix filter, and the MIMO white-noise equivalent channel becomes

\[ \tilde{P}(f) = S_n^{-1/2}(f) \cdot P(f) , \]  

(3.552)

which has inverse transform $\tilde{P}(t)$. Then the noise-whitened channel output becomes

\[ y_p(t) = \sum_k \tilde{P}(t - kT) \cdot x_k + n_p(t) , \]  

(3.553)

where $n_p(t)$ is now the desired white noise with variance $\frac{N_0}{2}$ in each and every (real) dimension.

The result of the matched-filter matrix operation then is

\[ y(t) = \tilde{P}^\top(-t) * y(t) \]  

(3.554)

separate quantities that might otherwise notionally confused - it can mean matrix multiply or scalar multiply in this text, depending on context.

\(^{26}\text{AWGN means white Gaussian noise on each dimension, but also of equal power and independent of all other } L_y - 1 \text{ dimensions for every dimension.}\)
The unnormalized \( L_x \times L_x \) ISI-crosstalk characterizing matrix function will then be defined as
\[
R_p(t) \triangleq \tilde{P}^* (-t) * \tilde{P}(t) .
\] (3.555)

Then, finally the sampled channel output vector is (with \( R_{p,k} \triangleq R_p(kT) \) and noting therefore that \( R_{p,k} = R^*_{p,-k} \) or \( R_p(D) = R^*_p(D^{-*}) \))
\[
y_k = \sum_m R_{p,k-m} \cdot x_m + n_k .
\] (3.556)

The vector D-Transform (see Appendix A.4) is correspondingly
\[
Y(D) = R_p(D) \cdot X(D) + N(D) .
\] (3.557)

The matched-filter-bound (MFB) measures if a receiver’s performance is close to an upper bound (that may not always be attainable). This MFB concept makes more sense for MIMO if each dimension of a simple slicer were used in each output dimension (following a MIMO matched filter of size \( L_x \times L_y \)). Thus, the MFB is really a concept to be individually applied to each of these \( L_x \) dimensions. An \( L_x \times L_x \) diagonal matrix with the norms of individual matched-filter-output scalings on the diagonal will also be useful and is defined as
\[
\| \tilde{P} \| \triangleq \begin{bmatrix}
\| \tilde{P}_{L_x} \| & 0 & \ldots & 0 \\
0 & \| \tilde{P}_{L_x-1} \| & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \| \tilde{P}_1 \|
\end{bmatrix} .
\] (3.558)

Each of (3.558)'s diagonal norm elements \( \| P_l \|, l = 1, \ldots, L_x \) is the sum of \( L_y \) integrals, each corresponding to one of the pulse-response matrix', \( P(t)'s, L_x \) columns, so
\[
\| P_l \|^2 = \sum_{i=1}^{L_y} \int_{-\infty}^{\infty} |p_{i,l}(t)|^2 \cdot dt ,
\] (3.559)

where
\[
\tilde{P}_l(t) \triangleq \begin{bmatrix}
p_{L_y,l}(t) \\
\vdots \\
p_{1,l}(t)
\end{bmatrix} l = 1, \ldots, L_x .
\] (3.560)

There is no intersymbol interference (nor crosstalk) when \( R_{p,k} = \| \tilde{P} \|^2 \cdot \delta_k \) or \( R_p(D) = \| \tilde{P} \|^2 \), basically a MIMO Nyquist criterion. In this Nyquist-satisfying case, the pulse response is such that dimension-by-dimension detection within symbol-by-symbol detection is optimum, and each dimension has a best SNR indexed by \( l \) as
\[
\text{SNR}_{MFB}(l) = \| \tilde{P}_l \|^2 \cdot \frac{\tilde{\varepsilon}_l}{N_0} l = 1, \ldots, L_x .
\] (3.561)

The corresponding diagonal SNR matrix is
\[
\text{SNR}_{MFB} = \| \tilde{P} \|^2 \cdot \frac{\tilde{\varepsilon}_x}{N_0} .
\] (3.562)

The next subsection explores further matrix SNRs.

3.10.2 Matrix SNRs

The single-user SNR concept for MIMO/vector transmission systems remains a scalar for each dimension,\(^{27}\) \( \text{SNR}_l, l = 1, \ldots, L_x \). Unlike the previous subsection’s scalar receivers, these different \( \text{SNR}'s \)’s could

---

\(^{27}\)It is possible to do an ML \( L_x \)-dimensional symbol-by-symbol detector that chooses \( \hat{x}_k \) to minimize the quantity \( \| \tilde{z}_k - B\hat{x}_k \|^2 \) over all choices for the \( L_x \)-dimensional \( x \) where \( B \) represents the crosstalk within the symbol’s spatial dimensions. Such a detector can be complex for \( L_x > 1 \) so this Chapter does not further pursue it because the main focus is simple detectors.
lead to different signal constellations $C_l$ on each spatial dimension and thus different dimensional $\hat{b}_l$ or $M_l = |C_l|$. When the gap $\Gamma = 0$ dB, a single equivalent SNR can replace the set and corresponds to the average $\hat{b}_{ave} = \frac{1}{L_x} \sum_{l=1}^{L_x} \hat{b}_l$. This single SNR’s biased version is the geometric average the (biased) SNR set, $\text{SNR}(R) = \left\{ \prod_{l=1}^{L_x} (1 + \text{SNR}_l) \right\}^{1/L_x} - 1$, as this subsection develops. The receiver’s objective will be to maximize this SNR and thus maximize the data-rate sum (or equivalently $\hat{b}_{ave}$ over all dimensions).

Figure 3.63 expands Figure 3.2 of Section 3.2 to the MIMO case. The MMSE solution minimizes Figure 3.63’s error-vector autocorrelation-matrix determinant $|\text{Re}_e|$ as proved in Appendix A.4. Appendix A.4 also shows that minimizing this determinant minimizes trace $\{\text{Re}_e\}$ so corresponds also to minimum sum of mean-squared spatial-dimension errors. As Figure 3.63 shows, an “SNR” can thus correspond to the ratio of the input-autocorrelation-to-error-autocorrelation determinants.

Chapter 3’s transmit symbol vector has independent and equal energy per spatial dimension $\hat{E}_k$ so that $R_{xx}(D) = \hat{E}_x \cdot I$. If $R_{ee}(D)$ is also constant diagonal $^{28}$ $R_{ee}(D) = \text{Diag}\{S_{e,1}, \ldots, S_{e,1}\}$, then (with numerator and denominator autocorrelation matrices both normalized consistently to the same number of dimensions, 1 for real or 2 for complex):

$$\text{SNR}(R) = \frac{|R_{xx}(D)|}{|R_{ee}(D)|} = \prod_{l=1}^{L_x} \frac{\hat{E}_x}{\hat{E}_{e,l}} = \prod_{l=1}^{L_x} \text{SNR}_l \tag{3.563}$$

If $R_{ee}(D)$ is not diagonal, Figure 3.63’s symbol-by-symbol detector will still independently detect each of the $L_x$ dimensions, so (3.563) has consistent practical meaning, but then the dimensional SNR’s simply use the diagonal elements of $R_{ee}(D = 0)$ as the entries of $S_{e,l}$, so $\text{Diag}\{R_{ee}(0)\} = S_{e,l}$. As in Appendix A.4, the individual optimization problems are separable if the receiver uses a MMSE criterion, and each

---

$^{28}$The notation “Diag (R)” with a capital letter “D” for an $N \times N$ square matrix means a square diagonal matrix that has a diagonal matching $R$ and all zeros off-diagonal. The notation, “diag (R)” means an $N \times 1$ vector formed from the diagonal elements of $R$, so similar to the Matlab command. The notation “Diag \{a, b, …\}” will mean forming a diagonal matrix from the set of elements in the obvious way.

284
dimension is an equalizer with a bias that can be removed, so

\[ \text{SNR}_{U,t} = \text{SNR}_t - 1 \]  

(3.564)

define the unbiased SNR’s that can be obtained in the usual way with bias removal in MMSE equalization. Each SNR corresponds to one of the \( L_x \) MIMO dimensions so the data rate is the sum of the individual data rates for the entire \( L_x \)-dimensional symbol\(^{29}\). The sum of the spatial dimensions’ bits per symbol entries is then (with gap \( \Gamma = 0 \) dB)

\[ \tilde{b} = \sum_{l=1}^{L_x} \tilde{b}_l , \]  

(3.565)

recalling that the notation \( \tilde{b} \) normalizes to the number of complex (temporal) dimensions when QAM is used and to the number of real temporal dimensions when PAM is used. The quantity \( \tilde{b}_{ave} \) then normalizes to the number of spatial dimensions \( L_x \), so \( \tilde{b}_{ave} = \frac{\tilde{b}}{L_x} \).

\[ \tilde{b} = \sum_{l=1}^{L_x} \tilde{b}_l = \sum_{l=1}^{L_x} \log_2 (1 + \text{SNR}_{U,l}) \]  

(3.566)

\[ = \log_2 \prod_{l=1}^{L_x} (1 + \text{SNR}_{U,l}) \]  

(3.567)

\[ = \log_2 \left[ \frac{|RX(D)|}{|Ree(D)|} \right] , \]  

(3.568)

which also validates the utility of the ratio of determinants. Further, the average number of bits per spatial symbol defines a geometric SNR through

\[ \tilde{b}_{ave} = \log_2 (1 + \text{SNR}_{geo,u}) \]  

(3.569)

or

\[ \text{SNR}_{geo,u} = \left[ \prod_{l=1}^{L_x} (1 + \text{SNR}_{l,u}) \right]^{1/L_x} - 1 . \]  

(3.570)

**3.10.3 The MIMO MMSE DFE and Equalizers**

Figure 3.64 redraws Figure 3.6 for the case of MIMO filters. All vector processes are shown in boldface, while the matrix filters are also in boldface. Dimensions have been included to facilitate understanding. This section will follow the earlier infinite-length scalar DFE development, starting immediately with MMSE.

The feedback section eliminates intersymbol interference from previous \( L_x \)-dimensional symbols, but also eliminates some crosstalk from “earlier” decided spatial dimensions. Thus \( B_0 \) is monic and upper triangular. \( B(D) \) is not generally upper triangular, but is causal. This structure specifically permits a first decision is made on dimension 1 after all previous ISI corresponding to \( x_{k-n} \) for all \( n \geq 1 \) can be made by subtracting ISI from that sole dimension (since the feedback section is also upper triangular and then has only one non-zero coefficient in the bottom row). That first dimension’s current decision, plus all its previous decisions, can be used respectively to subtract current symbol-instance crosstalk and ISI from the bottom dimension into the next dimension up. This process continues up the triangular feedback section where the top row uses the current decisions from all lower dimensions to eliminate current crosstalk along with all previous-decision dimensions’ crosstalk and ISI from its dimension.

\(^{29}\)Or \( 2L_x \) real dimensions if the system is complex baseband, but consists of \( L_x \) complex dimensional channels.
The MIMO MMSE Decision Feedback Equalizer (M4-DFE) jointly optimizes the settings of both the infinite-length \( L_x \times L_y \) matrix-sequence feedforward filter \( W_k \) and the causal infinite-length \( L_x \times L_z \) monic (\( B_0 \) is monic and upper triangular) matrix-sequence feedback filter \( B_0 \cdot \delta_k - B_k \) to minimize the MSE. Appendix ?? develops the orthogonality principal for vector processes and MIMO.

**Definition 3.10.1 (Minimum Mean Square Error Vector (for M4-DFE))** The M4-DFE error signal is

\[
e_k = x_k - z'_k .
\]  

The MMSE for the M4-DFE is

\[
\sigma^2_{MMSE-DFE} = \Delta \min_{W(D), B(D)} E \{|Ree(D)|\} .
\]  

The \( L_x \times 1 \) vector error sequence can be written as:

\[
E(D) = X(D) - W(D) \cdot Y(D) - [I - B(D)] \cdot X(D) = B(D) \cdot X(D) - W(D) \cdot Y(D) .
\]  

For any fixed \( B(D) \), \( E [E(D)Y^*(D^*)] = 0 \) to minimize MSE, which leads to the relation

\[
B(D) \cdot R_{xy}(D) - W(D) \cdot R_{yy}(D) = 0 .
\]  

Two correlation matrices of interest are (assuming the input has equal energy in all dimensions, \( \vec{\varepsilon}_x \))

\[
R_{xy}(D) = \vec{\varepsilon}_x \cdot R_p(D)
\]  

\[
R_{yy}(D) = \vec{\varepsilon}_x \cdot R^2_p(D) + \frac{N_0}{2} \cdot R_p(D)
\]  

The MMSE-MIMO LE (M4-LE) matrix filter then readily becomes

\[
W_{M4-LE}(D) = R_{xy}(D) \cdot R_{yy}^{-1}(D)
\]  

Figure 3.64: MIMO decision feedback equalization.
where $\text{SNR} \triangleq \frac{\hat{\mathbf{x}}^*}{\hat{\mathbf{x}}^*} \cdot I$. The corresponding MMSE MIMO (M4-LE) matrix is (with $\mathbf{B}(D) = I$)

\[
S_{M4-LE}(D) = \min_{W(D)} E \left\{ RE_{M4-LE} e_{M4-LE}(D) \right\} \quad (3.582)
\]

\[
= R_{xx}(D) - R_{xy}(D) \cdot R_{yy}(D) \cdot R_{yx}(D) \quad (3.583)
\]

\[
= \hat{\mathbf{x}}^* \cdot I - \hat{\mathbf{x}}^* \cdot R_p(D) \left[ \hat{\mathbf{x}}^* \cdot R_p^2(D) + \frac{N_0}{2} \cdot R_p \right]^{-1} \cdot R_p(D) \hat{\mathbf{x}}^* \quad (3.584)
\]

\[
= \hat{\mathbf{x}}^* \cdot I - [R_p(D) + \text{SNR}^{-1}]^{-1} \cdot R_p(D) \cdot \hat{\mathbf{x}}^* \quad (3.585)
\]

\[
= \hat{\mathbf{x}}^* \cdot \left\{ I - [R_p(D) + \text{SNR}^{-1}]^{-1} \cdot R_p(D) \right\} \quad (3.586)
\]

\[
= \hat{\mathbf{x}}^* \cdot [R_p(D) + \text{SNR}^{-1}]^{-1} \cdot [R_p(D) + \text{SNR}^{-1} - R_p(D)] \quad (3.587)
\]

\[
= \frac{N_0}{2} \cdot [R_p(D) + \text{SNR}^{-1}]^{-1} \quad (3.588)
\]

\[
R_{M4-LE}(D) = \frac{\frac{N_0}{2}}{|R_p(D) + \text{SNR}^{-1}|} \quad (3.589)
\]

A form similar to the scalar MMSE-LE version would define

\[
Q(D) \triangleq \| \hat{\mathbf{P}} \|^{-1} \cdot R_p(D) \cdot \| \hat{\mathbf{P}} \|^{-1} \quad (3.590)
\]

and then

\[
W_{M4-LE}(D) = \| \hat{\mathbf{P}} \|^{-1} \cdot [Q(D) + \text{SNR}_M^{-1}]^{-1} \cdot \| \hat{\mathbf{P}} \|^{-1} \quad (3.591)
\]

where the trailing factor $\| \hat{\mathbf{P}} \|^{-1}$ could be absorbed into the preceding matrix-matched filter to “normalize” it in Figure 3.64, and thus this development would exactly then parallel the scalar result (noting that multiplication by a non-constant diagonal matrix does not in general commute). Similarly,

\[
S_{M4-LE}(D) = \frac{N_0}{2} \cdot \| \hat{\mathbf{P}} \|^{-1} \cdot [Q(D) + \text{SNR}_M^{-1}]^{-1} \cdot \| \hat{\mathbf{P}} \|^{-1} \quad (3.592)
\]

\[
R_{M4-LE}(D) = \frac{\frac{N_0}{2}}{\| \hat{\mathbf{P}} \|^2 \cdot [Q(D) + \text{SNR}^{-1}]} \quad (3.593)
\]

The MMSE, as in Appendix ??, occurs for either choice of determinant or norm (they are not equal, but both minimized) and the more convenient form for the M4-LE is the determinant:

\[
\sigma_{M4-LE}^2 = R_{M4-LE}(0) \quad (3.594)
\]

\[
= \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |R_{M4-LE}(e^{-j\omega T})| d\omega \quad (3.595)
\]

\[
= \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |R_p(e^{-j\omega T}) + \text{SNR}^{-1}| d\omega \quad (3.596)
\]

\[
= \frac{N_0}{2} \cdot W_0^{-1} \quad (3.597)
\]
For dimension-by-dimension detection, \( W_0 \) is factored as \( Q_w S_w Q_w^\dagger \) (an eigenvalue decomposition of this symmetric positive definite matrix). Then,

\[
W(D) \rightarrow Q_w^\dagger W(D)
\]

so as to diagonalize the MMSE and allow each dimension independent decisions (instead of an ML detector for each symbol over \( L_x \) dimensions). This orthogonal transformation does not change the MMSE although it may redistribute error energy over the spatial dimensions. The overall SNR per spatial dimension then is

\[
\text{SNR}_{M4-LE} = \frac{\tilde{\epsilon}_x}{\left( \prod_{l=1}^{L_x} S_w(l) \right)^{1/L_x}} = \left[ \prod_{l=1}^{L_x} (1 + \text{SNR}_{M4-LE}(l)) \right]^{1/L_x} \quad \text{(3.599)}
\]

and of course \( \text{SNR}_{M4-LE, U} = \text{SNR}_{M4-LE} - 1 \). By linearity of MMSE estimates, then the FF filter matrix for any \( B(D) \) becomes

\[
W(D) = B(D) \cdot R_{xy}(D) \cdot R_{yy}^{-1}(D) = B(D) \cdot [R_p(D) + \text{SNR}]^{-1} = B(D) \cdot \| \tilde{P} \|^{-1} \cdot \left[ Q(D) + \text{SNR}^{-1}_{MFB} \right]^{-1} \cdot \| \tilde{P} \|^{-1}
\]

for any causal \( B(D) \). Again, since the MMSE estimate is linear then

\[
E(D) = B(D) \cdot E_{M4-LE}(D)
\]

The autocorrelation function for the error sequence with arbitrary monic \( B(D) \) is

\[
R_{ee}(D) = B(D) \cdot R_{e_{M4-LE}e_{M4-LE}}(D) \cdot B^*(D^{-*})
\]

which has a form very similar to the scalar quantity. In the scalar case, a canonical factorization of the scalar \( \tilde{Q}(D) \) was used and was proportional to the (inverse of the) autocorrelation function of the MMSE of the LE, but not exactly equal. Here because matrices do not usually commute, the MIMO canonical factorization will be directly of \( R_{e_{M4-LE}e_{M4-LE}}(D) \), or via (3.592) of \( \frac{N_L}{2} \| \tilde{P} \| \cdot \left[ Q(D) + \text{SNR}^{-1}_{MFB} \right] \| \tilde{P} \| \), so

\[
R_{e_{M4-LE}e_{M4-LE}}(D)^{-1} = G(D) \cdot S_e \cdot G^*(D^{-*}) \quad \text{or}
\]

\[
\left[ Q(D) + \text{SNR}^{-1}_{MFB} \right] = \| \tilde{P} \|^{-1} \cdot G(D) \cdot S_e \cdot G^*(D^{-*}) \| \tilde{P} \|^{-1}
\]

where \( G(D) \) has \( \text{Diag} \{ G(0) \} = I \), and is causal minimum-phase (causally invertible); \( S_e \) is diagonal with all positive elements. Such a factorization exists when the MIMO PW Criterion is satisfied (as in Appendix A.4), which 3 will be the case if \( R_{xx} = \mathcal{E} \cdot I \) and \( \mathcal{E} > 0 \) and \( \frac{N_L}{2} > 0 \). Substitution of the inverse factorization into (3.604) produces

\[
R_{ee}(D) = \frac{N_0}{2} \cdot B(D) \cdot G^{-1}(D) \cdot S_e^{-1} \cdot G^{-1}(D^{-*}) \cdot B^*(D^{-*})
\]

The determinant \( |R_{ee}(D)| \) is the product of \( (\frac{N_L}{2})^{L_x} \) with the determinants of the 5 matrices above, each of which is the product of the elements on the diagonal. Since \( B(D) \) and \( G(D) \) are both causal and both \( B_0 \) and \( G_0 \) are monic upper triangular, their product \( F_0 = B_0 \cdot G_0 \) shares these same properties. The product \( F = B(D) \cdot G(D) \) has a causal monic form \( F_0 + F_1 D + F_2 D^2 + \ldots \) with \( F_0 \) monic upper triangular. The product \( F(D) S_e^{-1} F^*(D^{-*}) \) will have a weighted sum of positive terms with \( F_0 S_e^{-1} F_0^* \) as one of them (time zero-offset term) and will have all other terms as positive definite matrices \( F_k \cdot F_k^* \) and thus is minimized when when \( F_k = 0 \forall k > 0 \), so \( B(D) = G(D) \), and in particular \( B_0 = G_0 \) for crosstalk elimination, is the MMSE solution. Further the minimum value is \( R_{ee}(D) = S_e^{-1} \).

\[
R_{ee}(D) = S_e^{-1}
\]

\[
\sigma_{M4-DFE}^2 \triangleq |S_e^{-1}| \triangleq S_e^{-1}
\]

(3.609)
Lemma 3.10.1 (MMSE-DFE) The MMSE-DFE has feedforward matrix filter

\[ \mathbf{W}(D) = \mathbf{W}(D) = \mathbf{S}_e^{-1} \cdot \mathbf{G}^{-\ast}(D^{-\ast}) \]  

(3.610)

(realized with delay, as it is strictly noncausal) and feedback section

\[ \mathbf{B}(D) = \mathbf{G}(D) \]  

(3.611)

where \( \mathbf{G}(D) \) is the unique canonical factor of the following equation:

\[ \frac{N_0}{2} \cdot \| \mathbf{P} \| \cdot [\mathbf{Q}(D) + \mathbf{SNR}_{\text{MF}}^{-1}] \cdot \| \mathbf{P} \| = \mathbf{G}(D) \cdot \mathbf{S}_e \cdot \mathbf{G}^\ast(D^{-\ast}) \]  

(3.612)

This text also calls the joint matched-filter/sampler/\( \mathbf{W}(D) \) combination in the forward path of the DFE the “MIMO Mean-Square Whitened Matched Filter (M4-WMF)”. These settings for the M4-DFE minimize the MSE as was shown above in (3.609).

The matrix signal to noise ratio is

\[ \mathbf{SNR}_{\text{M4-DFE}} = \frac{|\mathbf{R}_{xx}|}{|\mathbf{S}_e|} \]  

(3.613)

\[ = \frac{\langle \mathbf{\tilde{e}}_x \rangle}{\prod_{l=1}^{L_x} \mathbf{S}_{e,l}} \]  

(3.614)

\[ = \prod_{l=1}^{L_x} \mathbf{SNR}_{\text{M4-DFE}(l)} \]  

(3.615)

and correspondingly each such individual SNR is biased and can be written as

\[ \mathbf{SNR}_{\text{M4-DFE}(l)} = \mathbf{SNR}_{\text{M4-DFE,U}(l)} + 1 \]  

(3.616)

while also

\[ \mathbf{SNR}_{\text{M4-DFE}} = \mathbf{SNR}_{\text{M4-DFE,U}} + I \]  

(3.617)

The overall bits per symbol (with zero-dB gap) is written as (with \( \bar{l} = 2 \) when the signals are real baseband, and \( \bar{l} = 1 \) when complex baseband)

\[ b = \frac{1}{\bar{l}} \cdot \log_2 \left\{ \frac{|\mathbf{R}_{xx}|}{|\mathbf{S}_e|} \right\} \]  

(3.618)

\[ = \frac{1}{\bar{l}} \cdot \sum_{l=1}^{L_x} \log_2 \left[ 1 + \mathbf{SNR}_{\text{M4-DFE,U}(l)} \right] \]  

(3.619)

and with non-zero gap,

\[ \tilde{b} = \sum_{l=1}^{L_x} \cdot \log_2 \left[ 1 + \frac{\mathbf{SNR}_{\text{M4-DFE,U}(l)}}{I} \right] \]  

(3.620)

The reader might note that equal energy on each dimension probably is not optimum for transmission. Chapter 4 studies such optimization or loading.

The process of current symbol previous-order crosstalk subtraction can be explicitly written by defining the ISI-free (but not current crosstalk-free) output

\[ z'_k = z_k - \sum_{l=1}^{\infty} \mathbf{G}_l \cdot \mathbf{\hat{x}}_{k-l} \]  

(3.621)

Any coefficient matrix of \( \mathbf{G}(D) \) can be written \( \mathbf{G}_k \) with its \((m,n)^{th}\) element counting up and to the left from the bottom right corner where \( m \) is the row index and \( n \) is the column index, and \( m = 0, ..., L_x \)
and also \( n = 0, ..., L_x \). Then, the current decisions can be written as (with \( \text{sbs} \) denoting simple one- (or two- if complex) dimensional slicing detection):

\[
\hat{x}_{0,k} = \text{sbs}_0 (z'_{0,k}) \quad (3.622)
\]

\[
\hat{x}_{1,k} = \text{sbs}_1 (z'_{1,k} - G_0(1, 0) \cdot \hat{x}_{0,k}) \quad (3.623)
\]

\[
\hat{x}_{2,k} = \text{sbs}_2 (z'_{2,k} - G_0(2, 1) \cdot \hat{x}_{1,k} - G_0(2, 0) \cdot \hat{x}_{0,k}) \quad (3.624)
\]

\[
\vdots = \vdots \quad (3.625)
\]

\[
\hat{x}_{L_x-1,k} = \text{sbs}_{L_x-1} (z'_{L_x-1,k} - \sum_{l=0}^{L_x-2} G_0(L_x - 1, l) \cdot \hat{x}_{l,k}) \quad . (3.626)
\]

**3.10.4 MIMO Precoding**

Tomlinson Precoders follow exactly as they did in the scalar case with the \( \text{sbs} \) above in Equations (3.622) - (3.626) being replaced by the modulo element in the transmitter. The signal constellation on each dimension may vary so the modulo device corresponding to that dimension corresponds to the signal constellation used. Similar Laroia/Flexible precoders also operate in the same way with each spatial dimension having the correct constellation. To form these systems, an unbiased feedback matrix filter is necessary, which is found by computing the new unbiased feedback matrix filter

\[
G_U(D) = [\text{SNR}_{M4-DFE} - I]^{-1} \cdot \text{SNR}_{M4-DFE} [G(D) - \text{SNR}_{M4-DFE}]^{-1} \quad . (3.627)
\]

\( G_U(D) \) may then replace \( G(D) \) in feedback section with the feedforward matrix-filter output in Figure 3.64 scaled up by \( [\text{SNR}_{M4-DFE} - I]^{-1} \cdot \text{SNR}_{M4-DFE} \). However, the precoder will move the feedback section using \( G_U(D) \) to the transmitter in the same way as the scalar case, just for each dimension.

**3.10.5 MIMO Zero-Forcing**

The MIMO-ZF equalizers can be found from the M4 Equalizers by letting \( \text{SNR} \to \infty \) in all formulas and figures so far. However, this may lead to infinite noise enhancement for the MIMO-ZFE or to a unfactorable \( R_{ee}(D) \) in the ZF-DFE cases. When these events occur, ZF solutions do not exist as they correspond to channel and/or input singularity. Singularity is otherwise handled more carefully in Chapter 5. In effect the MIMO Channel must be “factorizable” for these solutions to exist, or equivalently the factorization

\[
|\hat{P}| \cdot Q(D) \cdot \|\hat{P}\| = P_c(D) \cdot S_c \cdot P^*_c(D^{--}) \quad , (3.628)
\]

exists. This means basically that \( Q(D) \) satisfies the MIMO Paley-Weiner Criterion in Appendix A.4.
3.11 Information Theoretic Approach to Decision Feedback

Subsection 3.2.1’s signal-to-noise ratio and Section 3.6’s filter settings for the MMSE-Decision Feedback Equalizer also follow from some basic information-theoretic results for Gaussian sequences. Chapter 2, and also Appendix ?? define information measures that this section uses to establish an information-theoretic approach canonical equalization. Subsection 3.11.1 revisits the basic information measures of entropy and mutual information for Gaussian sequences, particularly that MMSE estimation is fundamentally related to conditional entropy for jointly Gaussian stationary sequences.

Combination of the MMSE results with the information measures then simply relate the “CDEF” result of decision feedback equalization in Section 3.11.2, which suggests that good codes in combination with an appropriate (set of) DFE(s) can reliably allow the highest possible transmission rates, even though the receiver is not MAP/ML. Such a result is surprising as equalization methods as originally conceived are decidedly suboptimal, but with care of the input spectrum choice can be made to perform at effectively optimum, or “canonical” levels.

3.11.1 MMSE Estimation and Conditional Entropy

Given two complex jointly Gaussian random variables, \( x \) and \( y \), the conditional probability density \( p_{x/y} \) is also a Gaussian density and has mean \( E[x/y] \) equal to the MMSE estimate of \( x \) given \( y \). (This result is easily proved as an exercise by simply taking the ratio of \( p_{x,y} / p_y \), which are both Gaussian with the general \( N \)-complex-dimensional \((2N \) real dimensions\)) form. The (complex) Gaussian vector distribution is \( 1/(\pi^N |R_{xx}|) \cdot e^{-\langle x-u_x \rangle R_{xx}^{-1} \langle x-u_x \rangle^*} \), where \( R_{xx} \) is the covariance matrix and \( u_x \) is the mean.)

The entropy (see Chapter 2) of a complex Gaussian random variable is

\[
H_x = \log_2(\pi e E_x) \quad (3.629)
\]

where \( E_x \) is the mean-square value of \( x \). Thus, the conditional entropy of \( x \) given \( y \) is

\[
H_{x/y} = \log_2(\pi e \sigma_{x/y}^2) \quad (3.630)
\]

where \( \sigma_{x/y}^2 \) is the mean-square value of \( x \) given \( y \) because \( p_{x/y} \) is also a Gaussian distribution. Entropy can be normalized to the number of real dimensions, and complex random variables in passband transmission are designed to have the same variance in both real and imaginary dimensions, which is \( 1/2 \) the value of the complex variance. Thus, \( H_x = \frac{1}{2} \cdot \log_2(2\pi e E_x) \) whether \( x \) is real or complex. These results generalize to jointly \( \bar{N} \)-complex-dimensional non-singular\(^{30} \) Gaussian random vectors (so \( N = 2\bar{N} \) real dimensions) as

\[
H_x = \log_2\{(\pi e)^{\bar{N}} |R_{xx}|\} \quad (3.631)
\]

and

\[
H_{x/y} = \log_2\{(\pi e)^{\bar{N}} |R_{x/y}|\} \quad (3.632)
\]

respectively, where \( R_{x/y} = R_{xx} - R_{xy} R_{yy}^{-1} R_{yx} \) is the autocorrelation matrix of the error associated with the vector MMSE estimate of \( x \) from \( y \). As always \( R_{xx} = \frac{1}{T} \cdot R_{xx} \). Again, per-dimensional quantities are found by dividing by the number of real dimensions \( \bar{N} = 2\bar{N} \): \( H_x = \frac{1}{2\bar{N}} \cdot \log_2\{(\pi e)^{\bar{N}} |R_{xx}|^{1/\bar{N}}\} \). If \( x = x \), i.e., a scalar, then \( R_{xx} = 2E_x \) in the entropy formula with \( \bar{N} = 1 \) and all per-dimensional results are consistent.\(^{31} \)

\(^{30}\)The Gaussian probability distribution for any singular dimension would essentially be independent with probability 1 and thus eliminate itself from the overall probability density in practical terms [factorizing out 1], leaving only those dimensions that are random, so the entropy in the singular case would simply be the entropy of the non-singular dimensions or products of the non-zero eigenvalues of the matrix \( R_{xx} \), which here is simply denoted as \( |R_{xx}| \).

\(^{31}\)For the interested in alternative expressions (that provide the same entropy): If \( x \) is real, then \( H_x = \frac{1}{2} \cdot \log_2\{(2\pi e)^N |R_{xx}|\} \) or

\[
H_x = \frac{1}{2\bar{N}} \cdot \log_2\{(2\pi e)^{\bar{N}} |N R_{xx}|\} \quad (3.633)
\]

\[
H_{x/y} = \frac{1}{2} \cdot \log_2\{(2\pi e)^{\bar{N}} |R_{xx}|^{1/\bar{N}}\} \quad (3.634)
\]
For stationary Gaussian sequences, the chain rule of entropy allows computation of the entropy of a Gaussian sequence. The chain rule for a Gaussian random vector \( \mathbf{x} = [x_k, x_{k-1}, \ldots, x_0] \) is

\[
H(\mathbf{x}) = H(x_k/x_{k-1}, \ldots, x_0) + H(x_{k-1}/x_{k-2}, \ldots, x_0) + \ldots + H(x_1/x_0) + H(x_0) = \sum_{n=0}^{k} H(x_n/x_{n-1}, \ldots, x_0).
\]  
(3.638)

\( (H(\mathbf{x})_n \triangleq H_{x_n}) \) The first \( k \) terms in the sum above are conditional entropies that each equal the logarithm of \( \pi e \) times the MMSE associated with prediction of a component of \( \mathbf{x} \) based on its past values. The entropy of a stationary Gaussian sequence is defined by the limit of the entropy of the vector \([x_k, \ldots, x_0]\), normalized to the number of dimensions:

\[
H_X(D) = \lim_{k \to \infty} \frac{1}{k+1} \sum_{n=0}^{k} H(x_n/x_{n-1}, \ldots, x_0) \text{ bits/complex dimension.}
\]  
(3.639)

For an infinite-length stationary sequence, essentially all the terms in the sum above must be the same, so

\[
H_X(D) = \log_2 ([\pi e]S_x) ,
\]  
(3.640)

where \( S_x \) is the MMSE associated with computing \( x_k \) from its past, which MMSE linear prediction can be implemented as a monic causal filter:

\[
V(D) = A(D) \cdot X(D) \text{ where } A(D) = 1 + a_1 D + a_2 D^2 + \ldots
\]  
(3.641)

so the product corresponds to

\[
v_k = x_k + a_1 x_{k-1} + a_2 x_{k-2} + \ldots.
\]  
(3.642)

Then, the mean-square prediction error is \( E[|v_k|^2] \), which is the time-zero value of the autocorrelation function

\[
R_{vv}(D) = A(D) \cdot R_{xx}(D) \cdot A^*(D^{-*}) .
\]  
(3.643)

For Gaussian processes, the MMSE estimate is linear, and is best found by canonical factorization of its autocorrelation function:

\[
R_x(D) = S_x \cdot G_x(D) \cdot G_x^*(D^{-*}) ,
\]  
(3.644)

where \( S_x \) is a positive constant and \( G_x(D) \) is causal, monic, and minimum phase. The time-zero value of \( R_{vv}(D) \) is then found as

\[
E[|v_k|^2] = S_x \cdot \|A/G_x\|^2 ,
\]  
(3.645)

which is minimized for monic causal choice of \( A \) when \( A(D) = 1/G_x(D) \). This MMSE linear prediction filter is then shown in the lower filter in Figure 3.11.1. \( S_x \) is the MMSE. The output process \( V(D) \) is generated by linear prediction, and is also Gaussian and sometimes called the innovations sequence. This process carries the essential information for the process \( X(D) \), and \( X(D) \) can be generated causally from \( V(D) \) by processing with \( G_x(D) \) to shape the power spectrum, alter the power/energy, but not change the information content of the process, as in Figure 3.11.1. When \( X(D) \) is white (independent and identically distributed over all time samples \( k \)), it equals its innovations.

\[
\text{which checks with one-dimensional formula. If } \mathbf{x} \text{ is complex, then } H(\mathbf{x}) = \log_2 \left( |(\pi e)^N \cdot |R_{\mathbf{x}\mathbf{x}}| \right) \text{ or}
\]

\[
H(\mathbf{x}) = \frac{1}{2N} \log_2 \left( |(\pi e)^N \cdot |2N \cdot R_{\mathbf{x}\mathbf{x}}| \right) \]  
(3.635)

\[
= \frac{1}{2N} \log_2 \left( |(\pi e)^N \cdot (2N)^N \cdot |R_{\mathbf{x}\mathbf{x}}| \right) 
\]  
(3.636)

\[
= \frac{1}{2} \log_2 \left( (2N\pi e)^N \cdot |R_{\mathbf{x}\mathbf{x}}|^{1/N} \right) 
\]  
(3.637)

which also checks with the one dimensional formulae. When a complex vector is modeled as a doubly-dimensional real vector, one can see the two formulae for normalized entropy are the same as they should be.

\( \text{presuming Appendix A’s Paley-Wiener Criterion is satisfied for the process.} \)
From Chapter 2, the **mutual information** between two random variables $x$ and $y$ is

$$I(x; y) \triangleq H_x - H_{x/y} \tag{3.646}$$

$$= \log_2 \left( \frac{S_x}{\sigma^2_{x/y}} \right) \tag{3.647}$$

$$= \log_2 \left( \frac{S_y}{\sigma^2_{y/x}} \right) \tag{3.648}$$

showing a symmetry between $x$ and $y$ in estimation, related through the common SNR that characterizes MMSE estimation,

$$1 + \text{SNR}_{\text{mmse,u}} = \frac{S_x}{\sigma^2_{x/y}} = \frac{S_y}{\sigma^2_{y/x}} \tag{3.650}$$

Equation 3.647 uses the unbiased SNR. For an AWGN, $y = x + n$, $S_y = \mathcal{E}_x + \sigma^2 = \mathcal{E}_x + \sigma^2_{y/x}$. Since $\text{SNR} \triangleq \frac{\mathcal{E}_x}{\sigma^2}$, then Equation 3.650 relates that

$$1 + \text{SNR}_{\text{mmse,u}} = 1 + \text{SNR} \tag{3.651}$$

and thus

$$\text{SNR}_{\text{mmse,u}} = \text{SNR} \tag{3.652}$$

Thus the unbiased SNR characterizing the forward direction of this AWGN channel is thus also equal to the unbiased SNR ($\text{SNR}_{\text{mmse,u}}$) in estimating the backward channel of $x$ given $y$, a fact well established in Subsection 3.2.2. The result will extend to random vectors where the variance quantities are replaced by determinants of covariance matrices as in the next subsection.

### 3.11.2 The relationship of the MMSE-DFE to mutual information

In data transmission, the largest reliably transmitted data rate for a given input sequence covariance/spectrum is the mutual information between the sequence and the channel output sequence (see
Chapter 2). This result presumes maximum-likelihood detection after observing the entire output sequence \( Y(D) \) for the entire input sequence \( X(D) \). For the ISI channel, this mutual information is

\[
\bar{I}(X(D);Y(D)) = \bar{H}_{X(D)} - \bar{H}_{X(D)/Y(D)} .
\]  

(3.653)

The entropy of a stationary Gaussian sequence is determined by the innovations process, or equivalently its MMSE estimate given its past, thus (3.653) becomes

\[
\bar{I}(X(D);Y(D)) = \bar{H}_{x_k/x_{k-1},...} - \bar{H}_{x_k/Y(D),x_{k-1},...} = \frac{1}{2} \cdot \log_2(\pi e S_x) - \frac{1}{2} \cdot \log_2(\pi e \sigma_{MMSE-DFE}^2)
\]

(3.654)

\[
= \frac{1}{2} \cdot \log_2(\text{SNR}_{MMSE-DFE})
\]

(3.655)

\[
= \frac{1}{2} \cdot \log_2(1 + \text{SNR}_{MMSE-DFE,U})
\]

(3.656)

\[
= \frac{1}{2} \cdot \log_2(1 + \text{SNR}_{MMSE-DFE,U})
\]

(3.657)

The main observation used in the last 3 equations above is that the conditional entropy of \( x_k \), given the entire sequence \( Y(D) \) and the past of the sequence \( x_k \) exactly depends upon the MMSE estimation problem that the MMSE-DFE solves. Thus the variance associated with the conditional entropy is then the MMSE of the MMSE-DFE. This result was originally noted by four authors\(^\text{33}\) and is known as the CDEF result (the CDEF result was actually proved by a more circuitous route than in this chapter, with the shorter proof being first shown to the author by Dr. Charles Rohrs of Tellabs Research, Notre Dame, IN).

**Lemma 3.11.1 (CDEF Result)** The unbiased SNR of a MMSE-DFE is related to mutual information for a linear ISI channel with additive white Gaussian noise in exactly the same formula as the SNR of an ISI-free channel is related to the mutual information of that channel:

\[
\text{SNR}_{MMSE-DFE,U} = 2^{\bar{I}} - 1 ,
\]

(3.658)

where the mutual information \( I \) is computed assuming jointly Gaussian stationary \( X(D) \) and \( Y(D) \), and only when the MMSE-DFE exists.

**Proof:** Follows development above in Equations (3.654)-(3.657). QED.

The CDEF result has stunning implications for transmission on the AWGN channel with linear ISI: It essentially states that the suboptimum MMSE-DFE detector, when combined with the same good codes that allow transmission at or near the highest data rates on the ISI-free channel, will attain the highest possible data rates reliably. This result will be the same as Shannon’s infinite-dimensional MT result of Subsection 3.12.2. In all cases, the form of the mutual information used depends on Gaussian \( X(D) \). Since a Gaussian \( X(D) \) never quite occurs in practice, all analysis is approximate to within the constraints of a finite non-zero gap, \( \Gamma > 0 \) dB as always in this Chapter, and one could write \( \bar{b} = \frac{1}{2} \log_2 \left( 1 + \frac{\text{SNR}_{MMSE-DFE,U}}{1} \right) \leq \bar{I} \) when the DFE exists. The equality holds when the gap is 0 dB\(^\text{34}\).

Figure 3.11.2 depicts the estimation problem associated with \( I(X(D);Y(D)) \). This particular interpretation shows that the MMSE-DFE structure, when it exists, can be used to approach capacity with \( \Gamma \to 0 \) dB with the same codes that are used to approach capacity on the AWGN. This implies the MMSE-DFE could then be canonical, especially if the optimum energy distribution were used by the transmitter to maximize \( \bar{I}(X(D);Y(D)) \) to capacity.


\( ^{34}\) Error propagation for non-zero gap codes can cause deviation
Figure 3.66: MMSE estimation problem associated with $I(X(D); Y(D))$ – the same as the MMSE-DFE.

Figure 3.67: CDEF canonical equivalent AWGN channel to MMSE-DFE where codes with gap $\Gamma$ from capacity on the AWGN channel can be re-used to have the same gap to capacity on the ISI channel.
Definition 3.11.1 (Canonical Performance) The performance of a transmission design is said to be canonical if the signal-to-noise ratio of the equivalent AWGN characterizing the system is $2^{2I} - 1$ when the gap is $\Gamma = 0$ dB.

However, there are restrictions on the above CDEF result that were not made explicit to simplify the developments. In particular, the designer must optimize the transmit filter for the MMSE-DFE to get the highest mutual information. This process can, and almost always does, lead to unrealizable filters. Happily, there are solutions, but the resulting structures are not the traditional MMSE-DFE except in special cases. Subsections 3.12.2 and 3.12.3 study the necessary modifications of the DFE structure.

### 3.11.3 Canonical Channel Models

The symmetry of mutual information between $X(D)$ and $Y(D)$ suggest two interpretations of the relationship between $X(D)$ and $Y(D)$, known as the canonical channel models of Figure 3.11.3. The channel autocorrelation is $r(t) \triangleq p(t) \ast p^*(-t)$ (where any minor transmit and receive analog filtering has been absorbed into the channel impulse/pulse response, while any innovations filtering remains separate), and $R(D) = \sum_k r(kT) \cdot D^k$. $Y(D)$ again corresponds to the overall channel shaping at the sampled matched-filter output.

Definition 3.11.2 (Canonical Channel) A canonical channel is one in which the linear function or matrix characterizing ISI (or more generally cross-dimensional interference as occurs from Sections 3.11.3 onward) is equal to the autocorrelation function (or matrix) of the additive independent interference. Interference can be Gaussian noise, or some combination of additive Gaussian noise and residual ISI.

Canonical channels can lead to canonical performance with the proper design of receiver. There are two canonical channels of interest in this chapter:

![Diagram of Canonical Channel models for same mutual information.](image)
The **forward canonical model** has
\[
Y(D) = R(D) \cdot X(D) + N'(D) \quad ,
\]
where \(N'(D)\) is the Gaussian noise at the output of the matched filter with autocorrelation function \(\frac{N_0}{2} R(D)\). Thus the noise power spectral density and the channel filtering have the same shape \(R(e^{-\omega T})\).

The first term on the right in \((3.659)\) is the MMSE estimate of \(Y(D)\) given \(X(D)\) and \(N'(D)\), the MMSE, and \(X(D)\) are independent. The **backward canonical model** is
\[
X(D) = W(D) \cdot Y(D) + E(D) \quad .
\]

where (modulo scaling by \(\|p\|^{-1}\) \(W(D)\) is the MMSE-LE, and where \(E(D)\) is the MMSE error sequence for the MMSE-LE. The first term on the right in \((3.660)\) is the MMSE estimate of \(X(D)\) given \(Y(D)\). The shape of the equalizer and the error-sequence’s power spectral density are both \(\bar{R}_{ee}(e^{-\omega T})\), since \(\bar{R}_{ee}(D) = \frac{N_0}{2} / (\|p\|^2 \cdot Q(D) + \frac{N_0}{2} / (|Q|g\|p\|^2)) = \frac{N_0}{2} \cdot W(D)\). Canonical channels always exhibit a response with the same shaping as the power spectral density of the additive Gaussian noise.

It is relatively simple to construct DFE’s from canonical models: For the forward model, the receiver processes \(Y(D)\) by the inverse of the anticausal spectral factor of \(R(D) = (S_r \cdot G_r(D) \cdot G_r^*(D^{-*}))\) to obtain
\[
\frac{Y(D)}{S_r \cdot G_r^*(D^{-*})} = G_r(D) \cdot X(D) + N''(D) \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \ quad
3.12 Construction of the Optimized DFE Input

The CDEF result shows a relationship between mutual information $\bar{I}(X(D); Y(D))$ and $\text{SNR}_{\text{MMSE-DFE,U}}$ for an AWGN equivalent of the ISI-channel that is the same as the relationship between $\bar{I}(x; y)$ and SNR as for the (ISI-free) AWGN channel. Thus, these same good codes that bring performance to within gap $\Gamma$ of capacity on the AWGN channel can then be applied (ignoring error propagation\textsuperscript{35}) to a MMSE-DFE system to achieve the same gap from the ISI-channel’s capacity since it too looks like an AWGN with the capacity-achieving SNR.

To this point, the MMSE-DFE has used an i.i.d. input sequence $x_k$, which does not usually maximize $\bar{I}(X(D); Y(D))$. Maximization of $\bar{I}(X(D); Y(D))$ in Subsection 3.12.2 develops a best “water-filling (WF)” spectrum. This WF power spectrum maximizes $\bar{I}(X(D); Y(D))$. A designer might then assume that MMSE-DFE transmit-filter construction with the optimum power-spectral density would then maximize $\text{SNR}_{\text{MMSE-DFE,U}} = 2^{2\bar{I}(X(D); Y(D))} - 1$. This is correct if the measure of the optimum spectrum’s band is equal to the Nyquist bandwidth, i.e. $|\Omega_{\text{opt}}| = 1/T$, or equivalently, all frequencies (except a countable number of infinitesimally narrow notches) must be used by water-filling. This rarely occurs by accident\textsuperscript{36}, and so the designer must be careful in selecting the symbol rates and carrier frequencies of a minimum-size set of QAM/MMSE-DFE channels that can be made almost equivalent\textsuperscript{37}.

It is important to distinguish coding, which here is restricted to mean the use of codes like trellis codes, block codes, turbo codes, etc. (see Chapters 10 and 11 for detailed coding development) on an AWGN channel from the concept of spectrum design (sometimes often also called “coding” in the literature in a more broad use of the term “coding”). This chapter focuses on spectrum or input design and presumes use of good known codes for the AWGN channel in addition to the designed spectrum. The two effects are here made independent for the infinite-length MMSE-DFE. This section investigates the designed spectra or input to channels. Chapter 2 investigates code use.

This section addresses transmit-signal spectrum optimization, which finds best settings for the transmit filter(s), symbol rate(s), and carrier frequency(les). Subsection 3.12.1 begins with a review of the discrete-time Paley-Wiener (PW) criterion (see Appendix A.3), which is necessary for a canonical factorization to exist for a random sequence’s power spectral density. With this PW criterion in mind, Subsection 3.12.2 maximizes $\text{SNR}_{\text{MMSE-DFE,U}}$ over the transmit filter to find a desired transmit power spectral density, which has a “water-fill shape” when it exists. Subsections 3.12.3 then studies the choice of symbol rates, and possibly carrier frequencies, for a continuous-time channel so that the PW criterion will always be satisfied. Such optimization in these 3 subsections for different increasingly more general cases enables implementation of a countable set realizable transmit filters in each disjoint continuous-time/frequency water-filling band of nonzero measure. Subsection 3.12.5 culminates in the definition of the transmit-optimized MMSE-DFE, as a set of MMSE-DFEs.

3.12.1 The Discrete-Time Paley-Wiener Criterion (PWC) and Filter Synthesis

From Appendix A.3, the discrete-time PWC is\textsuperscript{38}

\begin{quote}
\textbf{Theorem 3.12.1 (Discrete-Time Paley Wiener)} The canonical factorization
\[ R_{xx}(D) = S_x \cdot G_x(D) \cdot G_x^*(D^{-*}) \]
of the power spectral density of a stationary random sequence $x_k$ exists if and only if
\[ \int_{-\pi/T}^{\pi/T} |\ln R_{xx}(e^{-j\omega T})| d\omega < \infty \quad . \]  
\end{quote}

\textsuperscript{35}To ensure avoidance of error-propagation with a code having $\Gamma > 0$ dB, a full maximum-likelihood detector on the entire sequence that would compare the channel-output sequences against all possible noise-free channel-filtered versions of the known possible transmitted codeword sequences. This would ensure the same performance as no error propagation and thus achieve the gap-reduced capacity for this channel at the target $P_e$.

\textsuperscript{36}Some designers of single-carrier QAM systems often err in assuming the full-spectrum-measure use does occur by accident only to find their system does not perform as expected.

\textsuperscript{37}Almost“ because error propagation, or roughly equivalently precoder loss, always reduce even the best-designed DFE systems’ performance slightly.
Proof: The proof appears in Appendix ??.

From Section 3.11, the factor \( G_x(D) \) is monic, causal, and minimum phase, and there exists a white innovations process \( v_k \) with energy per sample \( \mathbb{E}_x = S_x \) and autocorrelation function \( R_{vv}(D) = S_x \).

The process \( X(D) \) can be constructed from \( V(D) \) according to Figure 3.11.1 as

\[
X(D) = V(D) \cdot G_x(D).
\]

(3.666)

Any process so constructed can be inverted to get the innovations process as also shown in Figure 3.11.1 as (even when \( G_x(D) \) has zeros on the unit circle or equivalently the spectrum has a countable number of “notches”)

\[
V(D) = \frac{X(D)}{G_x(D)}.
\]

(3.667)

as also shown in Figure 3.11.1. There is a causal and causally invertible relationship between \( X(D) \) and \( V(D) \) that allows realizable synthesis or filter implementation relating the two random sequences. Random sequences with power spectra that do not satisfy the PW criterion can not be so constructed from a white innovations sequence. As shown in Subsection 3.11.1, the innovations sequence determines the information content of the sequence. In data transmission, the innovations sequence is the message sequence to be transmitted.

For the MMSE-DFE, the sequence \( X(D) \) was white, and so equals its innovations, \( X(D) = V(D) \).

**EXAMPLE 3.12.1 (White Sequence)** A real random sequence taking equiprobable values \( \pm 1 \) with \( R_{xx}(D) = \mathbb{E}_x = 1 \) satisfies the Paley Wiener Criterion trivially and has innovations equal to itself \( X(D) = V(D) \), \( G_x(D) = 1 \). The entropy would also be \( H_x = 1 \) for this binary sequence. As in Appendix ??, if this sequence is instead selected from a Gaussian distribution, then the factorizations remain the same, but this entropy is \( H_x = .5 \log_2(2\pi \cdot e \cdot \mathbb{E}_x) \) bits/dimension, and \( H_x = \log_2(\pi \cdot e \cdot S_x) \) bits per symbol if this random sequence is complex. The spectrum is shown in Figure 3.11.1(a).

**EXAMPLE 3.12.2 (AMI Sequence)** The AMI sequence was discussed early in Subsection 3.8.3 on partial-response channels. The AMI encoder transmits successive differences of the actual channel inputs (this “encoder” can happen naturally on channels that block DC so the encoder is in effect the channel prior to the addition of WGN). This action can be described or approximated by the D-transform \( 1 - D \). The consequent AMI sequence then has \( R_{xx}(D) = -1D^{-1} + 2 - D \) clearly factors into \( R_{xx}(D) = 1 \cdot (1 - D)(1 - D^{-1}) \) and thus must satisfy the Paley Wiener Criterion. The innovations sequence is \( V(D) = X(D)/(1 - D) \) and has unit energy per sample.\(^{38}\) The energy per sample of \( X(D) \) is \( r_{xx,0} = 2 \). For the binary case, the entropy is that of the input \( v(D) \), which is 1 bit/dimension if \( v_k = \pm 1 \). For the Gaussian case, the entropy is \( H_x = .5 \log_2(2\pi \cdot e \cdot \mathbb{E}_x) \) bits/dimension – the same entropies as in Example 3.12.1, even though the spectrum in Figure 3.12.1(b) is different.

The last example illustrates that the innovations never has energy/sample greater than the sequence from which it is derived.

**EXAMPLE 3.12.3 (Ideal Lowpass Process)** A sequence with \( \zeta < 1 \) is a positive constant

\[
R_{xx}(e^{-j\omega T}) = \begin{cases} 
1 & |\omega| < \frac{\zeta \pi}{T} \\
0 & |\omega| \geq \frac{\zeta \pi}{T}
\end{cases}
\]

(3.668)

does NOT satisfy the PW criterion and has power-spectral density shown in Figure 3.12.1(c). This process cannot be realized by passing a white information sequence through a “brickwall

\(^{38}\)This filter is marginally stable in that a pole is on the unit circle - however, the input to it has zero energy at this location (DC).
Figure 3.69: Illustration of two spectra that satisfy discrete PW in (a) and (b) while (c) does not satisfy the PW criterion.
lowpass filter” because that filter is always noncausal even with arbitrarily large delay in implementation. An approximation of such a filter with a causal filter, which would lead to a slightly different power spectral density, could satisfy the PW criterion. The designer might better consider a new sampling rate of $\zeta/T$, and the resulting new sampled random sequence would then truly satisfy the PW criterion trivially as in Example 3.12.1, rather than incur the complexity of such a lowpass filter.

When optimizing a transmit complex transmit filter, an adjustment to the real basis function of Chapter 1 needs to occur. The reader may recall that the QAM basis functions from Chapter 1 were of the form:

$$\varphi_1(t) = \sqrt{2/T} \cdot \varphi(t) \cdot \cos(\omega_c t)$$

$$\varphi_2(t) = -\sqrt{2/T} \cdot \varphi(t) \cdot \sin(\omega_c t)$$  

(3.669)  

(3.670)

where $\varphi(t)$ was real. Optimization of the spectrum in this case reduces to optimization of $|\Phi(e^{-j\omega T})|^2 = \mathcal{F}\{\varphi(t) \ast \varphi^*(-t)|_{t=kT}\}$ and thus insensitive to phase as far as performance is concerned so while complex basis functions with unequal imaginary and real parts could also be found, there is always one basis function with constant amplitude on both inphase and quadrature and has the magnitude of the optimum spectra and any phase desired.

### 3.12.2 Theoretical multi-tone system tool for transmit optimization

Multi-Tone (MT) transmission systems are discussed at length in Chapter 4, including the heavily used and optimum Discrete MultiTone (DMT) for stationary channels and Coded-OFDM methods for statistically model channels (statistically modeled channels being more completely addressed also in Chapter 1). To progress the optimization of transmit filters in equalized system, it is convenient here to model a theoretically optimum MT system that was first addressed by Shannon in the extension of his basic information-theoretic bounds to ISI channels. DMT approximates very closely this theoretical system in practical implementations (as in Chapter 4).

The MT system uses a set of $N \to \infty$ infinitesimally narrow QAM channels (and one PAM channel at baseband/DC). The MT system independently excites each tone as if it were its own isolated AWGN channel with the gain being the channel gain

$$H_n \triangleq H(\omega)|_{\omega=2\pi n/T} \quad \forall n = -\infty, ..., 0, 1, \ldots \infty$$  

(3.671)

where $T$ is the symbol rate and $n/T$ is the carrier frequency where $T \to \infty$. The $n^{th}$ QAM system uses the band $f \in \frac{n}{T} + [\frac{-1}{2T}, \frac{1}{2T}]$. The limit means that the carriers get infinitesimally narrow and the symbol length becomes all time. Also $g(\omega) \triangleq H(\omega)/\sigma^2$. The transmit filter gain at the frequency to be optimized is

$$\Phi_n \triangleq \Phi(\omega)|_{\omega=2\pi n/T}$$  

(3.672)

The noise power spectral density is $\sigma^2 = N_0/2$ at all frequencies (or the equivalently white-noise channel can replace $H$). The transmit filter is presumed to have no energy outside its band, so

$$\Phi_n(\omega) = \begin{cases} 0 & \omega \notin 2\pi \cdot (\frac{n}{T} + [-\frac{1}{2T}, \frac{1}{2T}]) \\ \Phi_n \neq 0 & \omega \in 2\pi \cdot (\frac{n}{T} + [-\frac{1}{2T}, \frac{1}{2T}]) \end{cases}$$

(3.673)

so that all the tones’ subchannels are independent.

The interval of non-zero-energized frequencies may be a set of continuous frequencies, or several such sets, and is denoted by $\Omega$. The measure of $\Omega$ is the total bandwidth used

$$|\Omega| = \int_{\Omega} d\omega$$  

(3.674)
An optimum bandwidth $\Omega^{opt}$ will then be that corresponding to the subchannels used in optimization as $T \to \infty$. The data rate is
\[
R = \lim_{T \to \infty} \frac{b}{T} .
\]  
(3.675)

Continuous frequency can then replace the frequency index $n$ according to
\[
\omega = 2\pi \cdot \lim_{T \to \infty} \frac{n}{T} , \quad n = -\infty, \ldots, -1, 0, 1, \ldots \infty ,
\]  
(3.676)
and the width of a tone becomes
\[
d\omega = \lim_{N \to \infty} \frac{2\pi}{NT} .
\]  
(3.677)
If $1/T'$ is sufficiently large to be at least twice the highest frequency that could be conceived of for use on any given band-limited channel, then $\Omega^{opt}$ becomes the true optimum band for use on the continuous channel. The two-sided power spectral density at frequency $\omega$ corresponding to $n/NT$ then is
\[
S_x(\omega) = \lim_{T \to \infty} \tilde{E}_n .
\]  
(3.678)

When using infinite positive and negative time and frequency as here, there is no need for complex baseband equivalents, and thus all dimensions are considered real (and QAM is just then one real dimension at positive frequency and one real dimension at that the negative image of that frequency). The data rate then becomes
\[
R = \frac{1}{2\pi} \int_{\Omega^{opt}} \frac{1}{2} \log_2 \left( 1 + \frac{S_x(\omega) \cdot g(\omega)}{\Gamma} \right) d\omega .
\]  
(3.679)
The input power constraint is
\[
P_x = \frac{1}{2\pi} \cdot \int_{\Omega^{opt}} S_x(\omega) d\omega .
\]  
(3.680)
Calculus of variations (see Chapter 4) produces Shannon’s famous water-filling equation for continuous frequency with discrete time as:
\[
\tilde{E}_n + \frac{\Gamma}{g_n} \to S_x(\omega) + \frac{\Gamma}{g(\omega)} = \lambda \text{ (a constant)}
\]  
(3.681)
where the the value $\lambda$ is chosen to meet the total power constraint in (3.680), recalling that $S_x(\omega) > 0$ for all $\omega \in \Omega^{opt}$ and $S_x(\omega) = 0$ at all other frequencies. The data rate then can be also written
\[
R = \frac{1}{2\pi} \int_{\Omega^{opt}} \frac{1}{2} \log_2 \left( \frac{\lambda \cdot g(\omega)}{\Gamma} \right) d\omega ,
\]  
(3.682)
These results do extend to MIMO, but require some notational effort that will be facilitated by the developments of Chapter 4.

**EXAMPLE 3.12.4 (1 + .9D channel)** The channel with impulse response $h(t) = \text{sinc}(t) + .9 \cdot \text{sinc}(t - 1)$ has the same performance as the $1 + .9D^{-1}$ channel studied throughout this book, if the analog transmit filter (without spectrum shaping included as it will be optimized) is $\frac{1}{\sqrt{T}} \text{sinc}(t/T)$. A system with optimized MT basis functions of infinite length (as $T \to \infty$) would have an optimum bandwidth $\Omega^{opt} = [-W, W]$ as in Figure 3.12.2 Then, continuous water-filling with $\Gamma = 1$ produces
\[
P_x = \int_{-W}^{W} \left( \lambda' - \frac{.181}{1.81 + 1.8 \cos(\omega)} \right) \frac{d\omega}{2\pi}
\]  
(3.683)
where $W$ is implicitly in radians/second for this example. If $P_x = 1$ with $\frac{\lambda_0}{\pi} = .181$, the integral in (3.683) simplifies to
\[
\pi = \int_{0}^{W} \left( \lambda' - \frac{.181}{1.81 + 1.8 \cos(\omega)} \right) d\omega
\]  
(3.684)
\[
= \chi W - .181 \left\{ \frac{2}{\sqrt{1.81^2 - 1.8^2}} \arctan \left[ \frac{\sqrt{1.81^2 - 1.8^2}}{1.81 + 1.8 \tan \left( \frac{W}{2} \right)} \right] \right\}
\]  
(3.685)
At the bandedge $W$, 

$$\lambda' = \frac{.181}{1.81 + 1.8 \cos(W)}.$$  \hfill (3.686)

leaving the following transcendental equation to solve by trial and error:

$$\pi = \frac{.181W}{1.81 + 1.8 \cos(W)} - 1.9053 \arctan (.0526 \tan(W/2))$$  \hfill (3.687)

$W = .88\pi$ approximately solves (3.687), and the corresponding value of $\lambda$ is $\lambda = 1.33$.

The highest data rate with $1/T' = 1$ is then

$$C = \frac{2}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{1}{2} \log_2 \left( \frac{1.33}{.181} (1.81 + 1.8 \cos \omega) \right) d\omega$$  \hfill (3.688)

$$= \frac{1}{2\pi} \int_{0}^{\pi} \log_2 7.35 d\omega + \frac{1}{2\pi} \int_{0}^{\pi} \log_2 (1.81 + 1.8 \cos \omega) d\omega$$  \hfill (3.689)

$$= 1.266 + .284$$  \hfill (3.690)

$$\approx 1.55 \text{bits/second}.$$  \hfill (3.691)

This exceeds the 1 bit/second transmitted on this channel in this chapter’s earlier examples where $T = T' = 1$. The MT system has no error propagation, and is also an ML detector.

### 3.12.3 Discrete-Time Water-filling and the Paley-Wiener Criterion

As in Subsection 3.12.1, transmission is often implemented in discrete time. The mutual information $\bar{I}(X(D);Y(D))$ is a function only of the power spectral densities of the discrete time processes $X(D)$ and $Y(D)$ if they are Gaussian. Thus maximization of $\bar{I}(X(D);Y(D))$ over the power spectral density of $X(D)$ will produce a symbol-rate-dependent water-fill spectrum for the transmit filter, which in turn has as in put a white (and “near-Gaussian”) message sequence.

Figure 3.12.3 shows the implementation of a transmitter that includes a digital filter $\phi_k$ (with transform $\Phi(D)$). This filter precedes the modulator that converts the symbols at its output $x_k$ into the modulated waveform $x(t)$, typically QAM or PAM. The discrete-time filter input is $v_k$, which now becomes the message sequence. Recalling that $q(t) \triangleq p(t) * p^*(-t)/\|p\|^2$, the mutual information or maximum data rate from Section 4.4.3 is

$$\bar{I}(X(D);Y(D)) = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \log_2 \left( \frac{\|p\|^2 \cdot \bar{E}_m}{\frac{\Delta_0}{T} \cdot |\Phi(e^{-j\omega T})|^2 \cdot Q(e^{-j\omega T}) + 1} \right) d\omega.$$  \hfill (3.692)
The transmit energy constraint is the (Γ = 0 dB)-sum of data rates on an infinite number of infinitessi-
mally small tones:

\[ \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \bar{E}_v \cdot |\Phi(e^{-j\omega T})|^2 d\omega = \bar{E}_x . \]  (3.693)

The maximization of \( \bar{I}(X(D); Y(D)) \) is then achieved by classic water fill solution in

\[ |\Phi(e^{-j\omega T})|^2 + \frac{N_0}{\bar{E}_v \cdot \|p\|^2 \cdot Q(e^{-j\omega T})} = K . \]  (3.694)

As always, energy at any frequency \( \omega \) must be nonnegative, so

\[ |\Phi(e^{-j\omega T})|^2 \geq 0 . \]  (3.695)

There is a set of frequencies \( \Omega^{opt} \) for which transmit energy is nonzero and for which the discrete-time
transmit filter output energy satisfies water-filling. The corresponding capacity is then

\[ \bar{C} = \max_{\|\Phi(e^{-j\omega T})\|^2 = 1} \bar{I}(X(D); Y(D)) . \]  (3.696)

The measure of \( \Omega^{opt} \) for the discrete-time case is similarly

\[ |\Omega^{opt}| = \frac{T}{2\pi} \int_{\Omega^{opt}} d\omega . \]  (3.697)

If |\( \Omega^{opt} \) = 1, then a realizable transmit filter exists by the PW criterion. In this case, SNR_{MMSE-DFE,U} = 2^{2C} - 1 and the MMSE-DFE transmission system with water-fill transmit power spectral density can achieve, with powerful known AWGN-channel codes, the highest possible rates, so the MMSE-DFE in this case only is effectively optimal or “canonical.” If the water-fill band does not have unit measure, the transmitter is not realizable, and the MMSE-DFE is not optimal. The non-
unit-measure case is very unlikely to occur unless the sampling rate has been judiciously chosen. This
is the single most commonly encountered mistake of QAM/PAM designers who attempt to match the
performance of MT systems - only in the rare case of |\( \Omega^{opt} \) = 1 can this design work.

3.12.4 The baseband lowpass channel with a single contiguous water-fill band

The baseband lowpass channel has the transfer characteristic of Figure 3.72. The continuous-time
waterfill band \( \Omega^{opt} \) is also shown. Clearly sampling of this channel at any rate exceeding 1/\( T^{opt} = |\Omega^{opt}| \)
should produce a discrete-time channel with capacity in bits/dimension

\[ \bar{C}(T < T^{opt}) = C \cdot T , \]  (3.698)
where the explicit dependency of $\bar{C}$ on choice of symbol period $T$ is shown in the argument $\bar{C}(T)$. The capacity in bits/second remains constant while the capacity in bits per symbol decreases with increasing symbol rate to maintain the constant value $C = \bar{C}(T)/T$. At symbol rates below $1/T < 1/T_{\text{opt}} = \Omega_{\text{opt}}$, capacity $C$ may not be (and usually is not) achieved so

$$\bar{C}(T > T_{\text{opt}}) \leq C \cdot T \quad .$$

(3.699)

To achieve the highest data rates on this lowpass channel with good codes, the designer would like to choose $1/T \geq 1/T_{\text{opt}}$. However, the transmit filter is not realizable unless $1/T = 1/T_{\text{opt}}$, so there is an optimum symbol rate $1/T_{\text{opt}}$ for which

$$\text{SNR}_{\text{MMSE-DFE,U}}(T_{\text{opt}}) = 2^{2C_{\text{opt}}} - 1 \quad .$$

(3.700)

At this symbol rate only can the MMSE-DFE achieve the best possible performance for the lowpass channel and indeed achieve the MT performance levels, see also Chapter 4. This example correctly suggests that an ideal symbol rate exists for each channel with “PAM/QAM-like” transmission and a MMSE-DFE receiver.

For rates below capacity where a code with gap $\Gamma > 1$ is used, the continuous-time water-filling can be solved with the transmit power $P_x$ reduced by $\Gamma$. This would produce a slightly lower optimum symbol rate $1/T_{\Gamma_{\text{opt}}} \leq 1/T_{\text{opt}}$. This slightly lower symbol rate should then be used for transmission with gap $\Gamma$, and the bit rate of the MMSE-DFE system would be

$$R = \frac{\bar{b}}{T_{\Gamma_{\text{opt}}}} = \frac{1}{2T_{\Gamma_{\text{opt}}}} \log_2 \left( 1 + \frac{\text{SNR}_{\text{MMSE-DFE,U}}(T_{\Gamma_{\text{opt}}})}{\Gamma} \right) \quad .$$

(3.701)

Two examples are provided to illustrate the effects and terminology of symbol-rate optimization:

**EXAMPLE 3.12.5 (1 + .9D−1 channel)** The 1 + .9D−1 ISI-channel has been revisited often, and is again revisited here with $\xi_x \|p\|^2/N_0 = 10$ dB. This channel corresponds to the real baseband lowpass channel discussed above. The capacity in bits per real sample/dimension has been previously found in Example 4.4.1 to be $C(T = 1) = 1.55$ bits/dimension, so that $C = 1.55$ bits/second with $T = 1$. This capacity corresponds to an overall MM SNR

---

39Less water to pour means more, or same, narrow water-fill band always.
Figure 3.73: Block diagram of optimized DFE system for the $1 + .9D^{-1}$ channel.

of 8.8 dB, but could not be directly realized with a MMSE-DFE, which instead had an SNR\textsubscript{MMSE-DFE,U} = 8.4 dB at the symbol rate of $1/T = 1$. The optimum water-fill band has width $|\Omega_{opt}| = .88\pi$ (positive frequencies). Then, the optimum symbol rate is

$$1/T_{opt} = .88 .$$

Thus, the capacity in bits/dimension for the system with optimum symbol rate is

$$C(T_{opt}) = 1.55 \frac{.88}{.88} = 1.76 \text{ bits/dimension} .$$

The resultant SNR is then

$$\text{SNR}_{\text{MMSE-DFE,U}}(T_{opt}) = \text{SNR}_{\text{MMSE-DFE,U}}^{opt} = 10.2 \text{ dB} .$$

This optimum MMSE-DFE at this new symbol rate is the only MMSE-DFE that can match the MT system in performance (if there is no loss from error propagation or precoding). The SNR\textsubscript{MFB} of 10 dB in previous invocations of this example is no longer a bound unless the symbol rate is $1/T = 1$ since that bound was derived for $T = 1$ and needs rederivation if another symbol rate is used. For verification, the capacity in bits/second is consistently the same at both sampling rates, but only $1/T_{opt} = .88$ can be used with a MMSE-DFE\textsuperscript{40}. The capacity in bits/second remains

$$C = \frac{.88}{2} \log_2(1 + 10^{1.02}) = \frac{1}{2} \log_2(1 + 10^{-88}) = 1.55 \text{ bits/second} .$$

\textbf{EXAMPLE 3.12.6 (The “half-band” ideal lowpass channel)} A brickwall lowpass channel has cut-off frequency .25 Hz as shown in Figure 3.74. Clearly the water-filling bandwidth is $\Omega_{opt} = 2\pi(-.25, .25)$. The capacity is given as $C = 2$ bits per second for this channel. Thus, at $T_{opt} = 2$, then $C(T_{opt}) = 4$ bits/dimension. By the CDEF result, a MMSE-DFE with this symbol rate will have performance given by $\text{SNR}_{\text{MMSE-DFE,U}}(T_{opt} = 2) = 2^{34} - 1 \approx 24$ dB. The MMSE-DFE in this case is clearly trivial and the equalized channel is equivalent to the original ISI-free channel at this symbol rate.

Suppose instead, a designer not knowing (or not being able to know for instance in broadcast or one-directional transmission) the channel’s Fourier transform or shape instead used $T = 1$ for a DFE with flat transmit energy. While clearly this is a poor choice, one could easily

\textsuperscript{40}At the optimum symbol rate, $E_x = 1/.88$, $\|p\|^2 = 1.81/.88$, and the noise per sample increases by $1/.88$. Thus, $\text{SNR}_{\text{MFB}}(.88) = 10.6 \text{ dB}$.
Figure 3.74: Example channel.

Figure 3.75: Margin difference (improvement) for optimized transmission over unoptimized MMSE-DFE on a “half-band” channel.
envison situations that approximate this situation in applications with variable channels. An immediate loss that is obvious is the 3 dB loss in received energy that is outside the passband of the channel, which in PAM/QAM is equivalent to roughly to .5 bit/dimension loss. This represents an upper bound SNR for the MMSE-DFE performance. The mutual information (which is now less than capacity) is approximately then
\[
\bar{I}(T = 1) \approx \frac{\bar{I}(T^{\text{opt}} = 2) - .5}{2} = 3.5 = 1.75 \text{bits/dimension .} \tag{3.706}
\]
The .5 bit/dimension loss in the numerator \(\bar{I}(T^{\text{opt}} = 2) - .5 = \frac{1}{2} \log_2(1 + \frac{\text{SNR}}{2})\) approximates the 3 dB loss of channel-output energy caused by the channel response. Clearly, the channel has severe ISI when \(T = 1\), but the CDEF result easily allows the computation of the MMSE-DFE SNR according to
\[
\text{SNR}_{MMSE-DFE,U}(T = 1) = 2^{2.175} - 1 = 10.1 \text{ dB .} \tag{3.707}
\]
The equalized system with \(T = 1\) is equivalent to an AWGN with SNR=10.1 dB, although not so trivially and with long equalization filters in this \(T = 1\) case, because ISI will be severe.

The system with \(T = 1\) then attempts to transmit at \(\bar{b}(T = 1)\) with an SNR of 10.1 dB, while the optimized system transmits \(\bar{b}(T^{\text{opt}} = 2) = 2\bar{b}(T = 1)\) with SNR of 24 dB. The margin as a function of \(\bar{b}(T = 1)\) is then
\[
\gamma_m(T = 1) = 10 \cdot \log_{10} \left( \frac{2^{2.175} - 1}{2^{2\bar{b}(T=1)} - 1} \right) = 10.1 \text{ dB} - 10 \cdot \log_{10} \left( 2^{2\bar{b}(T=1)} - 1 \right) . \tag{3.708}
\]
For \(T^{\text{opt}} = 2\), the corresponding margin (basically against an increase in the flat noise floor) is
\[
\gamma_m(T^{\text{opt}} = 2) = 10 \cdot \log_{10} \left( \frac{2^{2.4} - 1}{2^{2\bar{b}(T^{\text{opt}}=2)} - 1} \right) = 24 \text{ dB} - 10 \cdot \log_{10} \left( 2^{4\bar{b}(T=1)} - 1 \right) . \tag{3.709}
\]
The ratio of the margins for \(T = 1\) and \(T^{\text{opt}} = 2\) is then plotted in Figure 3.75 as a function of the number of bits per dimension. The difference decreases as the systems approach capacity, meaning codes with smaller gap are used, but is always nonzero in this case because the system with \(T = 1\) essentially “wastes 3 dB (a factor of 2) of bandwidth.” Indeed, it is not possible for the \(T = 1\) system to transmit beyond 1.75 bits/second and is thus at best 3 dB worse than the \(T^{\text{opt}} = 2\) system because half the energy is wasted on a channel band that cannot pass signal energy. One could infer that at fixed \(P_e\), the gap decreases with increasing \(\bar{b}(T = 1)\) on the horizontal access until the \(\Gamma = 0 \text{ dB limit is reached at 1.75 bps.}\)

Some designers might attempt transmission with a low-pass filter (with gain 3 dB) to correct the situation, thus regaining some of the minimum 3 dB energy loss when \(\Gamma = 0 \text{ dB}\). As the transmit filter becomes more tight, it becomes difficult to implement - in the limit, such a filter is not realizable because it does not meet the PW criterion. However, leaving \(T = 1\), no matter how complex the transmit filter, forces ISI and makes both transmitter and receiver very complex with respect to using the correct \(T^{\text{opt}} = 2\). As the gap increases, the performance difference is magnified between \(T = 1\) and \(T^{\text{opt}} = 2\), so even a very good brickwall transmit filter at \(T = 1\) loses performance when \(\Gamma > 0\). An implementable system must have a non-zero gap. By contrast, the optimum transmit filter is simple flat passband at \(T^{\text{opt}} = 2\), and the DFE receiver filters are trivial. Also as the capacity/SNR increases, the curve in Figure 3.75 will show larger difference at reasonable data rates, but more rapidly fall to again 3dB at the point where \(T = 1\) system achieves capacity with \(\Gamma = 0\) dB. It is thus very inadvisable design to use the incorrect symbol rate from the perspectives of performance or complexity.

\[^{41}\text{In fact, this upper bound can be very closely achieved for very long filters in MMSE-DFE design, which could be verified by DFE design for instance according to the DFE program in Section 3.7.6.}\]
3.12.5 The single-band bandpass case

The channels of the previous subsection (3.12.3) were baseband (i.e., real one-dimensional) lowpass channels. A complex lowpass channel tacitly includes the effect of a carrier frequency. Figure 3.76 illustrates the passband channel and the choice of best carrier (or really “center”) frequency $f_c$ and then consequently the best symbol rate $1/T_{opt}$. Clearly, to select the optimum symbol rate, the QAM carrier frequency must be in the center of the water-filling band. Any other choice of carrier frequency results in an asymmetric use of water-filling band frequencies around DC, which means that the PW criterion will not be satisfied.\footnote{Note Carrierless AMPM (CAP) systems of Chapter 1 do not use a carrier, but have a “center” frequency. The optimum center frequency is computed exactly the same way as the optimum carrier frequency in this subsection.}

Thus, for the MMSE-DFE on a passband channel, the carrier frequency must be centered in the water-filling band and the symbol rate is chosen to be the measure of the resulting (positive-frequency for passband case) continuous-time water-filling band. Again, a gap $\Gamma > 0$ dB causes a slight decrease in the water-fill band’s measure and will also alter the carrier/center frequency slightly (unless the band is conjugate symmetric with respect to the choice of carrier/center frequency).

**EXAMPLE 3.12.7 (V.34 Voiceband Modems)** The International Telecommunications Standardized v.34 modem, better known as the “28.8 kbps” voiceband modem used optimized...
Figure 3.77: Illustration of water-filling and transmit optimization for a general channel with more than one passband.

decision feedback equalization. These older modems initially trained by using multitone transmission technology, specifically 25 equally spaced tones are sent with fixed amplitude and phase at the frequencies $n(150\text{Hz})$ where $n = 1...25$. The frequencies 900, 1200, 1800, and 2400 are silenced, leaving 21 active tones for which the SNR’s are measured (and interpolated for 900, 1200, 1800, and 2400) Water-filling or other spectrum-setting procedures determine an optimum bandwidth, which is reduced to a choice of symbol rate and carrier frequency for QAM. The v.34 choices for symbol rate and carrier must be from the attached table:

<table>
<thead>
<tr>
<th>$1/T^{opt}$</th>
<th>$f_{c1}$</th>
<th>$f_{c2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2400</td>
<td>1600</td>
<td>1800</td>
</tr>
<tr>
<td>2743</td>
<td>1646</td>
<td>1829</td>
</tr>
<tr>
<td>2800</td>
<td>1680</td>
<td>1867</td>
</tr>
<tr>
<td>3000</td>
<td>1800</td>
<td>2000</td>
</tr>
<tr>
<td>3200</td>
<td>1829</td>
<td>1920</td>
</tr>
<tr>
<td>3429</td>
<td>1959</td>
<td>1959</td>
</tr>
</tbody>
</table>

Only the symbol rates in the table may be selected and after one is selected, only two carrier frequencies are then allowed as denoted in the same row. There is never more than one disjoint WF band, and examination of the choices with respect to optimum has revealed a loss of less than 1 dB. (Voiceband modems rarely have $M > 1$, so the issue of multiple band MMSE-DFE is not addressed in v.34, see Section 3.12.5). This special case of a single band being sufficient does not extend to DSL for instance, see Chapter 4. Designers became more aware after voiceband modems became obsolete, so this approach has been abandoned, but remains nevertheless a good example of MMSE-DFE transmit optimization.

3.12.6 The general multiple band case and the Optimum MMSE-DFE

Examples 3.72 and 3.76 were overly simple in that a single continuous transmission band was presumed. On ISI channels that exhibit notches (i.e., multipath fading or bridged-taps, multiple antennas, etc.) and/or band-selective noise (crosstalk, RF noise, etc.), $\Omega^{opt}$ may consist of many bands. This subsection addresses these more practical cases to find a minimal set of MMSE-DFEs that can be analyzed as a single system.
Practical ISI-channels often have water-filling solutions that consist of a countable number of bandpass (and perhaps one lowpass) water-fill frequency bands, which have nonzero transmit energy over a continuous set of frequencies within each band. To satisfy the PW criterion, the MMSE-DFE then must become multiple MMSE-DFE’s, each with its own independent water-filling solution. This situation appears in Figure 3.77 for 3 disjoint water-fill regions. Correspondingly, there would be 3 MMSE-DFE’s each operating at an SNR

\[ \text{SNR}_{\text{MMSE-DFE}} \]

for its band and at a data rate \( R_i \) for each band.

An overall SNR can be computed in this case of multiple bands. An overall symbol rate, presuming carriers centered in each band, is defined

\[ \frac{1}{T_{\text{opt}}} = \sum_{i=1}^{M} \frac{1}{T_{i}^{\text{opt}}} \]

where \( T_{i}^{\text{opt}} = (1 \text{ or } 2) \cdot T_{i}^{\text{opt}} \) for complex passband or real baseband respectively. The SNR in each band is \( \text{SNR}_i(T_{i}^{\text{opt}}) \). Each band has data rate

\[ R_i = \frac{1}{T_{i}^{\text{opt}}} \cdot \log_2 \left( 1 + \frac{\text{SNR}_i(T_{i}^{\text{opt}})}{\Gamma} \right) \]

Then, the overall data rate is in bits per second is

\[ R = \sum_{i=1}^{M} R_i \]

\[ = \sum_{i=1}^{M} R_i = \sum_{i=1}^{M} \frac{1}{T_{i}^{\text{opt}}} \cdot \log_2 \left( 1 + \frac{\text{SNR}_i(T_{i}^{\text{opt}})}{\Gamma} \right) \]

\[ = \frac{T_{\text{opt}}}{T_{\text{opt}}} \log_2 \prod_{i=1}^{M} \left( 1 + \frac{\text{SNR}_i(T_{i}^{\text{opt}})}{\Gamma} \right)^{1/T_{i}^{\text{opt}}} \]

\[ = \frac{1}{T_{\text{opt}}} \log_2 \left( 1 + \frac{\text{SNR}_{\text{MMSE-DFE}}(T_{\text{opt}})}{\Gamma} \right) \]

\[ = \frac{\bar{T}_{\text{opt}}}{T_{\text{opt}}} \]

where

\[ \text{SNR}_{\text{MMSE-DFE}} \equiv \Gamma \cdot \left\{ \prod_{i=1}^{M} \left( 1 + \frac{\text{SNR}_i(T_{i}^{\text{opt}})}{\Gamma} \right)^{T_{i}^{\text{opt}}/T_{i}^{\text{opt}}} \right\} - 1 \]

and

\[ \bar{T}_{\text{opt}} \equiv \frac{1}{2} \cdot \log_2 \left( 1 + \frac{\text{SNR}_{\text{MMSE-DFE}}}{\Gamma} \right) \] bits/dimension .

When all symbol rates and carrier frequencies are optimized, clearly this SNR is equivalent to the MT system of Subsection 3.12.2 because the rates/capacities for each water-filling band are the same. However, the MMSE-DFE now is often many MMSE-DFE’s, and each has a distinct variable carrier frequency and symbol rate. The reader is cautioned against the statements often encountered by as-yet-uninformed engineers who misinterpret the CDEF result to mean that a “single-carrier system performs the same as a multicarrier system” – misquoting the CDEF result. Such a statement is only true in the simplest cases where a single water-fill band is optimum, and the symbol rate and carrier frequencies have been precisely chosen as a function of the specific channel’s water-filling solution and coding gap. Otherwise, multicarrier will outperform single carrier – and further, single carrier can never outperform
multicarrier (with codes of the same gap $\Gamma$). More sophisticated DMT systems are developed in Chapter 4 for realistic implementation of MT.

**Lemma 3.12.1 (The Optimum MMSE-DFE)** The optimum MMSE-DFE is a set of $M$ independent DFE’s with $M$ equal to the number of disjoint water-filling bands. Each of the MMSE-DFE’s must have a symbol rate equal to the measure of a continuous water-fill band $1/T_{\text{opt}}^m = |\Omega_{\text{opt}}^m|$ and a carrier frequency where appropriate set exactly in the middle of this band ($f_{c,m} = (f_{\text{max},m} + f_{\text{min},m})/2 \forall m = 1, ..., M$). The number of bits per dimension for each such MMSE-DFE is

$$b_m = \frac{1}{2} \cdot \log_2 \left( 1 + \frac{\text{SNR}_m(T_{\text{opt}}^m)}{\Gamma} \right).$$

(3.720)

The overall SNR is provided in (3.718) and the (3.719)

A final simplified VDSL example provided to illustrate both proper design and a performance loss that might be incurred when designers try “to simplify” by only using one fixed band.

**EXAMPLE 3.12.8 (Simple Lost “Dead” Bands)** Figure 3.78 illustrates equivalent transmitters based on a multitone design of 8 equal bandwidth modulators. All tones are presumed passband uncoded QAM with gap $\Gamma = 8.8$ dB. This system chooses symbol rates to approximate what might happen certain low pass channel with notched frequency range. The transmit filter for each band is denoted by $H_n(f)$ in set A and $G_n(f)$ in set B. The $g_n$’s for these example subchannels are given in the table below.

Application of water-filling with 11 units of energy produces channel SNR’s that appear in the following table for a Gap of $\Gamma = 8.8$ dB and a water-filling constant of $\lambda' = 2$:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$g_n$</th>
<th>$f_n$</th>
<th>$b_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.2</td>
<td>1.50</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>30.3</td>
<td>1.75</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>121.4</td>
<td>1.9375</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>242.7</td>
<td>1.97</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>60.7</td>
<td>1.875</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>242.7</td>
<td>1.97</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Channel characteristics are such in this system that the optimum bandwidth use is only 6 of the 8 subchannels. 4 of the used bands (set A) are adjacent and another 2 of the bands (set B) are also adjacent. Sets A and B, however, are separated by an unused band. That unused band might be for example caused by a radio interference or perhaps notching from a reflected and delayed signals. The corresponding 2-DFE receiver is shown in Figure 3.79.

**First Multitone Design with 8 independent tones:**

The tones in Set A can carry $(2b_n =) 2, 3, 5, \text{ and } 6$ bits respectively corresponding to increasing signal-to-noise ratios in the corresponding channel, while the two bands in Set B can carry 4 and 6 bits respectively. The unused bands carry 0 bits. The average number of bits per tone in Set A is 4 while the average number in Set B is 5. If the symbol rate for each subchannel is 1 MHz, then the data rate is $(2+3+5+6+4+6)$ bits transmitted one million times per second for a total data rate of 26 Mbps. Equivalently, from Equation (3.710), the overall symbol rate is $1/T_{\text{opt}}^\prime = 6$ MHz, and there are thus $\tilde{b}(1/T_{\text{opt}}^\prime = 6$ MHz$)$ = $26 \text{ Mbps} / (1/T_{\text{opt}}^\prime) = 26/[2(6)] = 2.16 \text{ bits/dim or } 4.33 \text{ bits per equivalent tone.}$ For an uncoded-$\Gamma = 8.8$ dB multitone system $\tilde{b}(1/T^\prime = 8$ MHz$)$ = $26/16 = 1.625$. This rate is equivalent to
Figure 3.78: Simplified VDSL example multitone and equivalent 2-channel transmitters.
\( \bar{b}(1/T_{\text{opt}} = 6\text{MHz}) = 2.16 \) since both correspond to the same overall data rate of 26 Mbps. The SNR at a symbol rate of \( 1/T = 8 \text{ MHz} \) is

\[
\text{SNR}_{\text{mt}}(1/T' = 8\text{MHz}) = 10 \cdot \log_{10} \left\{ \Gamma \cdot \left[ 2^{2(1.625)} - 1 \right] \right\} = 18.1\text{dB} \, .
\] (3.721)

The SNR at a symbol rate of \( 1/T_{\text{opt}} = 6 \text{ MHz} \) is

\[
\text{SNR}_{\text{opt}}^{\text{MMSE-DFE,U}}(1/T^* = 6\text{MHz}) = 10 \cdot \log_{10} \Gamma \cdot (2^{4.31} - 1) = 21.6\text{dB} \,
\] (3.722)

The two MT SNRs are different because the symbol rates are different, but both correspond to a \( P_e = 10^{-6} \) at the data rate of 26 Mbps with gap 8.8 dB (same code) and \( P_e = 10^{-6} \).

2nd Design: Equivalent Two-tone QAM Design:

The corresponding best QAM receiver uses the same bands and is actually 2 QAM systems in this example. The first system has a symbol rate of \( 1/T_{\text{opt}} = 4 \text{ MHz} \) and carries 4 bits per QAM symbol (16 QAM) for a data rate of 16 Mbps. By the CDEF result, the MMSE-DFE receiver has an SNR of \( \text{SNR}_{\text{dfe,1}}(1/T_{\text{opt}} = 4\text{MHz}) = \Gamma \cdot (2^4 - 1) = 20.6 \text{ dB} \). The second system has symbol rate \( 1/T_{\text{opt}}^* = 2 \text{ MHz} \), and 5 bits per QAM symbol, which corresponds to 10 Mbps and \( \text{SNR}_{\text{dfe,2}}(1/T_{\text{opt}} = 2\text{MHz}) = \Gamma \cdot (2^5 - 1) = 23.7 \text{ dB} \). The total data rate of 26 Mbps is the same as the DMT system in the first design. The two systems (DMT or two-tone QAM) carry the same data rate of 26 Mbps with a probability of symbol error of \( 10^{-6} \).

Correct design thus finds a 2-band equivalent of the original 8-tone MT system, but absolutely depends on the use of the 2 separate QAM signals in the lower portions of Figures 3.78 and 3.79.

3rd Design - Single-tone QAM Designer:

Some designers might instead desire to see the loss of the single-band DFE system. The single-band (with \( 1/T = 8 \text{ MHz} \)) QAM system with MMSE-DFE then uses \( 2E_x = \bar{b}/8 = 1.375 \) at all frequencies and thus has

\[
\text{SNR}_{\text{dfe,flat}} = \Gamma \cdot \left\{ \prod_{n=1}^{8} \left[ 1 + \frac{1.375 \cdot g_n}{\Gamma} \right]^{1/8} - 1 \right\} = 17\text{dB} \,
\] (3.723)

which is about 1.1 dB below the multitone result. If Tomlinson Precoding was used at the \( \bar{b} = 1.625 \), the loss is another .4 dB, so this is about 1.5 dB below the multitone result in practice. This \( 1/T = 8 \text{ MHz} \) single-tone QAM does then perform below MT because two separate QAM/DFE systems are required to match MT performance.

As a continued example, this same transmission system is now applied to a more severely attenuated channel for which the upper frequencies are more greatly attenuated and \( g_6 = g_7 \leq 2 \), essentially zeroing the water-fill energy in band B. The first multitone system might attempt 16 Mbps data rate. Such a system could reload the 1.875+1.97 units of energy on tones 6 and 7 to the lower-frequency tones, and the new SNR_{mt} will be 14.7 dB, which is a margin of 1.2 dB with respect to 16 Mbps (2\bar{b} = 2 requires 13.5 dB). A new computation of the single-tone (\( 1/T = 8 \text{ MHz} \)) SNR according to (3.723), which now has only 4 terms in the product provides \( \text{SNR}_{\text{DFE}}(1/T = 8\text{MHz}) = 12.8 \text{ dB} \) that corresponds for the same data rate of 16 Mbps to a margin of -.7 dB (\( \gamma = \text{SNR}_{\text{DFE}}/\Gamma \cdot 2^{2(-1)} \)). The MT system is 1.9 dB better. Furthermore, precoder loss would be 1.3 dB for a Tomlinson or Laroia precoder, leading to a 3.2 dB difference. Thus there is always a loss when the single-band system is used, and the amount of loss increases with the inaccuracy of the symbol-rate with respect to best symbol rate, and with the gap as in Figure 3.75.

Since the margin of the QAM system is negative on the more severely attenuated channel, the practice in industry is to “design for the worst case” channel, which means a better symbol rate choice would be 4 MHz. In this case, on the channel, both MT and QAM at
Figure 3.79: Simplified VDSL example multi-tone and equivalent 2-channel receivers.
16 Mbps would have the SNR of the A band, 20.6 dB, increased by \((11/(11-1.97-1.875) = 1.86 \text{ dB})\) to 22.5 dB and a margin of 1.86 dB at \(10^{-6}\) probability of error with \(\Gamma = 8.8\) dB. However, if the short line with bands A AND ALSO B is now again encountered, the 4 MHz-worst-case-designed QAM system will remain at 1.86 dB margin at 16 Mbps. The MT system margin at 16 Mbps now improves to (Equation 4.7 in Chapter 4)

\[
\gamma_{mt} = \frac{2^{2b_{max}} - 1}{2^{2b} - 1} = \frac{2^{4.25} - 1}{2^{2} - 1} = 4.5 \text{ dB} \quad (3.725)
\]

or a 2.6 dB improvement in margin over the 1.9 dB margin of the QAM system.

As the used-bandwidth error grows, the loss of a single-band DFE increases, but is hard to illustrate with small numbers of tones. Used bandwidth ratios can easily vary by a factor of 10 on different channels, increasing the deviation (but not easily depicted with a few tones) between MT systems and DFE systems that use a fixed bandwidth (or single tone). For instance, if only tone 4 were able to carry data that the best rate-adaptive solution is about 8.5 Mbps, while the full-band single-QAM DFE would attain only 420 kbps, and performs approximately 7 dB worse.

For real channels, independent test laboratories hosted a “VDSL” Olympics in 2003 for comparing DMT systems with variable-symbol-rate QAM systems. The test laboratories were neutral and wanted to ascertain whether a single-carrier system (so only one contiguous band that could be optimized and placed anywhere) really could “get the same performance as DMT” – an abuse of the CDEF result then misunderstood and promulgated by single-carrier proponents). The results and channels appear in Figures 3.80 and 3.81 respectively. Clearly these fixed-margin (6 dB) tests for length of copper twisted pair at the data rate shown indicate that over a wider range of difficult channels as in DSL, the differences between DMT and single-carrier can be quite large. Generally speaking, the more difficult the channel, the greater the difference. These MT systems used a discrete multitone or DMT as in Chapter 4.

This section has shown that proper optimization of the MMSE-DFE may lead to several MMSE-DFEs on channels with severe ISI, but that with such a minimum-size set, and properly optimized symbol rate and carrier frequency for the corresponding transmitter of each such MMSE-DFE, determines a canonical transmission system, then matching the MT system.

### 3.12.7 Relations between Zero-Forcing and MMSE DFE receivers

In some special situations, the ZF-DFE and the MMSE-DFE can induce the same equivalent channel, although the performance will not be the same. Some caution needs to be exercised for equivalent channels. Some authors \(^{43}\) have made some serious errors in interpreting results like CDEF for the ZF-DFE.

---

**Theorem 3.12.2 (CDEF Derivative)** Over any of the continuous bands of water-filling, if the water-fill spectra are used, then the ZF-DFE and MMSE-DFE result in the same channel shaping before bias removal.

**proof:** This proof uses continuous bands, presuming the condition for choice of optimum sampling rate(s) and carrier frequencies have already been made so that all frequencies are used with non-zero energy (except possibly for countable set of infinitessimally narrow

---

\(^{43}\)One somewhat incorrectly derived result is due to MIT’s Professor Robert Price in early 70’s where he tacitly assumed zero noise, and then water-filling uses the entire band of transmission. Thus Price accidentally considered only one transmission band that was the entire band. Of course, trivially, ZF and MMSE are the same when there is no noise – this is hardly surprising, and Price erroneously arrived at a result resembling CDFE that is correct only when there is zero noise. This is sometimes referred to as “Price’s result,” but it is clearly Price at the time did not understand his small-equal-zero noise assumption would completely miss the important need for realizable filters.
Variable $f_c$ and $1/T$ single-carrier QAM results

DMT* results – exact same channels as QAM

*DMT is described in Chapter 4

Figure 3.80: Illustration of 2003 VDSL Olympics results.
<table>
<thead>
<tr>
<th>Config</th>
<th>Test</th>
<th>Service (Mbit/s) (Down/Up)</th>
<th>Reach (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>10/10</td>
<td>1800</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>10/10</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10/10</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10/10</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10/10</td>
<td>1500</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>10/10</td>
<td>1950</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>13/13</td>
<td>1350</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>13/13</td>
<td>1500</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>13/13</td>
<td>1500</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>13/13</td>
<td>1500</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>13/13</td>
<td>1350</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>13/13</td>
<td>1500</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>16/1</td>
<td>3300</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>16/1</td>
<td>3300</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>16/1</td>
<td>3300</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>16/1</td>
<td>2550</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>22/3</td>
<td>2700</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>22/3</td>
<td>2550</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>22/3</td>
<td>2400</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22/3</td>
<td>2400</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>22/3</td>
<td>2400</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>22/3</td>
<td>2550</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>22/3</td>
<td>2250</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>22/3</td>
<td>3150</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>22/3</td>
<td>2250</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>22/3</td>
<td>2700</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
<td>22/3</td>
<td>2700</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>22/3</td>
<td>2400</td>
</tr>
<tr>
<td>14</td>
<td>29</td>
<td>6/6</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>6/6</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>6/6</td>
<td>1800</td>
</tr>
<tr>
<td>15</td>
<td>32</td>
<td>6/6</td>
<td>2100</td>
</tr>
</tbody>
</table>

Figure 3.81: Channel types and speeds for 2003 VDSL Olympics.
notches) – that is the PW criterion is satisfied. The power spectrum for an ISI channel for which the symbol/sampling rate has been altered for exact match to one (any) water-filling band $1/T^{opt}$ that must satisfy

$$\bar{E}_x|\Phi(\omega)|^2 + \frac{\sigma^2}{|H(\omega)|^2} = K$$

(3.726)

for all frequencies $\omega$ in the band (zeros are only allowed at infinitesimally narrow notch points, so earlier resampling may have necessarily occurred to avoid singularity). $\hat{H}$ represents the equivalent-white-noise channel $\hat{H} = \frac{H}{\sqrt{\bar{S}_n}}$. The MMSE-DFE fundamentally is determined by canonical factorization, which in the frequency domain has magnitudes related by

$$\hat{E}_x|\Phi(\omega)|^2 \cdot |H(\omega)|^2 + \sigma^2 = \frac{\gamma_0}{||p||^2} \cdot |G(\omega)|^2 .$$

(3.727)

Insertion of (3.726) into (3.727) yields

$$K \cdot |H(\omega)|^2 = \frac{\gamma_0}{||p||^2} \cdot |G(\omega)|^2 ,$$

(3.728)

illustrating a very unusual event that the overall equivalent equalized channel has the same shape as the channel itself. This happens when the input is water-filling. Thus, since the shape of $H$ and $G$ are the same, they differ only by an all-pass (phase-only) filter. Equivalently, the receiver’s MS-WMF is essentially $\Phi^{opt}$ times an all-pass filter. Such an all-pass can be implemented at the transmitter without increasing the transmit energy, leaving only the filter $\tilde{S}_n^{-1/2} \cdot \Phi^{opt}$ in the receiver. Such an all-pass phase shift never changes the performance of the DFE (or any infinite-length equalizer). The receiver’s remaining noise whitening filter and matched-filter-to-transmit-water-fill filter constitute the entire MS-WMF, perhaps resulting in an opportunity to reduce receiver complexity. The MMSE-DFE has slightly non-unity gain through the feed-forward filter and the overall channel gain is $\frac{SNR_{MMSE-DFE,U}}{SNR_{MMSE-DFE,U+1}}$, which is removed by anti-bias scaling. However, ignoring the bias removal and consequent effect on signal and error sequence, water-filling creates an equivalent overall channel that has the same shaping of signal as the original channel would have had. That is the MS-WMF-output signal (ignoring bias in error) has the same minimum-phase equivalent ISI as the original channel would have, or equivalently as a ZF-DFE would have generated. QED.

The MMSE-DFE however does have a lower MMSE and better SNR, even after bias removal. The difference in SNR is solely a function of the error signal and scaling. The ZF-DFE has white noise and no bias. The MMSE-DFE with water-filling has the same minimum-phase shaping, that is achieved by filtering the signal, reducing noise power, but then ensuring that the components of the error (including bias) add in a way that the original channel reappears as a result of all the filtering.

The minimum-phase equivalent of the channel alone $H$ is equal to $G$ via Equation (3.728). However $P_x(D)$, the canonical equivalent of the pulse response $\Phi \cdot H$ (which includes the water-filling shaping) is not the same unless water-filling shape is the special case of being flat. The precoder settings or equivalently unbiased feedback-section settings are not necessarily determined by $G$, but by $G_u$. $G_u$ differs from the minimum-phase channel equivalent.

Thus, the MMSE-DFE results in same equivalent channel as ZF-DFE under water-filling, but still performs slightly better unless water-filling is flat (essentially meaning the channel is flat). Often water-filling is very close to flat in practical situations or there is very little loss. Thus, it may be convenient to implement directly a ZF-DFE once the exact water-filling band of frequencies is known.
Theorem 3.12.3 (Worst-Case Noise Equivalence) Over any of the continuous bands for which the PW Criterion holds, if the noise power spectral density is the worst possible choice that minimizes mutual information for the channel, then ZF and MMSE are the same even if the noise is not zero.

**proof:** The minimization of the mutual information

\[ I = \frac{1}{2\pi} \int_{\Omega} \log_2 \left( 1 + \frac{S_x(\omega) \cdot |H(\omega)|^2}{S_n(\omega)} \right) d\omega \]  

(3.729)

with noise power constrained at

\[ \sigma^2 = \frac{1}{2\pi} \int_{\Omega} S_n(\omega)d\omega \]  

(3.730)

leads easily by differentiation with LaGrange (calculus of variations) constraint to the equation/solution

\[ \frac{1}{S_x(\omega) \cdot |H(\omega)|^2 + S_n(\omega)} - \frac{1}{S_n(\omega)} = K \forall \omega \in \Omega . \]  

(3.731)

The magnitude of the MS-WMF with noise whitening included is

\[ |FF|^2 = \frac{|H(\omega)|^2 \cdot S_x(\omega)}{S_n^2(\omega) \cdot \gamma_0 \cdot ||p||^2 |G(\omega)|^2} = \frac{|H(\omega)|^2 \cdot S_x(\omega)}{(S_x(\omega) \cdot |H(\omega)|^2 + S_n(\omega)) \cdot S_n(\omega)} 
= \frac{1}{S_n(\omega)} - \frac{1}{S_x(\omega) \cdot |H(\omega)|^2 + S_n(\omega)} = -K \]  

(3.732)

Thus, the MS-WMF section is exactly an all pass filter, meaning then ZF and MMSE-DFEs (the latter with bias-removal scaling) are the same. QED.

The above result holds for any input spectrum \( S_x(\omega) \) (and clearly the worst-case noise may vary with this input spectrum choice). In particular, this result does hold for the water-filling spectrum also. The latter result has broad implications in multi-user optimization of systems as in Chapter 14. It suggests that if the noise is chosen to be the worst possible spectrum (for a given power), receiver processing is unnecessary (as the all-pass could be implemented at the transmitter without loss and a precoder could also be used). Worst-case noise (unlike water-filling transmit spectra) is not good – it means that receiver filtering is essentially useless. The absence of receiver processing enables derivation of the best possible transmit spectra for each of the individual users in what are called “broadcast” multi-user systems (think downlink to many devices in Wi-Fi or downstream to many homes in crosstalking DSL binders) where coordination of receiver dimensions (users) is not physically possible.

3.12.8 Optimized Transmit Filter with Linear Equalization

Linear equalization is suboptimal and not canonical (on all but trivial channels), but an optimized transmit filter can be found nonetheless, and will improve the performance of a MMSE-LE. Thus subsection develops that optimized transmit filter. All the realization concerns with PWC satisfaction still apply to any optimized transmit filter, thus there will also be an associated optimum symbol rate(s) and carrier frequency(ies). For the LE case, the MMSE is revisited without the receiver matched-filter normalization to simplify the mathematical development\(^{44}\). Again calling the data input to the transmit filter

\(^{44}\)The normalization factor \( ||p|| \) depends on the transmit filter to be optimized, so it is simpler to avoid that normalization in the development.
with the filter output \( x_k \), this development will assume tacitly that the symbol rate(s) (and carrier frequency(ies) when complex passband) will be set to exactly match the derived optimized transmit filter. This subsumes all the previous development for MMSE-DFE where continuous time was initially assumed to position discrete time. All that will happen together here in one step because the reader can refer if more mathematical precision is desired to the MMSE-DFE development earlier in this section. Thus we will momentarily drop frequency, continuous \( \omega \) or discrete \( e^{-j\omega T} \), from the frequency-transform notation.

The error signal will be
\[
E = V - W \cdot Y ,
\]
where the MMSE-LE (including any matched filter an sampling) \( W \) in (3.735) is
\[
W = R_{VV}^{-1} \cdot R_{VY} .
\]
The channel output is
\[
Y = H \cdot \Phi \cdot V + N .
\]
The input/output cross-correlation is
\[
R_{VY} = \bar{\epsilon}_x \cdot H \cdot \Phi ,
\]
while the channel output autocorrelation is
\[
R_{YY} = \bar{\epsilon}_x \cdot |H|^2 \cdot |\Phi|^2 + \frac{N_0}{2} .
\]
The MMSE power spectral density is (recalling that \( \text{SNR} \triangleq \bar{\epsilon}_x/\sigma^2 \))
\[
R_{EE} = \bar{\epsilon}_x - W \cdot R_{VV} = \bar{\epsilon}_x - \frac{\bar{\epsilon}_x \cdot |H|^2 \cdot |\Phi|^2}{\bar{\epsilon}_x \cdot |H|^2 \cdot |\Phi|^2 + \frac{N_0}{2}}
\]
\[
= \frac{\bar{\epsilon}_x}{1 + \text{SNR} \cdot |H|^2 \cdot |\Phi|^2 + \frac{N_0}{2}}
\]
\[
= \frac{\frac{N_0}{2}}{|H|^2 \cdot |\Phi|^2 + \frac{1}{\text{SNR}}} .
\]
The MMSE in (3.743) can be minimized over the squared magnitude of the input filter \( |\Phi|^2 \geq 0 \) subject to the energy (basis-function normalization) constraint:
\[
\frac{1}{2\pi} \int |\Phi(\omega)|^2 d\omega = 1 .
\]
Calculus of variations with respect to the squared transmit spectrum \( |\Phi|^2 \) provides \( \lambda' = \lambda^{-1} \)
\[
0 = \frac{-N_0}{2} \cdot \lambda' \cdot |H|^2 + \frac{1}{\text{SNR}} \cdot |H|^2 \cdot |\Phi|^2 + \frac{1}{\text{SNR}^2} - \lambda' \cdot \frac{N_0}{2} \cdot |H|^2
\]
\[
|\Phi|^2 = \frac{1}{|H|^2} \cdot \left[ \sqrt{\lambda' \cdot \frac{N_0}{2} \cdot |H|^2 - \frac{1}{\text{SNR}}} \right] .
\]

where the value of the Lagrange multiplier-related-constant $c$ and the optimum symbol period $T_{opt}$ are found from the energy constraint in (3.744), along with the need for non-negative power spectral density, as

$$c = \left[ \frac{N_0}{2} \cdot \frac{T_{opt}}{2\pi} \int_{-\frac{\pi}{T_{opt}}}^{\frac{\pi}{T_{opt}}} |H(\omega)| \cdot d\omega \right]^{-1} \left[ 1 + \frac{T_{opt}}{2\pi \cdot SNR} \int_{-\frac{\pi}{T_{opt}}}^{\frac{\pi}{T_{opt}}} |H(\omega)|^2 d\omega \right].$$

(3.750)

The optimized transmit filter has some interesting interpretations. Basically the optimum transmit filter shape is the difference between a scaled $\frac{1}{|H|}$ and $\frac{1}{SNR} |H|^2$ where the scale constant $c$ is chosen so that the transmit filter has unit norm. There will be again an optimum symbol rate and carrier frequency (exact middle), as illustrated in Figure 3.82.

This joint solution for $c$ and $T_{opt}$ partitions, for a particular initial guess of $c$, frequency into narrow MT spectra and computing (3.749) for each of these small tones or bands, then retaining and sorting those with non-negative spectra and then summing (computing the integrals) them to test the power constraint. If too much power, then $c$ is reduced and the process repeated until the first situation where $c$ is exactly achieved. If too little power, then $c$ is increased and the process repeated until the power constraint is met. The author has humorously referred to this process as “slush packing” instead of “water-filling” on numerous occasions. The idea being that slush ice (think “icee” for movie/carnival goers) can be molded as shown in Figure 3.82. There can of course be multiple bands when the sorting is undone, but only one band is shown in Figure 3.82. The optimized transmit energy band is clearly not water-filling, although it does tend to favor bands with higher channel gain and to avoid bands with low channel gain (essentially attempting to avoid large noise enhancement by inverting a poor band) in the MMSE-LE.

The corresponding power spectrum is

$$R_{EE} = \frac{\frac{N_0}{2}}{|H|^2 \cdot |\Phi|^2 + \frac{1}{SNR}}$$

(3.751)

$$= \frac{\frac{N_0}{2}}{c \cdot |H| - \frac{1}{SNR} + \frac{1}{SNR}}$$

(3.752)
\[
\begin{align*}
\sigma_{MMSE-LE, v}^2 \text{ opt} &= \frac{\sigma_{\text{MMSE-LE}}^2}{c \cdot |H|} \\
\text{(SNR)}_{MMSE-LE, v}^\text{opt} &= \frac{\tilde{E}_x}{(\sigma_{MMSE-LE}^2)^\text{opt}} - 1,
\end{align*}
\]

which implies the channel-dependent noise enhancement with optimized transmit spectrum grows only as the square-root of the noise enhancement that occurred with fixed-transmitter MMSE-LE systems. The frequency-dependent SNR can also be found as

\[
\text{SNR}(\omega) = \text{SNR} \cdot |H(\omega)| \cdot c - 1
\]

on those frequencies in the used optimum band (and zero elsewhere). The equalized channel output can be found as

\[
W \cdot Y = \frac{c \cdot \text{SNR} \cdot |H| - 1}{c \cdot \text{SNR} \cdot |H|} \cdot V + W \cdot N
\]

basically showing the input \( V \) with the MMSE bias well known at this point of this chapter and the filtered noise at each frequency going into the SBS detector.

For the zero-forcing case, the equalizer SNR tends to infinity, and the transmit filter becomes a scaled square-root pre-inverse of the channel \( H \), essentially splitting the equalizer’s channel inversion between the transmit filter and the receiver’s MMSE-LE filter that also will be a scaled square-root inverse of the channel \( H \). The symbol rate will be that corresponding to the measure of the frequency band between the two points shown in Figure 3.82 except that the lower curve will be the horizontal axis line (and the symbol rate is chosen to make the transmit basis-function filter unit norm).

Another situation of interest is when the the receiver has no ability to equalize, but the transmitter can alter the transmit filter. This case is readily resolves into the MMSE-LE moving to the transmitter with a scaling constant used to prevent transmit energy constraints being violated (or equivalently the transmit filter has a scaled MMSE-LE shape where the scaling causes the filter to have unit norm).

Chapter 4 continues with a more complete discussion of all forms of transmit optimization, while Chapter 5 develops further the equivalent of finite-length transmit filter optimization in what is called generalized decision feedback equalization (GDFE).
Exercises - Chapter 3

3.1 Single Sideband (SSB) Data Transmission with no ISI - 8 pts

Consider the quadrature modulator shown in Figure 3.83.

![Quadrature modulator](image_url)

a. What conditions must be imposed on \( H_1(f) \) and \( H_2(f) \) for the output signal \( s(t) \) to have no spectral components between \(-f_c\) and \( f_c\)? Assume that \( f_c \) is larger than the bandwidth of \( H_1(f) \) and \( H_2(f) \). (2 pts.)

b. With the input signal in the form,

\[
x(t) = \sum_{n=-\infty}^{\infty} a_n \cdot \phi(t - nT)
\]

what conditions must be imposed on \( H_1(f) \) and \( H_2(f) \) if the real part of the demodulated signal is to have no ISI. (3 pts.)

c. Find the impulse responses \( h_1(t) \) and \( h_2(t) \) corresponding to the minimum bandwidth \( H_1(f) \) and \( H_2(f) \) that simultaneously satisfy Parts a and b. The answer can be in terms of \( \phi(t) \). (3 pts.)

3.2 Sampling Time and Eye Patterns - 9 pts

The received signal for a binary transmission system is,

\[
y(t) = \sum_{n=-\infty}^{\infty} a_n \cdot q(t - nT) + n(t)
\]

where \( a_n \in \{-1, +1\} \) and \( Q(f) \) is a triangular, i.e. \( q(t) = \text{sinc}^2(\frac{t}{T}) \). The received signal is sampled at \( t = kT + t_0 \), where \( k \) is an integer and \( t_0 \) is the sampling phase, \( |t_0| < \frac{1}{2}T \).

a. Neglecting the noise for the moment, find the peak distortion as a function of \( t_0 \). \textbf{Hint:} Use Parseval’s relation. (3 pts.)

b. For the following four binary sequences \( \{u_n\}_{-\infty}^{\infty}, \{v_n\}_{-\infty}^{\infty}, \{w_n\}_{-\infty}^{\infty}, \{x_n\}_{-\infty}^{\infty} \):

\[
\begin{align*}
  u_n &= -1 \quad \forall n, \\
  v_n &= +1 \quad \forall n, \\
  w_n &= \begin{cases} +1, & \text{for } n = 0 \\
                         -1, & \text{otherwise} \end{cases}, \text{ and} \\
  x_n &= \begin{cases} -1, & \text{for } n = 0 \\
                         +1, & \text{otherwise} \end{cases},
\end{align*}
\]

use the result of (a) to find expressions for the 4 outlines of the corresponding binary eye pattern. Sketch the eye pattern for these four sequences over two symbol periods \(-T \leq t \leq T\). (2 pts.)
c. Find the eye pattern’s width (horizontal) at its widest opening. (2 pts.)

d. If the noise variance is \( \sigma^2 \), find a worst case bound (using peak distortion) on the probability of error as a function of \( t_0 \). (2 pts.)

3.3 The Stanford EE379 Channel Model - 12 pts

For the ISI-model shown in Figure 3.1.3 with PAM and symbol period \( T \), \( \varphi(t) = \frac{1}{\sqrt{T}} \cdot \text{sinc}(\frac{t}{T}) \) and \( h(t) = \delta(t) - \frac{1}{2} \delta(t - T) \):

a. Determine \( p(t) \), the pulse response. (1 pt)

b. Find \( ||p|| \) and \( \varphi_p(t) \). (2 pts.)

c. (2 pts.) Find \( q(t) \), the function that characterizes how one symbol interferes with other symbols. Confirm that \( q(0) = 1 \) and that \( q(t) \) is hermitian (conjugate symmetric). A useful observation is

\[
\text{sinc}\left(\frac{t + mT}{T}\right) \ast \text{sinc}\left(\frac{t + nT}{T}\right) = T \cdot \text{sinc}\left(\frac{t + (m + n)T}{T}\right).
\]

d. Use Matlab to plot \( q(t) \). This plot may use \( T = 10 \) and show \( q(t) \) for integer values of \( t \). With \( y(t) \) sampled at \( t = kT \) for some integer \( k \), how many symbols are distorted by a given symbol? How will a given positive symbol affect the distorted symbols? Specifically, will they be increased or decreased? (2 pts.)

e. The remainder of this problem considers 8 PAM with \( d = 1 \). Determine the peak distortion, \( D_p \). (2 pts.)

f. Determine the MSE distortion, \( D_{mse} \) and compare with the peak distortion. (2 pts.)

g. Find an approximation to the probability of error (using the MSE distortion). The answer will be in terms of \( \sigma \).(1 pt.)

3.4 Bias and SNR - 8 pts

This problem uses the results and specifications of Problem 3.3 with 8 PAM and \( d = 1 \) and sets

\[
\frac{1}{2} N_0 = \sigma^2 = 0.1.
\]

The detector first scales the sampled output \( y_k \) by \( \alpha \) and chooses the closest \( x \) to \( y_k \cdot \alpha \) as illustrated in Figure 3.2.2. Please express your SNR’s below both as a ratio of powers and in dB.

a. For which value of \( \alpha \) is the receiver unbiased? (1 pt.)

b. For the value of \( \alpha \) found in (a), find the receiver signal-to-noise ratio, \( SNR(R) \). (2 pts.)

c. Find the value of \( \alpha \) that maximizes \( SNR_R \) and the corresponding \( SNR(R) \). (2 pts.)

d. Show that the receiver found in the previous part is biased. (1 pt.)

e. Find the matched filter bound on signal-to-noise ratio, \( SNR_{MF} \). (1 pt.)

f. Discuss the ordering of the three \( SNR \)’s you have found in this problem. Which inequalities will always be true? (1 pt.)

3.5 Raised Cosine Pulses with Matlab - 8 pts

This problem uses the .m files from the instructors EE379A web page at http://web.stanford.edu/group/cioffi/ee379a/homework.html.
a. The formula for the raised cosine pulse is in Equation (3.79) and repeated here as:

\[ q(t) = \text{sinc} \left( \frac{t}{T} \right) \cdot \left[ \cos \left( \frac{\alpha \pi t}{T} \right) \right. \]

\[ \left. \frac{1}{1 - (2\alpha t)^2} \right] \]

There are three values of \( t \) that would cause a program, such as Matlab, difficulty because of a division by zero. Identify these trouble spots and evaluate what \( q(t) \) should be for these values. (2 pts)

b. Having identified those three trouble spots, a function to generate raised-cosine pulses is a straightforward implementation of the above equation and is in \texttt{mk_rcpulse.m} from the web site. Executing \( q = \text{mk_rcpulse}(a) \) will generate a raised-cosine pulse with \( \alpha = a \). The pulses generated by \texttt{mk_rcpulse} assume \( T = 10 \) and are truncated to 751 points. (2 pts)

Use \texttt{mk_rcpulse} to generate raised cosine pulses with 0\%, 50\%, and 100\% excess bandwidth and plot them with \texttt{plt_pls_lin}. The syntax is \texttt{plt_pls_lin(q\_50, 'title')} where \( q\_50 \) is the vector of pulse samples generated by \texttt{mk_rcpulse(0.5)} (50\% excess bandwidth). Show results and discuss any deviations from the ideal expected frequency-domain behavior. (3 pts)

d. The function \texttt{plt_qk (q, 'title')} plots \( q(k) \) for sampling at the optimal time and for sampling that is off by 4 samples (i.e., \( q(kT) \) and \( q(kT + 4) \) where \( T = 10 \)). Use \texttt{plt_qk} to plot \( q_k \) for 0\%, 50\%, and 100\% excess bandwidth raised cosine pulses. Discuss how excess bandwidth affects sensitivity of ISI performance to sampling at the correct instant. (3 pts)

### 3.6 Noise Enhancement: MMSE-LE vs ZFE - 23 pts

This problem explores the channel with

\[ \|p\|^2 = 1 + aa^* = 1 + |a|^2 \]

\[ Q(D) = a^*D^{-1} + \frac{\|p\|^2 + aD}{\|p\|^2} \]

\[ 0 \leq |a| < 1 \]

a. (2 pts) Find the zero forcing and minimum mean square error linear equalizers \( W_{\text{ZFE}}(D) \) and \( W_{\text{MMSE-LE}}(D) \). Use the variable \( b = \|p\|^2 \cdot \left( 1 + \frac{1}{\text{SNR}_{\text{MFB}}} \right) \) in your expression for \( W_{\text{MMSE-LE}}(D) \).

b. (6 pts) By substituting \( e^{-j\omega T} = D \) (with \( T = 1 \)) and using \( \text{SNR}_{\text{MFB}} = 10 \cdot \|p\|^2 \), use Matlab to plot (lots of samples of) \( W(e^{j\omega}) \) for both ZFE and MMSE-LE for \( a = .5 \) and \( a = .9 \). Discuss the differences between the plots.

c. (3 pts) Find the roots \( r_1, r_2 \) of the polynomial

\[ aD^2 + bD + a^* \]

\[ 0 \]

Show that \( b^2 - 4|a|^2 \) is always a real positive number (for \( |a| \neq 1 \)). \textit{Hint:} Consider the case where \( \frac{1}{\text{SNR}_{\text{MFB}}} = 0 \). Let \( r_2 \) be the root for which \( |r_2| < |r_1| \). Show that \( r_1 r_2^* = 1 \).

d. (2 pts) Use the previous results to show that for the MMSE-LE

\[ W(D) = \frac{\|p\|}{a} \frac{D}{(D - r_1)(D - r_2)} = \frac{\|p\|}{a(r_1 - r_2)} \left( \frac{r_1}{D - r_1} - \frac{r_2}{D - r_2} \right) \]  \hspace{1cm} (3.758)

e. (2 pts) Show that for the MMSE-LE, \( w_0 = \frac{\|p\|}{\sqrt{b^2 - 4|a|^2}} \). By taking \( \frac{1}{\text{SNR}_{\text{MFB}}} = 0 \), show that for the ZFE, \( w_0 = \frac{\|p\|}{1 - |a|^2} \).
f. (4 pts) For $\bar{E}_x = 1$ and $\sigma^2 = 0.1$ find expressions for $\sigma^2_{ZF-E}$, $\gamma_{MLSE}$, $\gamma_{ZFE}$, and $\gamma_{MMSE-LE}$.

g. (4 pts) Find $\gamma_{ZFE}$ and $\gamma_{MMSE-LE}$ in terms of the parameter $a$ and calculate for $a = 0, 0.5, 1$. Sketch $\gamma_{ZFE}$ and $\gamma_{MMSE-LE}$ for $0 \leq a < 1$.

3.7 DFE is Even Better - 8 pts

a. (2 pts) For the channel of Problem 3.6, show that the canonical factorization is

$$Q(D) + \frac{1}{SNR_{MFB}} = \gamma_0 \cdot (1 - r_2 D^{-1})(1 - r_2^* D).$$

What is $\gamma_0$ in terms of $a$ and $b$? Please don’t do this from scratch; use results of Problem 3.6.

b. (2 pts) Find $B(D)$ and $W(D)$ for the MMSE DFE.

c. (4 pts) Give an expression for $\gamma_{MMSE-DFE}$. Compute its values for $a = 0, 0.5, 1$ for the $\bar{E}_x$ and $\sigma^2$ of problem 3.6. Sketch $\gamma_{MMSE-DFE}$ as in problem 3.6. Compare with your sketches from Problem 3.6.

3.8 Noise Predictive DFE - 11 pts

![Figure 3.84: Noise-Predictive DFE](image)

With the equalizer shown in Figure 3.84 with $B(D)$ restricted to be causal and monic and all decisions presumed correct, use a MMSE criterion to choose $U(D)$ and $B(D)$ to minimize $E[|x_k - z_k'|^2]$. ($x_k$ is the channel input.)

a. (5 pts) Show this equalizer is equivalent to a MMSE-DFE and find $U(D)$ and $B(D)$ in terms of $G(D)$, $Q(D)$, $\|p\|$, $\bar{E}_x$, $SNR_{MFB}$ and $\gamma_0$.

b. (2 pts) Relate $U_{NPDFE}(D)$ to $W_{MMSE-LE}(D)$ and to $W_{MMSE-DFE}(D)$.

c. (1 pt) Without feedback ($B(D) = 1$), what does this equalizer become?

d. (1 pt) Interpret the name “noise-predictive” DFE by explaining what the feedback section is doing.

e. (2 pts) Is the NPDFE biased? If so, show how to remove the bias.

3.9 Receiver SNR Relationships - 11 pts
a. (3 pts) Recall that:

\[ Q(D) + \frac{1}{\text{SNR}_{MFB}} = \gamma_0 G(D) G^*(D^{-*}) \]

Show that:

\[ 1 + \frac{1}{\text{SNR}_{MFB}} = \gamma_0 \|g\|^2 \]

b. (3 pts) Show that:

\[ \|g\| \geq 1 \]

with equality if and only if \( Q(D) = 1 \) (i.e. if the channel is flat) and therefore:

\[ \gamma_0 \leq 1 + \frac{1}{\text{SNR}_{MFB}} \]

c. (3 pts) Let \( x_0 \) denote the time-zero value of the sequence \( X \). Show that:

\[
\begin{bmatrix}
1 \\
Q(D) + \frac{1}{\text{SNR}_{MFB}}
\end{bmatrix}
\geq \frac{1}{\gamma_0}
\]

and therefore:

\[ \text{SNR}_{MMSE-LE,U} \leq \text{SNR}_{MMSE-DFE,U} \]

(with equality if and only if \( Q(D) = 1 \), that is \( Q(\omega) \) has vestigial symmetry.)

d. (2 pts) Use the results of parts b) and c) to show:

\[ \text{SNR}_{MMSE-LE,U} \leq \text{SNR}_{DFE,U} \leq \text{SNR}_{MFB} \]

3.10 Bias and Probability of Error - 7 pts

This problem illustrates that the best unbiased receiver has a lower \( P_e \) than the (biased) MMSE receiver using the 3-point PAM constellation in Figure 3.85.

![Figure 3.85: A three point PAM constellation.](image)

The AWGN \( n_k \) has

\[ \frac{\mathcal{N}_0}{2} = \sigma^2 = 0.1. \]

The inputs are independent and identically distributed uniformly over the three possible values. Also, \( n_k \) has zero mean and is independent of \( x_k \).

a. (1 pt) Find the mean square error between \( y_k = x_k + n_k \) and \( x_k \). (easy)

b. (1 pt) Find the exact \( P_e \) for the ML detector of the unbiased values \( y_k \).
c. (2 pts) Find the scale factor \( \alpha \) that will minimize the mean square error

\[
E[e_k^2] = E[(x_k - \alpha \cdot y_k)^2].
\]

Prove that this \( \alpha \) does in fact provide a minimum by taking the appropriate second derivative.

d. (2 pts) Find \( E[e_k^2] \) and \( P_e \) for the scaled output \( \alpha \cdot y_k \). When computing \( P_e \), use the decision regions for the unbiased ML detector of Part b. The biased receiver of part (c) has a \( P_e \) that is higher than the (unbiased) ML detector of Part b, even though it has a smaller squared error.

e. (1 pt) Where are the optimal decision boundaries for detecting \( x_k \) from \( \alpha \cdot y_k \) (for the MMSE \( \alpha \)) in Part c? What is the probability of error for these decision boundaries?

3.11 Bias and the DFE - 4 pts

For the DFE of Problem 3.7, the canonical factorization’s scale factor was found as:

\[
\gamma_0 = \frac{b + \sqrt{b^2 - 4|a|^2}}{2(1 + |a|^2)}.
\]

a. (2 pts) Find \( G_u \) in terms of \( a \), \( b \), and \( r_2 \). Use the same \( SNR_{MFB} \) as in Prob 3.7.

b. (1 pt) Draw a DFE block diagram with scaling after the feedback summation.

c. (1 pt) Draw a DFE block diagram with scaling before the feedback summation.

3.12 Zero-Forcing DFE - 7 pts

This problem again uses the general channel model of Problems 3.6 and 3.7 with \( \bar{E}_x = 1 \) and \( \sigma^2 = 0.1 \).

a. (2 pts) Find \( \eta_0 \) and \( P_c(D) \) so that

\[
Q(D) = \eta_0 \cdot P_c(D)P_c^*(D^*)
\]

b. (2 pts) Find \( B(D) \) and \( W(D) \) for the ZF-DFE.

c. (1 pt) Find \( \sigma_{ZF-DFE}^2 \)

d. (2 pts) Find the loss with respect to \( SNR_{MFB} \) for \( a = 0, .5, 1 \). Sketch the loss for \( 0 \leq a < 1 \).

3.13 Complex Baseband Channel - 6 pts

The pulse response of a channel is band-limited to \( \frac{\pi}{T} \) and has

\[
\begin{align*}
p_0 &= \frac{1}{\sqrt{T}} \cdot (1 + 1.1j) \\
p_1 &= \frac{1}{\sqrt{T}} \cdot (0.95 + 0.5j) \\
p_k &= 0 \quad \text{for} \ k \neq 0, 1
\end{align*}
\]

where \( p_k = p(kT) \) and \( SNR_{MFB} = 15 \) dB.

a. (2 pts) Find \( \|p\|^2 \) and \( a \) so that

\[
Q(D) = \frac{a^*D^{-1} + \|p\|^2 + aD}{\|p\|^2}
\]

b. (2 pts) Find \( SNR_{MMSE-LE,U} \) and \( SNR_{MMSE-DFE,U} \) for this channel. Use the results of Problems 3.6 and 3.7 wherever appropriate. Since \( \|p\|^2 \neq 1 + |a|^2 \), some care should be exercised in using the results of Problems 3.6 and 3.7.
c. (2 pts) Use the SNR’s from Part b to compute the NNUB on $P_e$ for 4-QAM with the MMSE-LE and the MMSE-DFE.

### 3.14 Tomlinson Precoding - 13 pts

This problem uses a $1 + 0.9D$ channel (i.e. the $1 + aD$ channel that we have been exploring at length with $a = 0.9$) with $\sigma^2 = 0$ and with the 4-PAM constellation having $x_k \in \{-3, -1, 1, 3\}$.

a. (5 pts) Design a Tomlinson precoder and its associated receiver for this system. This system will be fairly simple. How would your design change if noise were present in the system?

b. (5 points) Implement the precoder and its associated receiver using any method you would like with no noise for the following input sequence:

$$\{3, -3, 1, -1, 3, -3, -1\}$$

Also compute the corresponding output of the receiver, assuming that the symbol $x' = -3$ was sent just prior to the input sequence.

c. (3 pts) Again with zero noise, remove the modulo operations from the precoder and the receiver. For this modified system, compute the output of the precoder, the output of the channel, and the output of your receiver for the same inputs as in the previous part.

Did the system still work? What changed? What purpose do the modulo operators serve?

### 3.15 Flexible Precoding - 13 pts

This problem uses a $1 + 0.9D$ channel (i.e. the $1 + aD$ channel that we have been exploring at length with $a = 0.9$) with $\sigma^2 = 0$ and with the 4-PAM constellation having $x_k \in \{-3, -1, 1, 3\}$.

a. (5 pts) Design a Flexible precoder and its associated receiver for this system. This system will be fairly simple. How would your design change if noise were present in the system?

b. (5 points) Implement the precoder and its associated receiver using any method you would like with no noise for the following input sequence:

$$\{3, -3, 1, -1, 3, -3, -1\}$$

Also compute the corresponding output of the receiver, assuming that the symbol $x' = -3$ was sent just prior to the input sequence.

c. (3 pts) Again with zero noise, remove the modulo operations from the precoder and the receiver. For this modified system, compute the output of the precoder, the output of the channel, and the output of your receiver for the same inputs as in the previous part.

Did the system still work? What changed? What purpose do the modulo operators serve?

### 3.16 Finite-Length Equalization and Matched Filtering - 5 pts

This problem explores the optimal FIR MMSE-LE without assuming an explicit matched filter. However, there will be a “matched filter” anyway as a result of this problem. This problem uses a system whose pulse response is band limited to $|\omega| < \frac{\pi}{T}$ that is sampled at the symbol rate $T$ after passing through an anti-alias filter with gain $\sqrt{T}$.

a. (3 pts)

Show that

$$w = R_x R_y^{-1} = (0, \ldots, 0, 1, 0, \ldots, 0) \Phi_p^* \left( ||p|| \cdot \left( Q + i \frac{1}{SNR_{MFB}} I \right) \right)^{-1}.$$  

where

$$Q = \frac{P^* P}{||p||^2}$$
b. (2 pts) Which terms in Part a correspond to the matched filter? Which terms correspond to the infinite length $W_{\text{MMSE-LE}}$?

3.17 Finite-Length Equalization and MATLAB - 5 pts
The $1 + 0.25 D$ channel with $\sigma^2 = .1$ has an input with $\tilde{E}_x = 1$.

a. (1 pt) Find $\text{SNR}_{\text{MMSE-LE},U}$ for the infinite-length filter.

b. (2 pts) This problem uses the MATLAB program `mmsele`. This program is interactive, so just type `mmsele` and answer the questions. When it asks for the pulse response, you can type $[1 \hspace{1em} 0.25]$ or $p$ if you have defined $p = [1 \hspace{1em} 0.25]$. Use `mmsele` to find the best $\Delta$ and the associated $\text{SNR}_{\text{MMSE-LE},U}$ for a 5-tap linear equalizer. Compare with the value from Part a. How sensitive is performance to $\Delta$ for this system?

c. (2 pts) Plot $|P(e^{jwT})|$ and $|P(e^{jwT})W(e^{jwT})|$ for $w \in [0, \frac{\pi}{T}]$. Discuss the plots briefly.

3.18 Computing Finite-Length equalizers - 11 pts
This problem uses the following system description:

$$\tilde{E}_x = 1 \quad \frac{N_0}{2} = \frac{1}{8} \quad \phi(t) = \frac{1}{\sqrt{T}} \text{sinc} \left(\frac{t}{T}\right) \quad h(t) = \delta(t) - 0.5\delta(t - T) \quad l = 1$$

Matlab may assist any matrix manipulations in completing this problem. Answers may be validated with `dfecolor.m`. However, this problem’s objective is to explore the calculations in detail.

a. (2 pts) The anti-alias filter is perfect (flat) with gain $\sqrt{T}$. Find $\tilde{p}(t) = (\phi(t) * h(t))$ and $\|\tilde{p}\|^2$, corresponding to the discrete-time channel

$$y_k = x_k - 0.5 \cdot x_{k-1} + n_k \quad \text{(3.759)}$$

Also, find $\|\tilde{P}(D)\|^2 = \sum_k |\tilde{p}_k|^2$.

b. (1 pt) Compute $\text{SNR}_{\text{MFB}}$ for this channel.

c. (2 pts) Design a 3 tap FIR MMSE-LE for $\Delta = 0$.

d. (1 pt) Find the $\sigma^2_{\text{LE}}$ for the equalizer of the previous part.

e. (2 pts) Design a 3 tap FIR ZF-LE for $\Delta = 0$.

f. (1 pt) Find the associated $\sigma^2_{\text{ZF}}$.

g. (2 pts) Design an MMSE-DFE which has 2 feedforward taps and 1 feedback tap. Again, assume that $\Delta = 0$.

h. (2 pts) Compute the unbiased SNR’s for the MMSE-LE and MMSE-DFE. Compare these two SNR’s with each other and the $\text{SNR}_{\text{MFB}}$.

3.19 Equalizer for Two-Band Channel with Notch - 10 pts

![Figure 3.86: Channel for Problem 3.19.](image)
Figure 3.86 has AWGN channel with pulse response $p(t) = \frac{1}{\sqrt{T}} \cdot \left[ \text{sinc} \left( \frac{t}{T} \right) - \text{sinc} \left( \frac{t-4T}{T} \right) \right]$. The receiver anti-alias filter has gain $\sqrt{T}$ over the entire Nyquist frequency band, $-1/2T < f < 1/2T$, and zero outside this band. The filter is followed by a $1/T$ rate sampler so that the sampled output has

$$y(D) = (1 - D^4) \cdot x(D) + n(D) \quad (3.760)$$

$x(D)$ is the $D$-Transform of the sequence of $M$-ary PAM input symbols, and $n(D)$ is the $D$-Transform of the Gaussian noise sample sequence. The noise autocorrelation function is $R_{nn}(D) = N_0^2$. Further equalization of $y(D)$ is in discrete time where (in this case) the matched filter and equalizer discrete responses (i.e., $D$-Transforms) can be combined into a single discrete-time response. The target $P_e$ is $10^{-3}$.

Let $\frac{N_0}{2} = -100$ dBm/Hz ($0$ dBm = 1 milliwatt = .001 Watt) and let $1/T = 2$ MHz.

a. Sketch $|P(e^{-j\omega T})|^2$. (2 pts)

b. What kind of equalizer should be used on this channel? (2 pts)

c. For a ZF-DFE, find the transmit symbol mean-square value (i.e., the transmit energy) necessary to achieve a data rate of 6 Mbps using PAM and assuming a probability of symbol error equal to $10^{-3}$. (4 pts)

d. For the transmit energy in Part c, how would a MMSE-DFE perform on this channel? (2 pts)

3.20 $FIR$ Equalizer Design - 14 pts

A symbol-spaced FIR equalizer ($l = 1$) is applied to a stationary sequence $y_k$ at the sampled output of an anti-alias filter, which produces a discrete-time IIR channel with response given by

$$y_k = a \cdot y_{k-1} + b \cdot x_k + n_k \quad (3.761)$$

where $n_k$ is white Gaussian noise.

The SNR (ratio of mean square $x$ to mean square $n$) is $\frac{\bar{E}_x}{N_0} = 20$ dB. $|a| < 1$, and both $a$ and $b$ are real. For all parts of this question, choose the best $\Delta$ where appropriate.

a. Design a 2-tap FIR ZFE. (3 pts)

b. Compute the SNR$_{zfe}$ for part a. (2 pts)

c. Compare the answer in Part b with that of the infinite-length ZFE. (1 pt)

d. Let $a = .9$ and $b = 1$. Find the 2-tap FIR MMSE-LE. (4 pts)

e. Find the SNR$_{MMSE-LE, U}$ for Part d. (2 pts)

f. Find the SNR$_{zfe}$ for an infinite-length ZF-DFE. How does the SNR compare to the matched-filter bound if there is no information loss incurred in the symbol-spaced anti-alias filter? Use $a = .9$ and $b = 1$. (2 pts)

3.21 ISI quantification - 8 pts

For the channel $P(\omega) = \sqrt{T}(1 + .9 \cdot e^{j\omega T}) \forall |\omega| < \pi/T$ studied repeatedly in this chapter, use binary PAM with $\bar{E}_x = 1$ and $\text{SNR}_m = 10$ dB. Remember that $q_0 = 1$.

a. Find the peak distortion, $D_p$. (1 pt)

b. Find the peak-distortion bound on $P_e$. (2 pts)

c. Find the mean-square distortion, $D_{MS}$. (1 pt)

d. Approximate $P_e$ using the $D_{MS}$ of Part c. (2 pts)
e. ZFE: Compare $P_e$ in part d with the $P_e$ for the ZFE. Compute SNR difference in dB between the
SNR based on mean-square distortion implied in Parts c and d and SNR$_{ZFE}$. (hint, see example
in this chapter for SNR$_{ZFE}$) (2 pts)

3.22 Precoding - 12 pts
For an ISI channel with

$$Q(D) + \frac{1}{\text{SNR}_{MFB}} = .82 \cdot \left[ \frac{j}{2} D^{-1} + 1.25 - \frac{j}{2} D \right],$$

and $\frac{\bar{E}_x}{2} = 100$.

a. Find SNR$_{MFB}$, $\|p^2$, and SNR$_{MMSE-DFE}$. (2 pts)

b. Find $G(D)$ and $G_U(D)$ for the MMSE-DFE. (1 pt)

c. Draw a Tomlinson-Harashima precoder, showing from $x_k$ through the decision device in the receiver
in your diagram (any $M$). (2 pts)

d. Let $M = 4$ in the precoder of Part c. Find $P_e$. (2 pts)

e. Design (show/draw) a Flexible precoder, showing from $x_k$ through the decision device in the
receiver in your diagram (any $M$). (2 pts)

f. Let $M = 4$ the precoder of Part e. Find $P_e$. (2 pts)

3.23 Finite-Delay Tree Search - 17 pts
A channel with multipath fading has one reflecting path with gain (voltage) 90% of the main path
with binary input. The relative delay on this path is approximately $T$ seconds, but the carrier sees a
phase-shift of $-60^\circ$ that is constant on the second path.

Use the model

$$P(\omega) = \begin{cases} \sqrt{T} (1 + ae^{-j\omega T}) & |\omega| < \pi/T \\ 0 & \text{elsewhere} \end{cases}$$

to approximate this channel with $\frac{N_0}{2} = .0181$ and $\bar{E}_x = 1$.

a. Find a. (1 pt)

b. Find SNR$_{MFB}$. (1 pt)

c. Find $W(D)$ and SNR$_{MMSE-LE,U}$ for the MMSE-LE. (3 pts)

d. Find $W(D)$, $B(D)$, and SNR$_{MMSE-DFE,U}$ for the MMSE-DFE. (3 pts)

e. Design a simple method to compute SNR$_{ZF-DFE}$ (2 pts).

f. A finite-delay tree search detector at time $k$ decides $\hat{x}_{k-1}$ is shown in Figure 3.87 and chooses $\hat{x}_{k-1}$
to minimize

$$\min_{\hat{x}_k, \hat{x}_{k-1}} |z_k - \hat{x}_k - a\hat{x}_{k-1}|^2 + |z_{k-1} - \hat{x}_{k-1} - a\hat{x}_{k-2}|^2,$$

where $\hat{x}_{k-2} = x_{k-2}$ by assumption.
Figure 3.87: Finite-delay tree search.

How does this FDTS compare with the ZF-DFE (better or worse)? (1 pt)

g. Find an SNR_{fds} for this channel. (3 pts)

h. Could you generalize SNR in part g for FIR channels with \( \nu \) taps? (3 pts)

3.24 Peak Distortion - 5 pts

Peak Distortion can be generalized to channels without receiver matched filtering (that is, \( \varphi_p(-t) \) is a lowpass anti-alias filter). A channel has

\[
y(D) = x(D) \cdot (1 - 0.5D) + N(D)
\]  

after sampling, where \( N(D) \) is discrete AWGN. \( P(D) = 1 - 0.5D \).

a. Write \( y_k \) in terms of \( x_k, x_{k-1} \) and \( n_k \). (hint, this is easy. 1 pt)

b. The peak distortion for such a channel is \( D_p = \Delta |x_{\text{max}}| \cdot \sum_{m \neq 0} |p_m| \). Find this \( D_p \) for this channel if \( |x_{\text{max}}| = 1 \). (2 pts)

c. Suppose \( \frac{N_0}{2} \), the mean-square of the sample noise, is .05 and \( \tilde{E}_x = 1 \). What is \( P_e \) for symbol-by-symbol detection on \( y_k \) with PAM and \( M = 2 \)? (2 pts)

3.25 Equalization of Phase Distortion - 13 pts

An ISI channel has \( p(t) = 1/\sqrt{T} \left[ \text{sinc}(t/T) - j\text{sinc}((t - T)/T) \right] \), \( \tilde{E}_x = 1 \), and \( \frac{N_0}{2} = .05 \).

a. What is SNR_{MFB}? (1 pt)

b. Find the ZFE. (2 pts)

c. Find the MMSE-LE. (2 pts)

d. Find the ZF-DFE. (2 pts)

e. Find the MMSE-DFE. (2 pts)

f. Draw a diagram illustrating the MMSE-DFE implementation with Tomlinson precoding. (2 pts)

g. For \( P_e < 10^{-6} \) and using square or cross QAM, choose a design and find the largest data rate you can transmit using one of the equalizers above. (2 pts)

3.26 Basic QAM Systems Revisited with Equalization Background - 6 pts

An AWGN channel has \( \tilde{E}_x = -40 \) dBm/Hz and \( \frac{N_0}{2} = -66 \) dBm/Hz. The symbol rate is initially fixed at 5 MHz, and the desired \( P_e = 10^{-6} \). Only QAM CR or SQ constellations may be used.
a. What is the maximum data rate? (1 pt)

b. What is the maximum data rate if the desired $P_e < 10^{-7}$? (1 pt)

c. What is the maximum data rate if the symbol rate can be varied (but transmit power is fixed at the same level as was used with a 5 MHz symbol rate)? (1 pt)

d. Suppose the pulse response of the channel changes from $p(t) = \frac{1}{\sqrt{T}} \cdot \text{sinc} \left( \frac{t}{T} \right)$ to $\frac{1}{\sqrt{T}} \cdot \text{sinc} \left( \frac{t}{T} \right) + 0.9 \cdot \frac{1}{\sqrt{T}} \cdot \text{sinc} \left( \frac{t + T}{T} \right)$, and Question a is revisited. With respect to your answer in Question a, compare the data rate for a ZF-DFE applied to this channel where $1/T = 5$ MHz. (1 pt)

e. Under the exact same change of Question d, compare the data rate with respect to your answer in Question c. (1 pt)

f. Suppose an optimum ML detector is used (observes all transmissions and makes one large decision on the entire transmitted message) with any pulse response that is a one-to-one transformation on all possible QAM inputs with $\|p\|^2 = 1$. Compare the probability of error with respect to that of the correct response to Question a. (1 pt)

3.27 Expert Understanding - 10 pts

A channel baseband transfer function for PAM transmission is shown in Figure 3.88. PAM transmission with sinc basis function is used on this channel with a transmit power level of 10 mW = $E_x/T$. The one-sided AWGN psd is -100 dBm/Hz.

a. Let the symbol rate be 20 MHz - find $\text{SNR}_{\text{MFB}}$. (1 pt)

b. For the same symbol rate as Part a, what is $\text{SNR}_{\text{MMSE-LE}}$? (1 pt)

c. For the equalizer in Part b, what is the data rate for PAM if $P_e \leq 10^{-6}$? (1 pt)

d. Let the symbol rate be 40 MHz - find the new $\text{SNR}_{\text{MFB}}$. (2 pts)

e. Draw a the combined shape of the matched-filter and feedforward filter for a ZF-DFE corresponding to the new symbol rate of Part d. (1 pt)
f. Estimate $\text{SNR}_{\text{MMSE-DFE,U}}$ for the new symbol rate, assuming a MMSE-DFE receiver is used. (2 pts) (Hint: the relationship of this transfer function to the channel is often used as an example in Chapter 3).

g. What is the new data rate for this system at the same probability of error as Part c - compare with the data rate of part c. (1 pt)

h. What can you conclude about incurring ISI if the transmitter is allowed to vary its bandwidth? (2 pts)

3.28 Infinite-Length EQ - 10 pts

An ISI channel with PAM transmission has channel correlation function given by

$$Q(D) = \frac{.19}{(1 + .9D)(1 + .9D^{-1})},$$

with $\text{SNR}_{\text{MFB}} = 10 \text{ dB}$, $\mathcal{E}_x = 1$, and $\|p\|^2 = \frac{1}{19}$.

a. (3 pts) Find $W_{\text{ZF}}(D)$, $\sigma^2_{\text{ZF}}$, and $\text{SNR}_{\text{ZF}}$.

b. (1 pt) Find $W_{\text{MMSE-LE}}(D)$

c. (3 pts) Find $W_{\text{ZF-DFE}}(D)$, $B_{\text{ZF-DFE}}(D)$, and $\text{SNR}_{\text{ZF-DFE}}$.

d. (3 pts) Find $W_{\text{MMSE-DFE}}(D)$, $B_{\text{MMSE-DFE}}(D)$, and $\text{SNR}_{\text{MMSE-DFE,U}}$.

3.29 Finite-Length EQ - 9 pts

A baseband channel is given by

$$P(f) = \begin{cases} \sqrt{T} \cdot (1 - .\mathcal{E} e^{2\pi f T}) & |f| < \frac{5}{T} \\ 0 & |f| \geq \frac{5}{T} \end{cases}$$

and finite-length equalization is used with anti-alias perfect LPR with gain $\sqrt{T}$ followed by symbol-rate sampling. After the sampling, a maximum complexity of 3 taps TOTAL over a feedforward filter and a feedback filter can be used. $\mathcal{E}_x = 2$ and $\frac{N_0}{2} = .05$ for a symbol rate of $1/T = 100kHz$ is used with PAM transmission. Given the complexity constraint, find the highest data rate achievable with PAM transmission when the corresponding probability of symbol error must be less than $10^{-5}$.

3.30 Infinite-length Channel and Equalizer - 10 pts

The pulse response on a channel with additive white Gaussian noise is $p(t) = e^{-at} \cdot u(t)$, where $u(t)$ is the unit step response. $a > 0$ and the two symbol inputs per dimension are $\pm 1$. Only QAM CR or SQ constellations allowed.

a. What is the argument to the Q-function in $\bar{P}_e = Q(x)$ at very low symbol rates? (1 pt)

b. For $\bar{P}_e \approx 10^{-6}$, $\frac{N_0}{2} = .05$, and very low symbol period, what is the value of $a$? (1 pt)

For the remainder of this problem, let $T = 1$.

c. Find $Q(D)$. (1 pt)

d. Find $\mathcal{D}_{m_s}$. (1 pt)

e. Find the receiver normalized matched filter prior to any symbol-rate sampling device. (1 pt)

f. Find $\text{SNR}_{\text{MFB}}$. Let $\text{SNR} = \mathcal{E}_x/\frac{N_0}{2}$. (1 pt)

g. Find the loss of the ZF-DFE with respect to the MFB. (1 pt)
h. Find SNR_{MMSE-DFE,U} for \( a = 1 \) and SNR = 20 dB. (3 pts)

i. What would an increase in \( a > 1 \) with respect to Part h do to the gap in performance between the ZF-DFE and the MMSE-DFE? (2 pts)

3.31 Diversity Channel - 6 pts

This problem compares the channel studied throughout this text with the same input energy \( \bar{E}_x = 1 \) and noise PSD of \( \frac{N_0}{T} = .181 \) versus almost the same channel except that two receivers independently receive:

- an undistorted delayed-by-\( T \) signal (that is \( P(D) = D \)), and
- a signal that is not delayed, but reduced to 90% of its amplitude (that is \( P(D) = .9 \)).

On both paths, the channel passes nonzero energy only between zero frequency and the Nyquist frequency.

a. Find \( Q(D), \| p \|, \) and SNR_{MFB} for the later diversity channel. (3 pts)

b. Find the performance of the MMSE-LE, MMSE-DFE for the later diversity channel and compare to that on the original one-receiver channel and with respect to the matched filter bound. (2 pts)

c. Find the probability of bit error on the diversity channel for the case of 1 bit/dimension transmission. (1 pt)

3.32 Do we understand basic detection - 9 pts

A filtered AWGN channel with \( T = 1 \) has pulse response \( p(t) = \text{sinc}(t) + \text{sinc}(t - 1) \) with SNR_{MFB} = 14.2 dB.

a. Find the probability of error for a ZF-DFE if a binary symbol is transmitted (2 pts).

b. The Tomlinson precoder in this special case of a monic pulse response with all unity coefficients can be replaced by a simpler precoder whose output is

\[
x'_k = \begin{cases} 
  x'_{k-1} & \text{if } x_k = 1 \\
  -x'_{k-1} & \text{if } x_k = -1
\end{cases}
\]  

Find the possible (noise-free) channel outputs and determine an SBS decoder rule. What is the performance (probability of error) for this decoder? (2 pts)

c. Suppose this channel (without precoder) is used one time for the transmission of one of the 8 4-dimensional messages that are defined by \( \pm \pm \pm \), which produce 8 possible 5-dimensional outputs. The decoder will use the assumption that the last symbol transmitted before these 4 was -1. ML detection is now used for this multidimensional 1-shot channel. What are the new minimum distance and number of nearest neighbors at this minimum distance, and how do they relate to those for the system in part a? What is the new number of bits per output dimension? (3 pts)

d. Does the probability of error change significantly if the input is extended to arbitrarily long block length \( N \) with now the first sample fixed at + and the last sample fixed at −, and all other inputs being equally likely + or − in between? What happens to the number of bits per output dimension in this case? (2 pts)

3.33 Equalization of a 3-tap channel - 10 pts

PAM transmission with \( \bar{E}_x = 1 \) is used on a filtered AWGN channel with \( \frac{N_0}{T} = .01, T = 1, \) and pulse response \( p(t) = \text{sinc}(t) + 1.8\text{sinc}(t - 1) + .81\text{sinc}(t - 2) \). The desired probability of symbol error is \( 10^{-6} \).

a. Find \( \| p \|^2, \) SNR_{MFB}, AND \( Q(D) \) for this channel. (2 pts)
b. Find the MMSE-LE and corresponding unbiased SNR for this channel. (2 pts)

c. Find the ZF-DFE detection SNR and loss with respect to SNR. (1 pt)

d. Find the MMSE-DFE. Also compute the MMSE-DFE SNR and maximum data rate in bits/dimension using the gap approximation. (3 pts)

e. Draw the flexible (Laroia) precoder for the MMSE-DFE and draw the system from transmit symbol to detector in the receiver using this precoder. Implement this precoder for the largest number of integer bits that can be transmitted according to your answer in Part d. (2 pts)

3.34 Finite-Length Equalizer Design - Lorentzian Pulse Shape- 12 pts

A filtered AWGN channel has the Lorentzian impulse response

\[ h(t) = \frac{1}{T} \cdot \frac{1}{1 + \left(\frac{10^7 t}{T}\right)^2} \]  

with Fourier Transform

\[ H(f) = \frac{3\pi}{10} e^{-6\pi \cdot 10^{-7} |f|} \].

This channel is used in the transmission system of Figure 3.89. QAM transmission with \(1/T = 1\) MHz and carrier frequency \(f_c = 600\)kHz are used on this channel. The AWGN psd is \(N_0/2 = -86.5\) dBm/Hz. The transmit power is \(E_x = 1\) mW. Square-root raised cosine filtering with 10% excess bandwidth is applied as a transmit filter, and an ideal filter is used for anti-alias filtering at the receiver. A matlab subroutine at the course web site may be useful in computing responses in the frequency domain.

a. Find \(\nu\) so that an FIR pulse response approximates this pulse response so that less than 5% error in \(\|p\|^2\). (3 pts)

---

455% error in approximation does not mean the SNR is limited to 1/20 or 13 dB. 5% error in modeling a type of pulse response for analysis simply means that the designer is estimating the performance of an eventual adaptive system that will adjust its equalizer parameters to the best MMSE settings for whatever channel. Thus, 5% modeling error is often sufficient for the analysis step to get basic estimates of the SNR performance and consequent achievable data rates.
b. Calculate SNR\(_{MFB}\). Determine a reasonable set of parameters and settings for an FIR MMSE-LE and the corresponding data rate at \(P_e = 10^{-7}\). Calculate the data rate using both an integer and a possibly non-integer number of bits/symbol. (4 pts)

c. Repeat Part b for an FIR MMSE-DFE design and draw receiver. Calculate the data rate using both an integer and a possibly non-integer number of bits/symbol. (3 pts)

d. Can you find a way to improve the data rate on this channel by changing the symbol rate and or carrier frequency? (2 pts)

3.35 Telephone-Line Transmission with “T1” - 14 pts

Digital transmission on telephone lines necessarily must pass through two “isolation transformers” as illustrated in Figure 3.90. These transformers prevent large D.C. voltages accidentally placed on the line from unintentionally harming the telephone line or the internet equipment attached to it, and they provide immunity to earth currents, noises, and ground loops. These transformers also introduce ISI.

![Figure 3.90: Illustration of a telephone-line data transmission](image)

a. The “derivative taking” combined characteristic of the transformers can be approximately modeled at a sampling rate of \(1/T = 1.544 \text{ MHz}\) as successive differences between channel input symbols. For sufficiently short transmission lines, the rest of the line can be modeled as distortionless. What is a reasonable partial-response model for the channel \(H(D)\)? Sketch the channel transfer function. Is there ISI? (3 pts)

b. How would a zero-forcing decision-feedback equalizer generally perform on this channel with respect to the case where channel output energy was the same but there are no transformers? (2 pts)

c. What are some of the drawbacks of a ZF-DFE on this channel? (2 pts)

d. Suppose a Tomlinson precoder were used on the channel with \(M = 2\), how much is transmit energy increased generally? Can you reduce this increase by good choice of initial condition for the Tomlinson Precoder? (3 pts)

e. Show how a binary precoder and corresponding decoder can significantly simplify the implementation of a detector. What is the loss with respect to optimum MFB performance on this channel with your precoder and detector? (3 pts)

f. Suppose the channel were not exactly a PR channel as in Part a, but were relatively close. Characterize the loss in performance that you would expect to see for your detector. (1 pt)

3.36 Magnetic Recording Channel - 14 pts

Digital magnetic information storage (i.e., disks) and retrieval makes use of the storage of magnetic fluxes on a magnetic disk. The disk spins under a “read head” and by Maxwell’s laws the read-head wire senses flux changes in the moving magnetic field. This read head thus generates a read current that is translated into a voltage through amplifiers succeeding the read head. Change in flux is often encoded to mean a “1” was stored and no change means a “0” was stored. The read head also has finite-band limitations.
a. Pick a partial-resonse channel with \( \nu = 2 \) that models the read-back channel. Sketch the magnitude characteristic (versus frequency) for your channel and justify its use. (2 pts)

b. How would a zero-forcing decision-feedback equalizer generally perform on this channel with respect to the best case where all read-head channel output energy conveyed either a positive or negative polarity for each pulse? (1 pt)

c. What are some of the drawbacks of a ZF-DFE on this channel? (2 pts)

d. Suppose a Tomlinson Precoder were used on the channel (ignore the practical fact that magnetic saturation might not allow a Tomlinson nor any type of precoder) with \( M = 2 \), how much is transmit energy increased generally? Can you reduce this increase by good choice of initial condition for the Tomlinson Precoder? (2 pts)

e. Show how a binary precoder and corresponding decoder can significantly simplify the implementation of a detector. What is the loss with respect to optimum MFB performance on this channel with your precoder and detector? (3 pts)

f. Suppose the channel were not exactly a PR channel as in part a, but were relatively close. Characterize the loss in performance that you would expect to see for your detector. (1 pt)

g. Suppose the density (bits per linear inch) of a disk is to be increased so one can store more files on it. What new partial response might apply with the same read-channel electronics, but with a correspondingly faster symbol rate? (3 pts)

3.37 Tomlinson Precoding and Simple Precoding - 8 pts

Section 3.8 derives a simple precoder for \( H(D) = 1 + D \).

a. Design the Tomlinson precoder corresponding to a ZF-DFE for this channel with the possible binary inputs to the precoder being \( \pm 1 \). (4 pts)

b. How many distinct outputs are produced by the Tomlinson Precoder assuming an initial state (feedback \( D \) element contents) of zero for the precoder. (1 pt)

c. Compute the average energy of the Tomlinson precoder output. (1 pt)

d. How many possible outputs are produced by the simple precoder with binary inputs? (1 pt)

e. Compute the average energy of the channel input for the simple precoder with the input constellation of Part a. (1 pt)

3.38 Flexible Precoding and Simple Precoding - 8 pts

Section 3.8 derives a simple precoder for \( H(D) = 1 + D \).

a. Design the Flexible precoder corresponding to a ZF-DFE for this channel with the possible binary inputs to the precoder being \( \pm 1 \). (4 pts)

b. How many distinct outputs are produced by the Flexible Precoder assuming an initial state (feedback \( D \) element contents) of zero for the precoder. (1 pt)

c. Compute the average energy of the Flexible precoder output. (1 pt)

d. How many possible outputs are produced by the simple precoder with binary inputs? (1 pt)

e. Compute the average energy of the channel input for the simple precoder with the input constellation of Part a. (1 pt)

3.39 Partial Response Precoding and the ZF-DFE - 10 pts

An AWGN has response \( H(D) = (1 - D)^2 \) with noise variance \( \sigma^2 \) and one-dimensional real input \( x_k = \pm 1 \).
a. Determine a partial-response (PR) precoder for the channel, as well as the decoding rule for the noiseless channel output. (2 pts)

b. What are the possible noiseless outputs and their probabilities? From these, determine the $P_e$ for the precoded channel. (4 pts)

c. If the partial-response precoder is used with symbol-by-symbol detection, what is the loss with respect to the MFB? Ignore nearest neighbor terms for this calculation since the MFB concerns only the argument of the Q-function. (1 pt)

d. If a ZF-DFE is used instead of a precoder for this channel, so that $P_c(D) = 1 - 2D + D^2$, what is $\eta_0$? Determine also the SNR loss with respect to the SNR_{ZF,FB} (2 pts)

e. Compare this with the performance of the precoder, ignoring nearest neighbor calculations. (1 pt)

3.40 Error propagation and nearest neighbors - 10 pts

A partial-response channel has $H(D) = 1 - D^2$ channel with AWGN noise variance $\sigma^2$ and $d = 2$ and 4-level PAM transmission.

a. State the precoding rule and the noiseless decoding rule. (1 pts)

b. Find the possible noiseless outputs and their probabilities. Find also $N_e$ and $P_e$ with the use of precoding. (4 pts)

c. Suppose a ZF-DFE is used on this system and that at time $k = 0$ an incorrect decision $x_0 - \hat{x}_0 = 2$ occurs. This incorrect decision affects $z_k$ at time $k = 2$. Find the $N_e$ (taking the error at $k = 0$ into account) for the ZF-DFE. From this, determine the $P_e$ with the effect of error propagation included. (4 pts)

d. Compare the $P_e$ in part (c) with that of the use of the precoder in Part a. (1 pt)

3.41 Forcing Partial Response - 6 pts

Consider a $H(D) = 1 + .9D$ channel with AWGN noise variance $\sigma^2$. The objective is to convert this to a $1 + D$ channel

a. Design an equalizer that will convert the channel to a $1 + D$ channel. (2 pts)

b. The received signal is $y_k = x_k + .9 \cdot x_{k-1} + n_k$ where $n_k$ is the AWGN. Find the autocorrelation of the noise after going through the receiver designed in Part a. Evaluate $r_0$, $r_{\pm 1}$, and $r_{\pm 2}$. Is the noise white? (3 pts)

c. Would the noise terms would be more or less correlated if the conversion were instead of a $1 + .1D$ channel to a $1 + D$ channel? You need only discuss briefly. (1 pt)

3.42 Equalization of a General Single-Pole Channel - 11 pts

PAM transmission on a filtered AWGN channel uses basis function $\varphi(t) = \frac{1}{\sqrt{T}} \cdot \text{sinc} \left( \frac{t}{T} \right)$ with $T = 1$ and undergoes channel impulse response with Fourier transform ($|\alpha| < 1$)

$$H(\omega) = \begin{cases} \frac{1}{1 + \alpha e^{-j\omega T}} & |\omega| \leq \pi \\ 0 & |\omega| > \pi \end{cases}$$ (3.767)

and SNR = $\frac{E_x}{\sigma^2}$ = 28 dB.

a. Find the Fourier Transform of the pulse response, $P(\omega) = ?$ (1 pt)

b. Find $\|p\|^2$. (1 pt)

c. Find $Q(D)$, the function characterizing ISI. (2 pts)
d. Find the filters and sketch the block diagram of receiver for the MMSE-DFE on this channel for $a = .9$. (3 pts)

e. Estimate the data rate for uncoded PAM transmission and $P_e < 10^{-6}$ that is achievable with your answer in part d. (2 pts)

f. Draw a diagram of the better precoder’s (Tomlinson or Laroia), transmitter and receiver, implementations with $d = 2$ in the transmitted constellation. (2 pts)

3.43 Finite Equalization of a “causal” complex channel – 15 pts

A channel sampled at symbol rate $T = 1$ (after anti-alias filtering) has output samples given by

$$y_k = p_k * x_k + n_k,$$

where $p_k = \delta_k + (-2 + 0.25j)\delta_{k+1}$, $E[n_k \cdot n_{k-l}] = \frac{\delta_l}{50}$ and $ex = 2$.

a. Compute SNR$_{MFB}$. (1 pt)

b. For a channel equalizer with 3 feedforward taps and no feedback, write the $P$ convolution matrix. Is this channel causal? If not, is it somehow equivalent to causal for the use of equalization programs like DFE color? (2 pts)

c. For the re-indexing of time to correspond to your answer in Part b, find the best delay for an MMSE-LE when $N_f = 3$. (1 pt)

d. Find $W_{MMSE-LE}$ for $N_{ff} = 3$ for the delay in Part c. (1 pt)

e. Convolve the equalizer for Part d with the channel and show the result. (1 pt).

f. Compute SNR$_{MMSE-LE,U}$ and the corresponding MMSE. (both per-dimension and per symbol). (2 pts).

g. Compute the loss w.r.t. MFB for the 3-tap linear equalizer. (1 pt).

h. Find a better linear equalizer using up to 10 taps and corresponding SNR$_{MMSE-LE,U}$ improvement. (2 pts)

i. For your answer in Part h, what is $P_e$ for 16 QAM transmission? (1 pt)

j. Design an MMSE-DFE with no more than 3 total taps that performs better than anything above. (3 pts)

3.44 Diversity Concept and COFDM - 7 pts

An additive white Gaussian noise channel supports QAM transmission with a symbol rate of $1/T = 100$ kHz (with 0 % excess bandwidth) anywhere in a total baseband-equivalent bandwidth of 10 MHz transmission. Each 100 kHz wide QAM signal can be nominally received with an SNR of 14.5 dB without equalization. However, this 10 MHz wide channel has a 100 kHz wide band that is attenuated by 20 dB with respect to the rest of the band, but the location of the frequency of this notch is not known in advance to the designer. The transmitters are collocated.

a. What is the data rate sustainable at probability of bit error $P_b \leq 10^{-7}$ in the nominal condition? (1 pt)

b. What is the maximum number of simultaneous QAM users can share this channel at the performance level of Part a if the notch does not exist? (1 pt)

c. What is the worst-case probability of error for the number of users in Part b if the notch does exist? (1 pt)
d. Suppose now (unlike Part b) that the notch does exist, and the designer decides to send each QAM signal in two distinct frequency bands. (This is a simple example of what is often called Coded Orthogonal Frequency Division Modulation or COFDM.) What is the worst-case probability of error for a diversity-equalization system applied to this channel? (1 pt)

e. For the same system as Part c, what is the best SNR that a diversity equalizer system can achieve? (1 pt)

f. For the system of Part c, how many users can now share the band all with probability of bit error less than $10^{-7}$? Can you think of a way to improve this number closer to the level of the Part b? (2 pts)

**3.45 Hybrid ARQ (HARQ) Diversity - 12 pts**

A time-varying wireless channel sends long packets of binary information that are decoded by a receiver that uses a symbol-by-symbol decision device. The channel has pulse response in the familiar form:

$$P(\omega) = \begin{cases} \sqrt{T} \cdot (1 + a_i \cdot e^{-j\omega T}) & |\omega| \leq \frac{\pi}{T} \\ 0 & |\omega| > \frac{\pi}{T} \end{cases}$$

(3.768)

where $a_i$ may vary from packet to packet (but not within a packet). There is also AWGN with constant power spectral density $N_0 = 1$. QPSK (4QAM) is sent on this channel with symbol energy $2$.

a. Find the best DFE SNR and corresponding probability of symbol error for the two cases or “states” when $a_1 = .9$ and $a_2 = .5$. Which is better and why? (2 pts)

Now, suppose the transmitter can resend the same packet of symbols to the receiver, which can delay the channel output packet from the first transmission until the symbols from the new packet arrive. Further suppose that the second transmission occurs at a time when the channel state is known to have changed. However, a symbol-by-symbol detector is still to be used by the receiver jointly on both outputs. This is a variant of what is called “Automatic Repeat reQuest” or ARQ. (Usually ARQ only resends when it is determined by the receiver that the decision made on the first transmission is unreliable and discarded. Hybrid ARQ uses both channel outputs.) For the remainder of this problem, assume the two $a_i$ values are equally likely.

b. Explain why this is a diversity situation. What is the approximate cost in data rate for this retransmission in time if the same symbol constellation (same $b$) is used for each transmission? (1 pt)

c. Find a new SNR$_{MFB}$ for this situation with a diversity receiver. (1 pt)

d. Find the new function $Q(D)$ for this diversity situation. (2 pts)

e. Determine the performance of the best DFE this time (SNR and $P_e$) and compare to the earlier single-instance receivers. (2 pts)

f. Compare the $P_e$ of Part e with the product of the two $P_e$’s found in Part a. Does this make sense? Comment. (2 pts)

Use the receiver with RAKE illustrated as well as DFE, unbiasing, and decision device. (phase splitter can be ignored in diagram and all quantities can be assumed complex.) (2 pts).

**3.46 Precoder Diversity - 7 pts**

A system with 4 PAM transmits over two discrete-time channels shown with the two independent AWGN channels shown in Figure 3.91. ($\bar{E}_x = 1$)
Figure 3.91: Transmission system for Problem ??.

a. Find the RAKE matched filter(s) for this system? (1 pt)
b. Find the single feedforward filter for a ZF-DFE? (1 pt)
c. Show the best precoder for the system created in Part b. (1 pt)
d. Find the SNR at the detector for this precoded system? (1 pt)
e. Find the loss in SNR with respect to the of either single channel. (1 pt)
f. Find the $P_e$ for this precoded diversity system. (1 pt)
g. Find the probability that a packet of 1540 bytes contains one or more errors. What might you do to improve? (1 pt)

3.47 Equalizer Performance and Means - 10 pts

Recall that arithmetic mean, geometric mean, and harmonic mean are of the form \( \frac{1}{n} \sum_{i=1}^{n} x_i \), \( \left( \prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \), \( \left( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1} \), respectively. Furthermore, they satisfy the following inequalities:

\[
\frac{1}{n} \sum_{i=1}^{n} x_i \geq \left( \prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \geq \left( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1},
\]

with equality when \( x_1 = x_2 = \cdots = x_n \)

a. Express $SNR_{ZFE}$, $SNR_{ZF-DFE}$, $SNR_{MFB}$ in terms of $SNR_{MFB}$ and frequency response of the autocorrelation function $g_k$. Prove that $SNR_{ZFE} \leq SNR_{ZF-DFE} \leq SNR_{MFB}$ using above inequalities. When does equality hold? hint: use $\int_{a}^{b} x(t)dt = \lim_{n \to \infty} \sum_{k=1}^{n} x \left( \frac{k-a}{n} + a \right)$.

(4 pts)

b. Similarly, prove that $SNR_{MMSE-LE,U} \leq SNR_{MMSE-DFE,U} \leq SNR_{MFB}$ using above inequalities. When does equality hold? (4 pts)

c. Compare $SNR_{ZFE}$ and $SNR_{MMSE-LE}$. Which scheme has better performance? (2 pts)
Use the gap approximation for this problem. A sequence of 16-QAM symbols with in-phase component $a_k$ and quadrature component $b_k$ at time $k$ is transmitted on a passband channel by the modulated signal

$$x(t) = \sqrt{2} \left\{ \left[ \sum_k a_k \cdot \varphi(t - kT) \right] \cdot \cos(\omega_c t) - \left[ \sum_k b_k \cdot \varphi(t - kT) \right] \cdot \sin(\omega_c t) \right\},$$

(3.769)

where $T = 40$ ns, $f_c = 2.4$ GHz, and the transmit power is 700 $\mu$W.

The transmit filter/basic function is $\varphi(t)$

$$\varphi(t) = \frac{1}{\sqrt{T}} \text{sinc} \left( \frac{t}{T} \right),$$

(3.770)

and the baseband channel response is a frequency translation given by $H(f) = e^{-j5\pi fT}$. The received signal is thus

$$y(t) = h(t) * x(t) + n(t)$$

(3.771)

where $n(t)$ is an additive white Gaussian noise random process with two-sided PSD of -100 dBm/Hz.

a. What is the data rate of this system? (1 pts)
b. What are $E_x$, $\bar{E}_x$ and $d_{\text{min}}$ for this constellation? (2 pts)
c. What is $\tilde{x}_{bb}(t)$? (1.5 pts)
d. What is the ISI characterizing function $q(t)$? what is $q_k$? (1.5 pts)
e. What are $P_e$ and $\bar{P}_e$ for an optimal ML detector? (2 pts)

For the remainder of this problem: The implementation of the baseband demodulator is faulty and the additive Gaussian noise power in the quadrature component is $\alpha$ times what it should be. That is, the noise variance of the quadrature component is $\alpha \frac{N_0}{2}$ while the noise variance in the in-phase dimension is $\frac{N_0}{2}$. Answer the following questions for this situation.

f. What is the receiver SNR? Is this SNR meaningful any longer in terms of the gap formula? (2.5 pts)
g. What are $P_e$ and $\bar{P}_e$? (3 pts)
h. Instead of using QAM modulation as above, two independent 4-PAM constellations (with half the total energy allocated for each constellation) are now transmitted on the in-phase and quadrature channels, respectively. What is $P_e$ averaged over both of the PAM constellations? (1.5 pts)

i. The system design may allocate different energies for the in-phase and quadrature channels (but this design still uses QAM modulation with the same number of bits on each of the in-phase and quadrature channels, and the transmit power remains at 700 $\mu$W). Use different inphase and quadrature energies to improve the $\bar{P}_e$ performance from Part (g). What is the best energy allocation for the in-phase and quadrature channels? For the given $P_e = 10^{-6}$, what is $\bar{b}$ and the achievable data rate? (3 pts)

j. Using the results from part (i) find the minimal increase in transmit power needed to guarantee $P_e = 10^{-6}$ for 16-QAM transmission? (2.5 pts)

The design may now reduce losses. Given this increased noise in the quadrature dimension, the design may use only the in-phase dimension for transmission (but subject to the original transmit power constraint of 700 $\mu$W).
k. What data rate can be then supported that gives $P_e = 10^{-6}$? (2.5 pts)

l. Using a fair comparison, compare the QAM transmission system from part (i) to the single-dimensional scheme ($N = 1$) from part (k). Which scheme is better? (3 pts)

m. Design a new transmission scheme that uses both the in-phase and quadrature channels to get a higher data rate than part (i), subject to the same $P_e = 10^{-6}$, and original symbol rate and power constraint of 700 $\mu W$. (3 pts)

3.49 A Specific One-Pole Channel (9 pts)
A transmission system uses the basis function $\phi(t) = \text{sinc}(t)$ with $\frac{1}{T} = 1$ Hz. The Fourier transform of the channel impulse response is:

$$H(f) = \begin{cases} 
\frac{1}{1-0.5j e^{-j\pi T}}, & \text{for } |f| \leq \frac{1}{2} \\
0, & \text{for } |f| > \frac{1}{2}.
\end{cases}$$

and $\text{SNR} = \frac{\bar{\epsilon}_x}{\sigma^2} = 10$ dB.

a. Find $||p||^2$ and $Q(D)$. (1 pt)
b. Find $W_{ZFE}(D)$ and $W_{MMSE-LE}(D)$. (1 pt)
c. Find $W(D)$ and $B(D)$ for ZF-DFE. (1 pt)
d. Find $W(D)$ and $B(D)$ for MMSE-DFE. (2 pts)
e. What is $\text{SNR}_{ZF-DFE}$? (0.5 pt)
f. Design a Tomlinson precoder based on the ZF-DFE. Show both the precoder and the corresponding receiver (for any $M$). (1.5 pts)
g. Design a Laroia precoder based on the ZF-DFE. Assume the system uses 4-QAM modulation. Show both the precoder and the corresponding receiver. The kind of constellation used at each SBS detector should be shown. (2 pts)

3.50 Infinite Diversity - Malkin (7 pts)
A receiver has access to an infinite number of diversity channels where the output of channel $i$ at time $k$ is

$$y_{k,i} = x_k + 0.9 \cdot x_{k+1} + n_{k,i}, \quad i = 0, 1, 2, \ldots$$

where $n_{k,i}$ is a Gaussian noise process independent across time and across all the channels, and with variance $\sigma_i^2 = 1.2^i \cdot 0.181$. Also, $\bar{\epsilon}_x = 1$.

a. Find $\text{SNR}_{MFB,i}$, the $\text{SNR}_{MFB}$ for channel $i$. (1 pt)
b. What is the matched filter for each diversity channel? (2 pts)
c. What is the resulting $Q(D)$ for this diversity system? (1.5 pts)
d. What is the detector SNR for this system for a MMSE-DFE receiver? (2.5 pts)

3.51 Infinite Feedforward and Finite Feedback - Malkin (9 pts)
A symbol-rate sampled MMSE-DFE structure, where the received signal is filtered by a continuous time matched filter, uses an infinite-length feedforward filter. However, the MMSE-DFE feedback filter has finite length $N_b$. 

a. Formulate the optimization problem that solves to find the optimal feedforward and feedback coefficients to minimize the mean squared error (Hint: Don’t approach this as a finite-length equalization problem as in Section 3.7).

Assume that $\hat{N}_b = .181$, $E_x = 1$, and $P(D) = 1 + .9 \cdot D^{-1}$. (and $\text{SNR}_{MFB} = 10$ dB). (2 pts)

b. Using only 1 feedback tap ($N_b = 1$), determine the optimal feedforward and feedback filters (Hint: This is easy - you have already seen the solution before). (2 pts)

c. Now consider the ZF-DFE version of this problem. State the optimization problem for this situation (analogous to Part a). (2 pts)

Now, change the channel to the infinite-impulse response of $P(D) = 1 + 0.9D^{-1}$ (and $\|p\|^2 = 1 - 0.9^2 = 0.19$).

d. Determine the optimal feedforward and feedback filters for the ZF-DFE formulation using only 1 feedback tap ($N_b = 1$). (3 pts)

3.52 Packet Processing - Malkin (10 pts)

Suppose that in a packet-based transmission system, a receiver makes a decision from the set of transmit symbols $x_1, x_2, x_3$ based on the channel outputs $y_1, y_2, y_3$, where

$$
\begin{bmatrix}
  y_3 \\
  y_2 \\
  y_1
\end{bmatrix} =
\begin{bmatrix}
  p_{11} & p_{12} & p_{13} \\
  0 & p_{22} & p_{23} \\
  0 & 0 & p_{33}
\end{bmatrix}
\begin{bmatrix}
  x_3 \\
  x_2 \\
  x_1
\end{bmatrix} +
\begin{bmatrix}
  n_3 \\
  n_2 \\
  n_1
\end{bmatrix} =
\begin{bmatrix}
  0.6 & 1.9 & -3.86 \\
  0 & 1.8 & 3.3 \\
  0 & 0 & 1.2
\end{bmatrix}
\begin{bmatrix}
  x_3 \\
  x_2 \\
  x_1
\end{bmatrix} +
\begin{bmatrix}
  n_3 \\
  n_2 \\
  n_1
\end{bmatrix}
$$

(3.773)

and where the noise autocorrelation matrix is given by

$$
R_n = E\left(\begin{bmatrix}
  n_3 \\
  n_2 \\
  n_1
\end{bmatrix}\begin{bmatrix}
  n_3^* \\
  n_2^* \\
  n_1^*
\end{bmatrix}\right) =
\begin{bmatrix}
  0.3 & 0 & 0 \\
  0 & 0.5 & 0 \\
  0 & 0 & 0.1
\end{bmatrix}.
$$

(3.774)

The inverse of the channel matrix is given by

$$
\begin{bmatrix}
  0.6 & 1.9 & -3.86 \\
  0 & 1.8 & 3.3 \\
  0 & 0 & 1.2
\end{bmatrix}^{-1} =
\begin{bmatrix}
  1.67 & -1.76 & 10.19 \\
  0 & 0.56 & -1.52 \\
  0 & 0 & 0.83
\end{bmatrix}.
$$

(3.775)

The transmit symbols are i.i.d with $\mathcal{E}_{x_1} = \mathcal{E}_{x_2} = \mathcal{E}_{x_3} = 20$.

a. What is the 3-tap zero-forcing equalizer for $x_1$ based on the observation $y_1, y_2, y_3$? Repeat for $x_2$ and $x_3$. (1.5 pts)

b. What is the matrix ZFE for detecting $x_1, x_2,$ and $x_3$? (Hint: This is easy - all the work occurred in the previous part) (1 pt)

c. What is the detection SNR for $x_1$ using a ZFE? Answer the same for $x_2$ and $x_3$. (1.5 pts)

d. An engineer now realizes that the performance could be improved through use previous decisions when detecting the current symbol. If the receiver starts by detecting $x_1$, then $x_2$, and finally $x_3$, describe how this modified ZF detector would work? (1 pt)

e. Assuming previous decisions are correct, what is now the detection SNR for each symbol? (1.5 pts)

f. The approach above does not necessarily minimize the MMSE. How would a packet MMSE-DFE work? How would you describe the feedback structure? (Don’t solve for the optimal matrices - just describe the signal processing that would take place). (1 pt)
g. Can the THP also be implemented for this packet based system? Describe how this would be done (a description is sufficient response). (1 pt)

3.53  **DFEcolor Program with Nontrivial Complex Channel – 9 pts**

Use of the dfecolor.m program in Matlab produces:

```matlab
>> p = [0 - 0.2500i 0.5000 + 0.5000i 1.0000 + 0.1250i 0.3000 - 0.2500i]
>> [snr,w]=dfecolor(1,p,7,3,6,1,.03*[1 zeros(1,6)])
```

snr = 15.2754
w = Columns 1 through 5
0.0006 + 0.0019i -0.0064 + 0.0002i 0.0103 + 0.0198i 0.0249 + 0.0341i 0.2733 - 0.0231i
Columns 6 through 10
0.6038 - 0.6340i -0.0000 + 0.2402i -0.6392 + 0.5137i 0.0074 + 0.1009i -0.0601 - 0.0721i

a. What is equalizer output (before bias removal), $E[z_k/x_k-\Delta]$, for the equalizer as output above? (1 pt)

b. If the noise vector in the matlab command above becomes .02*[1.5 -.5+.5*i], what will happen to DFE performance (improve/degrade)? (1 pt)

c. Find $B_U(D)$ for the original (white) noise. (1 pt)

d. What is the maximum data rate if the symbol rate is 100 MHz and $P_e=10^{-6}$? (Assume no error propagation and arbitrary constellations with SBS decisions in the MMSE-DFE.) (1 pt)

e. What is the maximum data rate that can be achieved if a Tomlinson precoder is used? (1 pt)

f. If the carrier frequency is 75 MHz, what are possible pulse-response samples for the actual channel? (1 pt)

g. Assuming the channel has the minimum bandwidth for the pulse response given, what is the loss with respect to the matched filter bound? (1 pt)

h. When bias is absorbed into the feedforward equalizer, what is the magnitude of the “center” tap? (1 pt)

i. For a flexible precoder in the transmitter, what is the maximum number of points in the first two-dimensional decision device used in the receiver? (1 pt)

3.54  **Finite-Length Equalizer Design with Diversity – 17 pts**

A receiver designer has two options for the two-path channel described below. The designer may use only a single antenna (Parts a to c) that receives both paths’ outputs added together with a single AWGN power spectral density of -110 dBm/Hz. Or as this problem progresses in Parts d and beyond, the designer may instead use two antennas that are directional with one receiver only on the first path with AWGN at -113 dBm/Hz and the second receiver only only the second path with AWGN also at -113 dBm/Hz (the noise that was present on the channel is presumably split evenly in this case). The transmit power is $P_x = \frac{E_x}{T} \leq 20$ dBm. The symbol rate is 1 MHz. The target probability of error is $10^{-6}$ and only integer number of bits/symbol are allowed. The system is sampled at instants $kT/2 = kT'$ for integer $k \geq 0$

The two paths’ sampled (at pulse responses are:

$$p_1(t') = e^{-10^6 t'} \cdot u(t)$$

$$p_2(t) = e^{-0.5 \cdot 10^6 \cdot (t-0.5\mu s)} \cdot u(t-0.5\mu s)$$

where $\mu s=10^{-6}$ s.
where \( N \) is the symbol sequence before transmission, though the filtered symbol sequence must have the same power as the transmit sequence, \( E \). In practice, an adaptive implementation would tune the equalizer to get roughly \( 3.55 \)dB error. Call the magnitude squared of the transmit filter \( \parallel p \parallel^2 \) occurs, and provide the larger value of \( \nu \). (2 pts)

Now find the Ex input for the dfecolor program, the vector \( p \) input that would be input to the DFECOLOR matlab program, and the CORRESPONDING noise vector that would be input. Please find these presuming that the anti-alias filter for \( 1/T \) sampling is normalized. (3 pts)

c. Design an FIR DFE that achieves within .1 dB of the infinite-length performance for Part b’s inputs. Record the smallest numbers of feedforward and feedback taps for which the equalizer can achieve this performance and the corresponding delay \( \Delta \). An acceptable response is the equalizer coefficient vector from Matlab decomposed separately into feedforward and feedback taps. Find the SNR and corresponding data rate for the design (3 pts).

d. Continue Part b except now find the two FIR pulse responses that approximates the pulse response so that less than .25 dB error in the overall \( \parallel p \parallel^2 \) occurs, and provide the larger value of \( \nu \). (2 pts)

e. Now find the Ex input for the dfecolor program, the vector (rows) of \( p \) input that would be input to the DFERAKE matlab program, and the CORRESPONDING noise vector that would be input. Please find these presuming that the anti-alias filter for \( 1/T' \) sampling is normalized on both antennas. (2 pts)

f. Design an FIR DFE that achieves within .1 dB of the infinite-length performance for Part d’s pulse responses. Record the smallest numbers of feedforward and feedback taps that can achieve this performance and the corresponding delay \( \Delta \). An acceptable response is the equalizer coefficient vector from Matlab. Find the SNR and corresponding data rate for your design. (3 pts).

3.55 Optimizing the DFE Transmit Filter – 3 pts

A discrete-time channel is described by \( y(D) = H(D) \cdot X(D) + N(D) \), where \( X(D) \) is the transmit symbol sequence, \( E \) and \( N(D) \) is AWGN with variance \( \frac{A_0}{2} \). These notes so far have always assumed that the transmit sequence \( x_k \) s white (messages are independent). Suppose the transmitter could filter the symbol sequence before transmission, though the filtered symbol sequence must have the same power \( E_2 \). Call the magnitude squared of the transmit filter \( \parallel PH(e^{-j\omega T})\parallel^2 \).

a. Given a MMSE-DFE receiver, find the transmit filter (squared magnitude) that maximizes the MMSE-DFE detection SNR. (2 pts)

b. Find the best transmit filter (squared magnitude) for the ZF-DFE (1 pt).

3.56 Multi-User Scenario – 22 pts

A single receiver receives transmissions from two independent users with symbol sequences \( X_1(D) \) and \( X_2(D) \), respectively. The two received signals are (with \( a \) and \( b \) real).

\[
Y_1(D) = (1 + a \cdot D)) \cdot X_1(D) - X_2(D) + N_1(D) \\
Y_2(D) = X_1(D) + (1 - b \cdot D^2) \cdot X_2(D) + (1 - b \cdot D^2) \cdot N_2(D) + b \cdot D^2 \cdot Y_2(D),
\]

where \( N_1(D) \) is AWGN with variance \( \frac{A_0}{2} \), and similarly \( N_2(D) \) is AWGN with variance \( \frac{A_0}{2} \). The quantites \( a \) and \( b \) are real. The following are the energies

\[
E \left[ X_1(D)X_1(D^{-*}) \right] = E_1 \\
E \left[ X_2(D)X_2(D^{-*}) \right] = E_2.
\]

Readers should note that .25 dB accuracy is only about 6% of the energy; however, this is close enough to estimate equalizer behavior of the actual system. In practice, an adaptive implementation would tune the equalizer to get roughly the same performance as is derived here for the truncated approximation to the response \( p(kT/2) \) over all time.

350
The noises $N_1(D)$ and $N_2(D)$ are independent of one another and all mutually independent of the two transmitted data sequences $X_1(D)$ and $X_2(D)$. For this problem treat all ISI and “crosstalk” interference as if it came from a normal distribution, unless specifically treated otherwise as in Parts f and beyond. Matlab may be useful in this problem, particularly the routines “conv” “roots,” and “integral.”

The receiver will use both $Y_1(D)$ and $Y_2(D)$ for detection in all parts.

a. The receiver here will use both $Y_1(D)$ and $Y_2(D)$ for optimum detection of user 1 ($X_1(D)$), but treat $X_2(D)$ as noise. Find the noise-equivalent channel $\mathbf{P}(D)$ and the resulting $Q(D)$ for the case when $\mathcal{E}_2 = \frac{N_0}{2}$. The Matlab Function “chol” may be useful in finding a noise-whitening filter. (3 pts)

b. For $\mathcal{E}_1 = 10$, $\mathcal{E}_2 = 1$, $\frac{N_0}{2} = 1$, $a = 1$, and $b = 0.7$, find $Q(D)$, $\text{SNR}_{MFB}, \tilde{Q}(D)$, and the feed-forward and feedback filters for an infinite length MMSE-DFE for Part a’s. The “integral function” may be convenient to compute the norm of channel elements. The “roots” and “conv” commands may also reduce labor to solve. (5 pts)

c. Show a Tomlinson precoder implementation of the structure in part b if M-PAM is transmitted. (1 pt)

d. Suppose, instead, the receiver instead first detects user 2’s signal $X_2(D)$ based on both $Y_1(D)$ and $y_2(D)$ by using a ZF-DFE receiver. Please specify all the filters and the $\text{SNR}_{ZF-DFE}$, given that $\mathcal{E}_1 = 1$, $\mathcal{E}_2 = 30$, $\frac{N_0}{2} = 1$, $a = 1$, and $b = 7$. Noise whitening may need to presume a triangular form as in Part a except that now the elements are functions of $D$ - just solve it, 3 equations in 3 unknown functions of D - not that hard for first two elements, 3rd is ratio of 3rd order polynomials. The integral and conv Matlab functions are again useful. (4 pts)

e. Show a Laroia implementation of the structure in Part d if M-PAM is transmitted. (2 pts)

f. This sub question investigates the system in Part b if $x_{2,k}$ is known and its effects removed at the receiver? Suppose a structure like that found in Parts b and c to detect $x_{2,k}$ reliably. (Since $\mathcal{E}_1 = 10$ is large, the data rate for the channel from $x_{2,k}$ to the receiver may be very low relative to the correct response for Part b.) Show a receiver for $x_{1,k}$ that makes use of a known $x_{2,k}$. Does this receiver look familiar in some way? What is roughly the new gap-based data rate achievable for $x_1$? (3 pts)

g. Repeat Part f for the numbers in Part d except that $x_{1,k}$ is detected and its affects removed. (2 pts)

h. Suppose time-sharing of the two systems ($\mathcal{E}_1 = 10$ or $\mathcal{E}_2 = 30$) is allowed (long blocks are assumed so any beginning or end effects are negligible). What data rate pairs for $x_1$ and $x_2$ might be possible (think of plot in two dimensions)? (2 pts)

3.57 Mutual Information and Parallel Independent AWGN Channels (3 pts)

This problem considers a set of parallel, independent, and one-real-dimensional AWGN channels of the form:

$$y_k = h_k \cdot x_k + n_k,$$

where the noises all have variance $\sigma^2$ with zero mean.

a. (1 pt) Use probability densities factoring to show that this channel set’s mutual information is the sum of the set’s individual mutual information quantities.

b. (1 pt) If the set of parallel channels has a total energy constraint that is equal to the sum of the energy constraints, what energy $\mathcal{E}_k$, $k = 1, ..., N$ should be allocated to each of the channels to maximize the mutual information. The answer may use the definition that the subchannel gains are given as $g_k = \frac{|h_k|^2}{2\sigma^2}$ (so that the individual SNRs would then be $\text{SNR}_k = \mathcal{E}_k \cdot g_k$).

c. (1 pt) Find the overall $\text{SNR}_{overall}$ for a single AWGN that is equivalent to this problem’s set of parallel channels in terms of mutual information.
3.58 Innovations (6 pts)
Find the innovations variance per real dimension, entropy per dimension, and linear prediction filter for the following Gaussian processes (let $T = 1$):

a. (2 pts) A real process with autocorrelation $R_{xx}(D) = 0.0619 \cdot D^{-2} + 0.4691 \cdot D^{-1} + 1 + 0.4691 \cdot D + 0.0619 \cdot D^2$.

b. (2 pts) A complex discrete-time process with power spectral density $10 \cdot [1.65 + 1.6 \cdot \cos(\omega)]$.

c. (2 pts) A complex process that has a maximum of only 3 nonzero terms in its autocorrelation function and a notch exactly in the middle of its band (i.e., at normalized frequency $1/4$) and total power 2.

3.59 Multibands (14 pts)
Three bands of transmission result from infinite-length water-filling as shown below with a gap of 5.8 dB at $P_e = 10^{-6}$. The lowest band uses baseband PAM transmission, and the other two bands use QAM transmission. Each band has an MMSE-DFE receiver with the SNR shown in Figure 3.92.

![Figure 3.92: Multiband transmission for Problem 3.59](image)

a. (3 pts) Find the optimum symbol rates $1/T^*_i$, $1/\bar{T}^*_i$, and the optimum carrier frequencies $f^*_c,i$ for each of the 3 bands shown in Figure 3.92.

b. (3 pts) Find $\bar{b}_1$, $\bar{b}_2$, and $\bar{b}_3$, as well as $b_1$, $b_2$, and $b_3$ for each of these 3 bands. Also find the data rate $\tilde{R}$ for the entire 3-band system.

c. (1 pt) Find $1/T^*$.

d. (2 pts) Find the overall best SNR and $\bar{b}$ for a single AWGN that is equivalent to the set of channels in terms of mutual information.

e. (2 pts) The noise is such that $\tilde{E}_x = -70$ dBm/Hz in all used bands. Find the total energy per symbol period $T^*$ that is used and the power used. The answers can be specified in terms of dBm/Hz and dBm.

Suppose baseband PAM with a symbol rate of $1/T = 200$ MHz is used instead with the same power, and energy equally divided on all dimensions (i.e., successive PAM symbols are independent).

f. (1 pts) What is $\tilde{E}_{x,PAM} =$?

g. (2 pts) Find approximate SNR$_{mmse-\text{df},PAM,u}$ and new data rate for this alternative single-band PAM design.

3.60 CDEF (10 pts)
An AWGN with intersymbol interference sustains successful transmission using a MMSE-DFE receiver with 256 QAM with Gap of $\Gamma = 8.8$ dB with a margin of 1.2 dB at $P_e = 10^{-6}$.
a. (2 pts) What is the mutual information in bits per symbol for this channel? In bits per dimension?

b. (2 pts) Suppose optimization of the transmit filter increased the margin to 2.5 dB. What is the capacity (bits/dimension) of this new transmit-optimized channel at this symbol rate?

Returning (for the remainder of this problem) to the well-known $1 + .9 D^{-1}$ channel with $\text{SNR}_{MFB} = 10$ dB with $T = 1$ and PAM transmission:

c. (2 pts) What is the capacity of this system if energy increases such that $T^* = 1$ becomes the optimum symbol rate?

d. (3 pts) What would be the new (smallest) symbol energy per dimension for the situation in Part c?

e. (1 pt) Suppose this optimized system transmits with code with gap of 6 dB for probability of error $10^{-6}$ at $\bar{b} = 1$. What is margin at this data rate? (fractional bits per symbol are ok for inclusion in results and use of all formulas.)
Appendix A

Linear Minimum Mean-Square Estimation, Normed Vector Spaces, Spectral Factorization, and Filter Realization

Section A.1 begins with the general MMSE estimate, before focusing rapidly on jointly Gaussian processes for which this estimate is always linear. Section A.2 develops the Orthogonality Principle that this text uses throughout to minimize mean square errors, both for scalar and vector processes. Section A.3 addresses scalar spectrum factorization and filter realization that is very important in solving many MSE problems also, especially for data transmission and noise whitening. The famed (scalar) Paley-Weiner Criterion is explained and derived in this Section. Section A.4 address a vector/matrix generalization of Paley Wiener Theory that this text calls MIMO Paley-Weiner, and which provides elegant solution generalizations to many MIMO situations of increasing interest. Section A.5 is short and lists the well known matrix inversion lemma.

A.1 General MMSE estimation

Generally, the MMSE estimate of one random variable or process \( x \) based on other random variables or processes, organized in a vector \( y \), is some function \( \hat{x} = f(y) \) that minimizes

\[
\arg \min_{f(y)} E \left[ (x - f(y))^2 \right].
\]  

(A.1)

In this text, the quantities \( x \) and \( y \) have zero mean\(^1\). The vector \( y \)'s dimensional elements generally have any organization, which can include time samples, frequency samples, spatial samples or others.

\begin{center}
\textbf{Theorem A.1.1 (Minimum-Mean-Square-Error Estimate)} The mean square error generally is \( \hat{x} = f(y) = E(x/y) \).
\end{center}

\textbf{Proof:}

\[
MSE = E \left[ (x - f(y))^2 \right] = E \left[ (x - E(x/y) + E(x/y) - f(y))^2 \right]
\]  

(A.2)

(A.3)

\(^1\)Any nonzero mean is non-information bearing and can be subtracted from the random process so for instance \( x \rightarrow x - E[x] \), without altering theoretical analysis.
\[
E \left\{ [x - E(x/y)]^2 + 2 \cdot [x - E(x/y)] \cdot [E(x/y) - f(y)] + [E(x/y) - f(y)]^2 \right\} \geq 0 \text{ when } f(y) = E(x/y).
\]

where the first left-side term is non-negative and does not depend on \( f(y) \), and the remaining terms are zero when \( f(y) = E(x/y) \). The MMSE is thus

\[ \sigma^2_{mmse} = E \left[ (x - E(y/x))^2 \right] \] QED. \hfill (A.5)

The MMSE estimate is \( E(x/y) \) in general for any \( x \) and \( y \) distribution(s).

### A.1.1 Gaussian MMSE, autocorrelation, and cross-correlation

Gaussian and also linear MMSE estimates have two constituent quantities:

**Definition A.1.1 (autocorrelation and cross-correlation matrices)**

The \( N \times N \) **autocorrelation matrix** for any \( N \)-dimensional random vector \( y \) is

\[ R_{yy} \triangleq E \left[ y \cdot y^* \right] \] \hfill (A.6)

When \( N = 1 \), the autocorrelation matrix becomes simply the random vector’s energy \( \sigma^2_y \). When stationary random vector process \( y_k \) has discrete (continuous) time index \( k \) (\( t \)), then the autocorrelation matrix becomes itself a time sequence (function)

\[ R_{yy,k} \triangleq E \left[ y_l \cdot y^*_{l-k} \right] \] \hfill (A.7)

\[ R_{yy}(t) \triangleq E \left[ y(u) \cdot y^*(u-t) \right] \] \hfill (A.8)

In this case when \( N = 1 \), the autocorrelation-matrix sequence becomes the random processes’ autocorrelation function. Similarly, The **cross-correlation matrix** for any two \( N_x \)- and \( N_y \)-dimensional random vectors \( x \) and \( y \) is the \( N_x \times N_y \) matrix

\[ R_{xy} \triangleq E \left[ x \cdot y^* \right] \] \hfill (A.9)

Also, clearly \( R_{yx} = R_{xy}^* \). When \( N_x = N_y = 1 \), the cross-correlation matrix becomes simply the two random vector’s correlation \( R_{x,y} \). The cross correlation between a scalar \( x \) and a vector \( y \) is thus a \( 1 \times N_y \) row vector \( R_{x,y} \). When the two jointly stationary random vector processes \( x_k \) and \( y_k \) have discrete (continuous) time index \( k \) (\( t \)), then the cross-correlation matrix becomes itself a time sequence (function)

\[ R_{xy,k} \triangleq E \left[ x_l \cdot y^*_{l-k} \right] \] \hfill (A.10)

\[ R_{xy}(t) \triangleq E \left[ x(u) \cdot y^*(u-t) \right] \] \hfill (A.11)

When \( N = 1 \), the autocorrelation-matrix sequence becomes the random processes’ cross-correlation function \( R_{xy}(t) \).

Often noise is Gaussian, which (see Chapter 2) leads to best transmitted signals also being Gaussian, and then all signals being Gaussian. Thus, the Gaussian distribution’s form has particular interest in
digital transmission. Multivariate Gaussian variables, organized in terms of two vectors \( x \) and \( y \) (each with possibly many elements), then depend only on the autocorrelation and cross-correlation matrices, organized into a single matrix \( R \) where

\[
R \Delta E \left\{ \begin{bmatrix} x \\ y \end{bmatrix} [x^* y^*] \right\} = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}
\]  
(A.12)

**Definition A.1.2 (Multivariate Gaussian Distribution)** It is helpful to provide both the real and complex forms. The real-valued multivariate Gaussian distribution is

\[
\text{real: } p(x, y) = (2\pi)^{-\frac{N_x + N_y}{2}} \cdot |R|^{-1/2} \cdot e^{-\frac{1}{2} \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \cdot R^{-1} \cdot [x^* y^*] \right\}} \quad \text{(A.13)}
\]

\[
\text{complex: } p(x, y) = (2\pi)^{-\frac{(N_x + N_y)}{2}} \cdot |R|^{-1/2} \cdot e^{-\frac{1}{2} \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \cdot R^{-1} \cdot [x^* y^*] \right\}} \quad \text{(A.14)}
\]

The two distributions are indeed equivalent when viewed in terms of this text’s per-dimensional normalizations.

For jointly Gaussian random variables and processes, the MMSE estimate is linear in \( y \), which can be found through somewhat tedious algebra by taking the ratio \( p(x, y)/p(y) \) and simplifying to a Gaussian form and reading the expected value:

\[
E[x/y] = R_{xy} \cdot R_{yy}^{-1} \cdot y \quad \text{(A.15)}
\]

The MMSE estimate’s autocorrelation is

\[
R_{x/y} = R_{xy} \cdot R_{yy}^{-1} \cdot R_{yx} \quad \text{(A.16)}
\]

The error \( x - E(x/y) \) has autocorrelation

\[
R_{x/y}^\perp = E_x - R_{xy} \cdot R_{yy}^{-1} \cdot R_{yx} \quad \text{(A.17)}
\]

In the case of scalar \( x = x \), then the result is readily seen to be the MMSE\(^2\)

\[
\sigma_{\text{mmse}}^2 \Delta \sigma_{x/y}^2 = E_x - R_{xy} \cdot R_{yy}^{-1} \cdot R_{yx} \quad \text{(A.18)}
\]

When \( R_{yy} \) is singular, then replace inverse with pseudoinverse in this text.

### A.2 The Orthogonality Principle and Linear MMSE estimation

This section shows that a linear MMSE estimator with any jointly stationary distributions leads to the same MMSE-estimate form as in (A.15) - (A.18). This section’s orthogonality principle provides a simple way to derive such linear MMSE estimates, but also leads to rich intuition, as later used throughout this text.

The random variable \( x \)'s linear MMSE estimate depends on a linear combination of observations of the random variables \( \{y_n=1,...,N\} \). That linear combination uses parameters \( w_n \), with the index \( n \) having one distinct value for each observation of \( y_n \) that is used. This linear MMSE minimizes

\[
E \left[ |e|^2 \right] \quad \text{(A.19)}
\]

\(^2\)This appendix will soon show the trace of the matrix \( R_{x/y} \) is corresponding vector MMSE.
where this error is
\[ e = x - \sum_{n=0}^{N-1} w_n \cdot y_n = x - w \cdot y. \]  
(A.20)

\( N \to \infty \) without difficulty. The row vector \( w \) is not random. The column vector \( y \) can be random, as can be the scalar \( x \). The orthogonality principle determines the values of \( w_n \):

**Theorem A.2.1 (Orthogonality Principle)** The minimum MSE must meet the following condition
\[ E[e \cdot y_n^*] = 0 \quad \forall \ n = 1, ..., N. \]  
(A.21)

**Proof:** Writing \( |e|^2 = |\Re(e)|^2 + |\Im(e)|^2 \) allows partial MSE differentiation with respect to both \( x_n \)’s real and imaginary parts for each \( n \). The real and imaginary derivatives pertinent parts are (realizing that all other \( w_i, i \neq n \), will drop from the corresponding partial derivatives)
\[ e_r = x_r - w_{r,n} \cdot y_{r,n} + w_{i,n} \cdot y_{i,n} \]  
(A.22)
\[ e_i = x_i - w_{i,n} \cdot y_{r,n} - w_{r,n} \cdot y_{i,n}, \]  
(A.23)

where subscripts of \( r \) and \( i \) denote real and imaginary part in the obvious manner. Then, optimization over \( w_{r,n} \) and \( w_{i,n} \) yields,
\[ \frac{\partial |e|^2}{\partial w_{r,n}} = 2e_r \frac{\partial e_r}{\partial w_{r,n}} + 2e_i \frac{\partial e_i}{\partial w_{r,n}} = -2(e_r y_{r,n} + e_i y_{i,n}) = 0 \]  
(A.24)
\[ \frac{\partial |e|^2}{\partial w_{i,n}} = 2e_r \frac{\partial e_r}{\partial w_{i,n}} + 2e_i \frac{\partial e_i}{\partial w_{i,n}} = 2(e_r y_{i,n} - e_i y_{r,n}) = 0. \]  
(A.25)

The desired result is found by taking expectations and rewriting the series of results above in vector form. Since the MSE is quadratic in the parameters \( w_{r,n} \) and \( w_{i,n} \), this setting must be a global minimum. QED.

If the autocorrelation matrix \( R_{yy} \) is strictly positive-definite, then the linear-MMSE estimator is unique.

### A.2.1 Some expansion of linear MMSE to vector estimates

While the scalar estimate follows simply as in Theorem A.2.1, estimates of multi-dimensional error signals require explanation. This subsection formalizes vector/Hilbert spaces to that objective.

**Definition A.2.1 (Vector Space)** A vector space \( \mathcal{V} \) for a scalar field \( \mathcal{F} \) and for all vectors \( v \in \mathcal{V} \) has two operations vector addition + and scalar multiplication \( \cdot \). The vector addition operation maps \( \mathcal{V} \otimes \mathcal{V} \to \mathcal{V} \) while scalar multiplication maps \( \mathcal{F} \otimes \mathcal{V} \to \mathcal{V} \). The vector space and operations must satisfy the following:

a. **closure of vector addition**, so that \( \forall u \in \mathcal{V} \) and \( v \in \mathcal{V} \), \( u + v \in \mathcal{V} \).

b. **closure of scalar-vector multiplication**, so that \( \forall \alpha \in \mathcal{F} \) and \( v \in \mathcal{V} \), \( \alpha \cdot v \in \mathcal{V} \).

c. **commutativity** \((u + v = v + u)\) and **associativity** \((\alpha (u + v) + w = u + (v + w))\) of vector addition.

d. **existence of vector addition’s zero element** \( 0 \in \mathcal{V} \) \( \ni \) \( 0 + v = v \) and **inverse element** \(-v \in \mathcal{V} \ni v + (-v) = 0\).
e. **Closure of scalar-scalar multiplication**, so that \( \forall \alpha, \beta \in \mathbb{F} \) and \( v \in \mathbb{V} \), \( \alpha \cdot \beta \cdot v \in \mathbb{V} \).

f. **Commutativity and associativity of scalar-scalar multiplication** so that \( \forall \alpha, \beta \in \mathbb{F} \) and \( v \in \mathbb{V} \), \( \alpha \cdot (\beta \cdot v) = (\alpha \cdot \beta) \cdot v = \beta \cdot (\alpha \cdot v) \).

item multiplicative identity \( \exists 1 \in \mathbb{F} \) such that \( 1 \cdot v = v \).

g. **Distributivity** of scalar multiplication over vector and field addition, \( \alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v \) and \( (\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v \).

A random vector space will select vectors \( v \in \mathbb{V} \) according to a probability density/distribution \( p(v) \), and then will have autocorrelation \( R_{uu} \) and cross-correlation \( R_{vv} \) matrices respectively. A vector space may have an infinite number of vector members. Simple examples are the \( N \)-dimensional vectors of real (or complex) numbers, \( \mathbb{V} = \mathbb{R}^N (\mathbb{C}^N) \), with scalar field of real \( \mathbb{R} \) and complex \( \mathbb{C} \) scalars respectively. Chapter 2 also had vector codewords selected from the vector space of \( N \)-dimensional integers (or Gaussian integers) with finite-field arithmetic and scalar multiplication.

A normed vector space assigns a non-negative real scalar \( \|v\| \in \mathbb{R}^+ \) to all \( v \in \mathbb{V} \) that measures length or size.

**Definition A.2.2 (Normed Vector Space)** A normed vector space has an additional functional mapping, or norm \( \|v\| \geq 0 \), to non-negative real numbers, \( \mathbb{V} \to \mathbb{R}^+ \), as well as positive-real scalar-multiplication norm simply written as the absolute value \( |\alpha| \geq 0 \).

The vector norm is \( \|v\| \forall v \in \mathbb{V} \) and has the following properties:

a. satisfaction of the **triangle inequality** \( \|u + v\| \leq \|u\| + \|v\| \).

b. **Uniform scaling** \( \|\alpha \cdot v\| = |\alpha| \cdot \|v\| \).

c. **Unique zero** norm (length) \( \|v\| = 0 \iff v = 0 \).

Normed vector spaces allow specification also of a **distance** between two vectors \( d(v, u) = \|v - u\| \), which is zero only if the two vectors are equal. A normed vector space is a **Hilbert Space** if it is complete, which basically means that any sum of the vector space’s norms is finite.

This text’s far most common norm is the **Euclidean norm** for real or complex vectors:

\[
\|v\| = \sqrt{\sum_{n=1}^{N} |v_n|^2} .
\]  
(A.26)

However, there are other norms and generally the \( p \)-norm is \( \|v_n\|_p = \left\{ \sum_{n=1}^{N} |v_n|^p \right\}^{1/p} \), which is the Euclidean norm for \( p = 2 \), the sum of the absolute values for \( p = 1 \) and the maximum element magnitude for \( p = \infty \). The Euclidean norm aligns particularly well with Gaussian and Linear MMSE developments.

A matrix can be viewed as a collection of column (or row) vectors from a vector space, so vector space properties and norms apply to each of the columns and therefore to the entire matrix, except the norm would translate to a set of real numbers. Matrix multiplication also expands the types of operations, but is of course not commutative. This text however focuses for matrix norm definition **only on positive-semi-definite square matrices**:

**Definition A.2.3** Matrix norm A matrix norm for an \( N \times N \) square positive-semi-definite matrix \( R \) is

\[
\|R\| = \sqrt{\text{trace}(R)} .
\]  
(A.27)

This matrix norm satisfies all the properties in Definition A.2.2.
Linear MMSE’s expansion to vectors involves semantic bookkeeping that is notationally convenient. The error becomes a vector \( e \), which might immediately raise the question of which element to minimize. However the error vector definition expands to

\[
e = x - W \cdot y ,
\]

where \( W \) becomes an \( N_x \times N_y \) matrix. Each row of \( W \) is a row vector \( w_n \) that estimates \( x_n \) from \( y \). The mean norm is

\[
E \left[ \|e\|^2 \right] = \sum_{n=0}^{N_x} |e_n|^2 .
\]

The (mean) squared error components are variables separable in that each depends only on its corresponding \( w_n \). Minimization of \( \|e\|^2 \) is therefore the same as separate minimization of each component. The orthogonality principle then applies to each, which can be written simply as

\[
W = R_{xy} \cdot R_{yy}^{-1} .
\]

Perhaps more interestingly, the minimized error components are also the diagonal elements of

\[
Ree = R_{xx} = R_{xx} - R_{xy} \cdot R_{yy}^{-1} \cdot R_{yx} ,
\]

which then clearly has a minimum \( \sqrt{\text{trace} \{Ree\}} = \|Ree\| \). Linear MMSE is thus the same as minimizing the error-autocorrelation matrix’ (squared) norm.

---

**Lemma A.2.1 (Generalized Pythagorean Theorem (GPT))** The following holds for vector MMSE estimation:

\[
R_{xx} = Ree + \hat{R}_{xx} .
\]

**Proof:** The proof follows directly from \( \hat{x} = R_{xy}R_{yy}^{-1} \cdot y \) and Equation A.31. QED.

The GPT conveniently allows direct analysis of many Chapter 3 receiver structures’ performance. The error vector \( e \)’s dimensionality equals the dimensionality of \( x \). In this text, transmitted dimensions may be in time, usually indexed by \( k \) for discrete time and infinite or semi-infinite. \( N \) often corresponds in Chapter 4 and beyond to a frequency index where the basic functions within a symbol are indexed to frequency. The error vector may also correspond to spatial dimensions \( l = 1, ..., L_x \). Whatever the index bookkeeping, the Orthogonality Principle and the GPT hold.

The MMSE minimizes the sum of the eigenvalues of \( Ree = Q \Lambda_e Q^* \) where \( \Lambda_e \) is a diagonal matrix of non-negative real eigenvalues (see Matlab eig command) and \( QQ^* = Q^*Q = I \), because the trace is the sum of the eigenvalues. This is immediately evident because multiplication by \( Q \) does not affect the error vector’s norm \( \|e\| = \|Q \cdot e\| = \|e\| \). So the MMSE estimator matrix then would be

\[
\hat{W} = Q \cdot W ,
\]

which is still now separable in the \( \hat{e} \) uncorrelated components. This means the determinant \( |Ree| \) has also been minimized because the product of the eigenvalues is the determinant (as should be obvious from observing \( |Q| = 1 \)), and indeed is also separable multiplicatively. Thus, often MMSE vector problems are written in terms of minimizing the determinant of the error autocorrelation matrix, but this is the same as minimizing the sum of component MSE’s.

---

\(^3\)sometimes a maximum number of time dimensions may be provided as \( N \) corresponding to a single transmission stream within a symbol, but sometimes the index \( k \) corresponds to a symbol and so there may be up to \( KN \) dimensions in a concatenation of \( K \) successive \( N \)-dimensional symbols.
A.3 Spectral Factorization and Scalar Filter Realization

This section describes scalar filter-based autocorrelation realization via spectral factorization. The realized filters will be causal, causally invertible (minimum phase), and monic. Such realization invokes the so-called Paley-Wiener Criterion (PWC) that is constructively developed and proven as part of the realization process, basing the proof on discrete sequences but covering also continuous signals\(^4\). Section A.4 generalizes these filters and/or process and Paley Wiener criterion to MIMO (matrix filters and/or processes), which will require understanding Cholesky Factorization of simple matrices first from Subsection A.3.6 that ends this section and precedes Section A.4.

A.3.1 Some Transform Basics: $D$-Transform and Laplace Transform

This subsection addresses use of $D$-transform notation for sequences\(^5\):

**Definition A.3.1 (D-Transforms)** A sequence $x_k \forall$ integer $k \in (-\infty, \infty)$ has $D$-Transform $X(D) = \sum_{k=-\infty}^{\infty} x_k \cdot D^k$ for all $D \in \mathcal{D}_x$, where $\mathcal{D}_x$ is the region of convergence of complex $D$ values for which the sum $X(D)$ converges, and $X(D)$ is analytic\(^6\) in $\mathcal{D}_x$. The inverse transform is a clockwise line/contour integral around any closed circle in the region of convergence $x_k = \frac{1}{2\pi j} \oint_{D \in \mathcal{D}_x} X(D) \cdot D^{1-k} \cdot dD$.\(^7\)

The sequence $x_k$ can be complex. The symbol rate will be presumed $T = 1$ in this appendix\(^8\). The sequence $x^*_k$ has a $D$-Transform $X^*(D^*) = \sum_{k=-\infty}^{\infty} x^*_k \cdot D^{-k}$. When the region of convergence includes the unit circle, the discrete-time sequence’s Fourier transform exists as $X(e^{-j\omega}) = X(D)|_{D=e^{-j\omega}}$, and such sequences are considered to be “stable” or “realizable” (non-causal sequences become realizable with sufficient delay, or infinite delay in some limiting situations).

A sufficient condition for the discrete-time sequence’s Fourier transform to exist (the sum $X(D)$ converges) is that the sequence itself be absolutely summable, meaning

$$\sum_{k=-\infty}^{\infty} |x_k| < \infty \quad , \quad (A.34)$$

or equivalently the sequence belongs to the (Hilbert infinite-dimensional vector) space of sequences $L^1$. Another sufficient condition is that the sequence belongs to $L^2$ or has finite energy according to

$$\sum_{k=-\infty}^{\infty} |x_k|^2 < \infty \quad . \quad (A.35)$$

---

\(^4\)In both cases, these proofs are applicable to deterministic magnitude-squared functions or to stationary random processes with consequently non-negative real power spectra.

\(^5\)The reader may insert $Z^{-1}$ for $D$ to relate to other developments where Z transforms are used.

\(^6\)Analytic here means the function and all its derivatives exist and are bounded in $\mathcal{D}_x$ - not to be confused with Chapter 2’s “analytic-equivalent” signals. The analytic function’s values or derivatives may be infinite only at a countable number of zero measure values known as the function’s poles.

\(^7\)Typically the inverse transform is implemented by partial fractions, which often arise in communication problems and in most approaches to contour integration anyway, whenever the $D$-transform is a rational fraction of polynomials in $D$.

\(^8\)The generalization for $T \neq 1$ is addressed in the specific text sections of this chapter and later chapters when necessary. For instance, see Table 3.1 for the generalization of transforms under sampling for all $T$. Sampling should not be confused with the bi-linear transform: the former corresponds to conversion of a continuous waveform to discrete samples, while the latter maps filters or functions from/to continuous to/from discrete time and thus allows use of continuous- (discrete-) time filter realizations in the other discrete- (continuous-) time domain.
The similarity of the form of transform and inverse then allows equivalently that the inverse Fourier Transform \((\frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega k} \cdot d\omega)\) exists if:

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{-j\omega})| \cdot d\omega < \infty ,
\]  
(A.36)

or if

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{-j\omega})|^2 \cdot d\omega < \infty .
\]  
(A.37)

While \(x_k \in L_1\) or \(x_k \in L_2\) are sufficient conditions, this text uses non-\(L_1\)-nor-\(L_2\) functions that can have Fourier transforms. These “generalized” functions include the Dirac Delta function \(\delta(t)\) or \(\delta(\omega)\), so that for instance \(\cos(\omega_0 k)\) has Fourier Transform \(\pi [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]\), or the series \(x_k = 1\) has transform \(2\pi \delta(\omega)\), even though neither of the sums above in Equations (A.34) and (A.35) converge for these functions. These types of generalized-function-assisted Fourier Transforms are “on the stability margin” where values of \(D\) in a convergence-region series that approaches closely the unit circle (outside, but not on) so that existence criteria have limiting values in (A.36) or (A.37) in a “generalized-function” sense.

A continuous-time sequence \(x(t)\) has a Laplace Transform \(X(s)\) defined over a convergence region \(S_x\) as

**Definition A.3.2 (Laplace Transform)** A function \(x(t)\) has Laplace-Transform \(X(s) = \int_{-\infty}^{\infty} x(t) e^{-st} dt\), which converges/exists for all \(s \in S_x\), where \(S_x\) is the convergence region of complex \(s \in \mathscr{C}\) values. The inverse transform is \(\frac{1}{2\pi i} \oint X(s) e^{st} ds\) when the closed contour of integration is in \(S_x\).

The time function \(x(t)\) can be complex. The function \(x^*(-t)\) has a LaplaceTransform \(\int_{-\infty}^{\infty} x^*(-t) e^{-st} dt = X^*(-s^*)\). When the region of convergence includes the \(j\omega\) axis, the Fourier transform exists as \(X(\omega) = X(s)|_{s=\omega}\), and such functions are considered to be “stable” or “realizable” (non-causal functions become realizable with sufficient delay, or infinite delay in some limiting situations).

A sufficient condition for the continuous Fourier transform to exist (the integral converges) is that the function be absolutely integrable, meaning

\[
\int_{-\infty}^{\infty} |x(t)| \cdot dt < \infty , \text{ or equivalently}\]  
(A.38)

\[
\frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| \cdot d\omega < \infty ,
\]  
(A.39)

or equivalently that the function \(x(t)\) belongs to the space of continuous functions \(L^1\). Another sufficient condition is that function \(x(t)\) belongs to \(L^2\) or has finite energy according to

\[
\int_{-\infty}^{\infty} |x(t)|^2 \cdot dt < \infty , \text{ or equivalently}\]  
(A.40)

\[
\frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 \cdot d\omega < \infty
\]  
(A.41)

Similar to the discrete-time D-Transform, “generalized” functions complete the capability to handle continuous Fourier Transforms that are “on the stability margin” where values of \(s\) in a convergence-region sequence arbitrarily close to the \(j\omega\) axis (left of, but not on this axis) will converge so the criteria satisfy (A.40) or (A.41) in a limiting or generalized sense.

**A.3.2 Autocorrelation and Non-Negative Spectra Magnitudes**

Of particular interest in this text, and generally in digital communication, are the autocorrelation functions and associated power spectra for stationary and wide-sense stationary processes. These concepts are revisited briefly here for discrete processes before returning to filter realization of a given specified non-negative Fourier Transform magnitude.
Definition A.3.3 (Autocorrelation and Power Spectrum for Sequences) If \( x_k \) is any stationary complex sequence, its autocorrelation function is \( r_{xx,j} = E[x_k x_{k-j}^*] \) with D-Transform \( R_{xx}(D) \); symbolically\(^9\)

\[
R_{xx}(D) \triangleq E \left[ X(D) \cdot X^*(D^{-*}) \right] .
\] (A.42)

By stationarity, \( r_{xx,j} = r_{xx,-j}^* \) and \( R_{xx}(D) = R_{xx}^*(D^{-*}) \). The power spectrum of a stationary sequence is the Fourier transform of its autocorrelation function, which is written as

\[
R_{xx}(e^{-j\omega}) = R_{xx}(D)|_{D=e^{-j\omega}} , \quad -\pi < \omega \leq \pi ,
\] (A.43)

which is real and nonnegative for all \( \omega \). Conversely, any function \( R(e^{-j\omega}) \) that is real and nonnegative over the interval \( \{-\pi < \omega \leq \pi\} \) is a power spectrum, and has an autocorrelation function satisfying \( R(D) = R^*(D^{-*}) \).

Generally, conjugate symmetric sequences with \( a_k = a_{-k}^* \) have real Fourier transforms \( A(e^{-j\omega}) \in \mathbb{R} \) that however can be negative. Thus, a necessary and sufficient condition to be an autocorrelation sequence is that \( A(e^{-j\omega}) \geq 0 \), or a positive real sequence. The term “positive real” used by mathematicians should not be confused to mean that each time-domain sequence value is positive and real\(^10\).

The quantity \( E \left[ |x_k|^2 \right] \) is \( \mathcal{E}_x \), or \( \mathcal{E}_x \) per dimension, and can be determined from either the autocorrelation function or the power spectrum as follows:

\[
\mathcal{E}_x = E \left[ |x_k|^2 \right] = r_{xx,0} \quad (A.45)
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} R_{xx}(e^{-j\omega})d\omega . \quad (A.46)
\]

If the matrix sequence \( R_k \), for instance perhaps formed as the inverse transform of a fixed filter \( \mathcal{F}^{-1}(H(e^{-j\omega}) \cdot H^*(e^{j\omega})) \), is deterministic, then the averages are not necessary. The power spectra is then essentially the magnitude squared of the Fourier Transform \( R(e^{-j\omega}) \triangleq |H(e^{-j\omega})|^2 \geq 0 \) for discrete time. These Fourier Transforms’ magnitudes can be thought of also as power spectra, and the corresponding inverse transforms as autocorrelation functions in this text.

Definition A.3.4 (Autocorrelation and Power Spectrum for Continuous-time Functions)
If \( x(t) \) is any stationary (or WSS, see Chapter 1’s Appendix A) complex function, its autocorrelation function is \( r_{xx,c}(t) = E[x_c(u)x_{c+}(u-t)] \) with Laplace Transform \( R_{xx,c}(s) \); symbolically\(^11\)

\[
R_{xx,c}(s) \triangleq E \left[ X(s) \cdot X^*(-s^*) \right] .
\] (A.47)

By stationarity, \( r_{xx,c}(t) = r_{xx,c}^*(-t) \) and \( R_{xx,c}(s) = R_{xx,c}^*(-s^*) \). The power spectrum of a stationary continuous-time process is the Fourier transform of its autocorrelation function, which is written as

\[
R_{xx,c}(\omega_c) = R_{xx,c}(s)|_{s=j\omega_c} , \quad -\infty < \omega_c \leq \infty ,
\] (A.48)

which is real and nonnegative for all \( \omega_c \). Conversely, any function \( R(\omega) \) that is real and nonnegative over the interval \( \{-\infty < \omega_c \leq \infty\} \) is a power spectrum and has autocorrelation function \( r(t) = r^*(-t) \).

---

\(^9\) The expression \( R_{xx}(D) \triangleq E \left[ X(D)X^*(D^{-1}) \right] \) is used in a symbolic sense, since the terms of \( X(D)X^*(D^{-1}) \) are of the form \( \sum_k x_k x_{k-j}^* \), implying the additional operation \( \lim_{N \to \infty} [1/(2N+1)] \sum_{-N \leq k \leq N} \) on the sum in such terms.

\(^{10}\) Positive values for all discrete time instants means that the corresponding Fourier Transform is an “autocorrelation” function of frequency.

\(^{11}\) The expression \( R_{xx,c}(s) \triangleq E \left[ X(s)X^*(-s^*) \right] \) is used in a symbolic sense, since the terms of \( X(s)X^*(-s^*) \) are of the form \( \int_{-\infty}^{\infty} E[x_c(u)x_{c+}^*(u-t)]du \), implying the additional operation \( \lim_{T \to \infty} [1/(2T)] \int_{-T \leq u \leq T} \) on the integral in such terms.
Generally, conjugate symmetric functions with $a(t) = a^*(-t)$ have real Fourier transforms $A(\omega) \in \mathbb{R}$ that however can be negative. Thus, a necessary and sufficient condition to be an autocorrelation sequence is that $A(\omega) \geq 0$, or a positive real sequence. The term “positive real” used by mathematicians should not be confused to mean that each time-domain function value is positive and real\(^{12}\), but instead refers to the Fourier Transform being “positive real.”

The quantity $E \left[ |x_c(t)|^2 \right]$ is $P_x$, or the power of the random continuous-time process, and can be determined from either the autocorrelation function or the power spectrum as follows:

$$P_x = E \left[ |x_c(t)|^2 \right] = r_{x,x}(0) \quad \text{ (A.49)}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} R_{x,x}(\omega_c) \cdot d\omega_c \quad \text{ (A.50)}$$

If the sequence in question (a fixed filter for instance) is deterministic, then the averages are not necessary above. The power spectra is then essentially the magnitude squared of the Fourier Transform

$$R(\omega_c) \triangleq |X_c(\omega_c)|^2 \geq 0 \quad \text{for continuous time. These Fourier Transforms' magnitudes can viewed as power spectra, and the corresponding inverse transforms as autocorrelation functions in this text.}$$

### A.3.3 The Bilinear Transform and Spectral Factorization

This section denotes a continuous-time function’s Fourier Transform radian frequency by $\omega_c$ while the discrete-time sequence’s Fourier Transform variable will be $\omega$ (with no subscript of $c$). Similarly all continuous-time quantities will use a subscript of $c$ to avoid confusion with discrete time. For the transforms, if a transform $X(D)$ or $X_c(s)$ exists in their respective regions of convergence, then the transforms $e^{X(D)}$ and $e^{X_c(s)}$ also exist in that same region of convergence\(^{13}\). Similarly then, ln$(X(D))$ and ln$(X_c(s))$ also have the same regions of convergence.

A filter-design technique for discrete-time filters uses what is known as the “bi-linear” transform to map a filter designed in continuous time into a discrete-time filter (or vice versa):

**Definition A.3.5 (Bilinear Transforms)** The bilinear transform maps between discrete-time $D$ Transforms and continuous-time Laplace Transforms according to

$$s = \frac{1 - D}{1 + D} \quad \text{ (A.52)}$$

and conversely

$$D = \frac{1 - s}{1 + s} \quad \text{ (A.53)}$$

The bilinear transform can also relate discrete-time and continuous-time Fourier Transforms by inserting $D = e^{-j\omega}$ and $s = j\omega_c$. The $j\omega_c$ (complex) axis from 0 to $\pm \infty$ corresponds to mapping $D = e^{-j\omega}$ along the unit circle (centered at origin of $D$ plane) from the (real, imaginary) D-plane point $[1,0]$ of 0 radians to the point of $\pi$ radians (or $[-1,0]$) clockwise for positive $\omega_c$ (counter clockwise for negative $\omega_c$). This bilinear transform scales or compresses the infinite range of frequencies $\omega_c \in (-\infty, \infty)$ for the continuous-time Fourier Transform to the finite range of frequencies $\omega \in [-\pi, \pi]$ for the discrete-time Fourier Transform. The bilinear transform does not correspond to sampling (and why $T = 1$ here to avoid confusion) to go from continuous time to discrete time. The filter designer needs to convert the filter cut-off frequencies with this compression/scaling in mind. A stable design in continuous time corresponds to all poles on/in the left-half plane of $s$ (or the region of convergence includes, perhaps in limit, the $j\omega_c$ axis). These poles will map (with the Bilinear Transform’s frequency-scale conversion) into corresponding points through the bilinear transform in the region outside (or in limiting sense on)

\(^{12}\)Positive values for all time means that the corresponding Fourier transform is an “autocorrelation” function of frequency.

\(^{13}\)This follows from the derivative of $e^{f(x)}$ for any function $f(x)$ is $e^{f(x)} \cdot f'(x)$ so if the function existed at $x$, then $e^{f(x)}$ also exists at all argument values, and then since $f'(x)$ also exists, then so does the derivative. This argument can be recursively applied to all successive derivatives that of course exist for $f(x)$ in its domain of convergence.
the unit circle $|D| = 1$, and vice-versa. Similarly a minimum-phase design (all poles and zeros in LHP) will map to all poles/zeros outside the unit circle, and vice-versa.

For the Fourier Transform, the blinear transformation maps frequency according to

$$j\omega_c = \frac{1 - e^{j\omega}}{1 + e^{j\omega}}$$  \hspace{1cm} (A.54)

$$= j\tan \left(\frac{\omega}{2}\right)$$  \hspace{1cm} (A.55)

$$\omega = 2\arctan(\omega_c)$$  \hspace{1cm} (A.56)

$$d\omega = \frac{d\omega_c}{1 + \omega_c^2}.$$  \hspace{1cm} (A.57)

The spectral factorization of a discrete-time autocorrelation function’s $D$-Transform is:

**Definition A.3.6 (Factorizability for Sequences)** An autocorrelation function $R_{xx}(D)$, or equivalently any non-negative real $R_{xx}(e^{j\omega})$ so that $r_k = r^*_{-k}$, will be called **factorizable** if it can be written in the form

$$R_{xx}(D) = S_{x,0}G_x(D)G_x^*(D^{-*}).$$  \hspace{1cm} (A.58)

where $S_{x,0}$ is a finite positive real number and $G_x(D)$ is a canonical filter response. A filter response $G_x(D)$ is called **canonical** if it is causal ($g_{x,k} = 0$ for $k < 0$), monic ($G_x(s = 0) = 1$), and minimum-phase (all of its poles and zeros are outside or on the unit circle). If $G_x(D)$ is canonical, then $G_x^*(D^{-*})$ is **anticanonical**; i.e., anticausal, monic, and maximum-phase (all poles and zeros inside or on the unit circle).

The region of convergence for factorizable $R_{xx}(D)$ clearly includes the unit circle, as do the regions for both $G_x(D)$ and $G_x^*(D^{-*})$. If $G_x(D)$ is a canonical response, then $\|g_x\|^2 \triangleq \sum_j |g_{x,k}|^2 \geq 1$, with equality if and only if $G_x(D) = 1$, since $G_x(D)$ is monic. Further, the inverse also factorizes similarly into

$$R_{xx}^{-1}(D) = \left(1/S_{x,0}\right)G_x^{-1}(D)G_x^{-*}(D^{-*}).$$  \hspace{1cm} (A.59)

Clearly if $R_{xx}(D)$ is a ratio of finite-degree polynomials in $D$, then it is factorizable (simply group poles/zeros together for inside and outside of the circle - any on unit circle will also appear in conjugate pairs so easily separated). For the situation in which $R_{xx}(D)$ is not already such a polynomial, the next section generalizes through the Paley-Wiener Criterion. Also, if $R_{xx}(D)$ is factorizable, then the corresponding $R_{x,x_\cdot}(s) = R_{xx}\left(\frac{1-s}{1+s}\right)$ is also factorizable into

$$R_{x,x_\cdot}(s) = S_{x,0}G_{x_\cdot}(s)G_{x_\cdot}^*(-s^*)$$  \hspace{1cm} (A.60)

**Definition A.3.7 (Factorizability for Continuous Functions)** An autocorrelation function $R_{x,x_\cdot}(s)$, or equivalently any non-negative real power spectrum $R_{x,x_\cdot}(\omega)$ so that $r(t) = r^*(-t)$, will be called **factorizable** if it can be written in the form

$$R_{x,x_\cdot}(s) = S_{x,0}G_{x_\cdot}(s)G_{x_\cdot}^*(-s^*)$$  \hspace{1cm} (A.61)

where $S_{x,0}$ is a finite positive real number and $G_{x_\cdot}(s)$ is a canonical filter response. A filter response $G_{x_\cdot}(s)$ is called **canonical** if it is causal ($g_{x_\cdot}(t) = 0$ for $t < 0$), monic ($g_{x_\cdot}(0) = 1$), and minimum-phase (all of its poles and zeros are in the left half plane). If $G_{x_\cdot}(s)$ is canonical, then $G_{x_\cdot}^*(-s^*)$ is **anticanonical**; i.e., anticausal, monic, and maximum-phase (all poles and zeros inside in the right half plane).

The region of convergence for factorizable $R_{x,x_\cdot}(s)$ clearly includes the $j\omega_c$ axis, as do the regions for both $G_{x_\cdot}(s)$ and $G_{x_\cdot}^*(-s^*)$.  

364
If $G_{x_c}(s)$ is a canonical response, then $\|g_{x_c}\|^2 \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} |g_{x_c}(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |G_{x_c}(\omega)|^2 d\omega$. Further, the inverse also factorizes similarly into

$$R_{x,x_c}^{-1}(s) = (1/S_{x,0}) \cdot G_{x,c}^{-1}(s) \cdot G_{x,c}^{-*}(-s^*) \ .$$  (A.62)

Clearly if $R_{xx}(s)$ is a ratio of finite-degree polynomials in $s$, then it is factorizable (simply group poles/zeros together for left and right half planes - any on imaginary axis will also appear in conjugate pairs, and so are easily separated). When not already such a polynomial, the next section generalizes this situation through the Paley-Wiener Criterion.

### A.3.4 The Paley-Wiener Criterion

Minimum-phase signals or filters are of interest in data transmission not only because they are causal and admit causal invertible inverses (one of the reasons for their study more broadly in digital signal processing) but because they allow best results with Decision Feedback as in Section 3.6. These minimum-phase filters are also useful in noise whitening.

Calculation of $S_{x,0}$ for a factorizable $D$-Transform follows Equations (3.232) to (3.233) in Section 3.6 as

$$S_{x,0} = e^{\frac{i}{2\pi} \int_{-\pi}^{\pi} \ln|R_{xx}(e^{i\omega})|\ d\omega} \quad \text{or} \quad \ln(S_{x,0}) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln|R_{xx}(e^{i\omega})|\ d\omega \ ,$$  (A.63)

which has components corresponding to the last two terms in (A.58) integrating to zero because they are periodic with no constant components and being integrated over one period of the fundamental frequency. The integral in (A.64) must be finite for the exponent in (A.63) to be finite (also, any exponent of a real number is a real positive number, so $S_{x,0} > 0$ and real, consistent with the power spectral density). The integral of a power spectral density’s natural log is fundamental in filter realization and in the Paley Wiener Criterion to follow. Again, $\mathcal{P}_x$ is the same for $R_{xx}(D)$ as for $\ln|G_x(D)|$ and includes the unit circle; this also means the region of convergence for $\ln |G_x(D)|$ also is the same as for $G_x(D)$ and includes the unit circle. Further the region of convergence for $G_x^{-1}(D)$ also includes the unit circle and is the same as for $\ln |G_x^{-1}(D)|$.

The calculation of $S_{x,0}^{-1}$ has a very similar form to that of $S_{x,0}$:

$$S_{x,0}^{-1} = e^{-\frac{i}{2\pi} \int_{-\pi}^{\pi} \ln|R_{xx}(e^{i\omega})|\ d\omega} \quad \text{or} \quad \ln(S_{x,0}^{-1}) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln|R_{xx}(e^{i\omega})|\ d\omega \ ,$$  (A.65)

Because (A.65) and (A.66) are similar, just differing in sign, and because any functions of $G_x$ (including in particular $\ln$ or $|\bullet|$) are all periodic, factorizability also implies

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln|R_{xx}(e^{i\omega})|\ d\omega < \infty$$  (A.67)

(or the function $\ln |R_{xx}(D)|$ exists because the autocorrelation in L1 is absolutely integrable). Essentially this integral’s finite value corresponds to a factorizable $R_{xx}(D)$ and means that the sequence’s Fourier Transform has no frequencies (except point frequencies of non-zero measure) at which it can be either zero or infinite. The non-infinite is consistent with the basic criterion (A.34) to be absolutely integrable, but the the non-zero portion corresponds intuitively to saying any filter that has some kind of “dead band” is singular. Any energy in these singular bands would be linear combinations of energy at other frequencies. Singular bands thus carry no new information, and can be viewed as useless: A signal with such a dead band is wasting energy on components transmitted already at other frequencies that exactly cancel. A filter with such a dead band would block any information transmitted in that band, making reliable data-detection/communication impossible (kind of an inverse to the reversibility concept and
Theorem A.3.1 (Paley Wiener Criterion) If $R_{xx}(e^{-j\omega})$ is any power spectrum such that both $R_{xx}(e^{-j\omega})$ and $\ln R_{xx}(e^{-j\omega})$ are absolutely integrable over $-\pi < \omega \leq \pi$, and $R_{xx}(D)$ is the corresponding autocorrelation function, then there is a canonical discrete-time response $G_x(D)$ that satisfies the equation

$$R_{xx}(D) = S_{x,0} \cdot G_x(D) \cdot G_x^*(D^{-*}), \quad (A.68)$$

where the finite constant $S_{x,0}$ is given by

$$\ln S_{x,0} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln R_{xx}(e^{-j\omega}) d\omega. \quad (A.69)$$

For $S_{x,0}$ to be finite, $R_{xx}(e^{-j\omega})$ must satisfy the discrete-time Paley-Wiener Criterion (PWC)

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\ln R_{xx}(e^{-j\omega})| d\omega < \infty. \quad (A.70)$$

The PWC’s continuous-time equivalent of this PWC is that the Fourier Transform of the continuous-time autocorrelation function is factorizable

$$R_{xx}(s) = S_{x,0} \cdot G_x(s) \cdot G_x^*(-s^*) , \quad (A.71)$$

where $G_x(s)$ is minimum phase (all poles and zeros in the left half plane or on axis in limiting sense) and “monic” $g_x(t)|_{t=0} = 1$, whenever

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\ln R_{xx}(e^{j\omega})}{1 + \omega^2} d\omega < \infty. \quad (A.72)$$

Constructive Proof: The equivalence of the two PW criteria in (A.70) and (A.72) (discrete-and continuous-time) follows directly from Equations (A.54) to (A.57). However, it remains to show that the condition is necessary and sufficient for the factorization to exist. The necessity of the criterion followed previously when it was shown that factorizability lead to the PWC being satisfied. The sufficiency proof will be constructive from the criterion itself.

The desired non-negative real in (A.63) and (A.64) frequency has a (positive or zero) real square root $R_{xx}^{1/2}(e^{j\omega})$ at each frequency, and this function in turn has a natural log

$$A(e^{-j\omega}) \triangleq \ln \left[R_{xx}^{1/2}(e^{-j\omega})\right]. \quad (A.73)$$

$A(e^{-j\omega})$ itself is also periodic and real, and by the PWC integral equation, is absolutely integrable and so has a corresponding Fourier representation

$$A(e^{-j\omega}) = \sum_{k=-\infty}^{\infty} a_k \cdot e^{-j\omega k}. \quad (A.74)$$
\[ a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(e^{-j\omega}) \cdot e^{j\omega} \cdot d\omega \quad (A.75) \]

Because this (Fourier Transform) \( A(e^{-j\omega}) \) is purely real, then \( a_k = a_{-k}^* \), and the D-Transform simplifies to

\[ A(D) = a_0 + \sum_{k=1}^{\infty} a_k \cdot D^k + \sum_{l=-1}^{-\infty} a_l \cdot D^l \quad (A.76) \]

and then by letting \( k = -l \) in the second sum,

\[ A(D) = a_0 + \sum_{k=1}^{\infty} [a_k + a_{-k}] \cdot D^k \quad (A.77) \]

\[ = a_0 + 2 \sum_{k=1}^{\infty} \Re[a_k] \cdot D^k \quad (A.78) \]

which defines a causal sequence \( a_k \) that corresponds to \( \ln \left[ R_{1/2}^{1/2}(D) \right] \). The sequence is causally invertible because \( \ln \left[ R_{-1/2}^{1/2}(D) \right] \) can be handled in the same way following (A.70). So,

\[ R_{1/2}(D) = e^{A(D)} \cdot e^{A^*(D^{-*})} \quad (A.80) \]

Then, the desired canonical factorization has the factors

\[ S_{x,0} = e^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[R_{1/2}(e^{-j\omega})] \cdot d\omega} \quad (A.81) \]

\[ G_x(D) = \frac{e^{A(D)}}{\sqrt{S_{x,0}}} \quad (A.82) \]

The corresponding continuous-time spectrum factorization then would be found with \( R_{xx}(s) = R_{xx} \left( \frac{1-s}{1+s} \right) \) and thus \( A_c(s) = A(\frac{1-s}{1+s}) \). Then, with \( s \to j\omega_c \)

\[ S_{x,c,0} = e^{\frac{1}{2\pi} \int_{-\infty}^{\infty} \ln[R_{xx,c}(j\omega)]} \cdot d\omega_c \quad (A.83) \]

\[ G_{x,c}(s) = \frac{e^{A_c(s)}}{\sqrt{S_{x,c,0}}} \quad (A.84) \]

If the original desired spectra were defined in continuous time, then it could be mapped into discrete time through \( \omega_c \to \tan(\frac{\pi}{2}) \) and then proceeding with that discrete-time mapped equivalent through the process above, ultimately leading to Equations (A.83) and (A.84). Sufficiency has thus been established in both discrete- and continuous-time.

\[ \text{QED.} \]

Minimum-phase functions and sequences have several interesting properties that can help understand their utility. If a “phasor” diagram were drawn from each pole and zero to a point on the unit circle (or imaginary axis for continuous time), the magnitude is the ratio of the zero-phasor-length products to the pole-phasor-length products while the phase is the sum of the zero phases minus the sum of the pole phases. For the minimum-phase D-Transform the phasor angle is measured from a horizontal line to the left (while for the minimum-phase Laplace Transform it is measured from a horizontal line to the right). These phase contributions are always the smallest with respect to the “other choice” of the zero/pole from the maximum-phase factor. Whence the name “minimum phase.” Perhaps more importantly, one can see as frequency increases the rate of change of the angle (the magnitude of delay) is smallest for this same choice. Equivalently, each frequency for the particular magnitude spectrum
of interest is delayed the smallest possible amount. With respect to all other pole/zero choices, the energy is maximally concentrated towards zero for any time period (among all waveforms with the same spectrum). In Section 3.6, the DFE feedback section thus has largest ratio of first tap magnitude to remaining taps’ summed magnitude, so thus smallest loss with respect to Chapter 3’s matched filter bound. Such minimal-energy delay allows inversion of the function with the also-minimum-phase/delay that is actually the negative of the first delay. In effect, this only occurs when the function is causal and causally invertible.

A.3.4.1 Illustrative Concepts

A first illustration investigates the general realization’s simplification when the function to be factored is already a ratio of finite polynomials. In this case, the following log power spectrum has a term like:

\[ \sum_k \left[ (1 + z_k D) \cdot (1 + z_k^* D^{-1}) \right] \quad (A.85) \]

for the numerator where \( z_k \) are the zeros, divided into min-phase set for first term and max-phase set for second term. There is a similar pole-term for the denominator. The example focuses on the log-of-square-root term

\[ \ln(1 + z_k D) = z_k D - \frac{(z_k D)^2}{2} + \frac{(z_k D)^3}{3} - \frac{(z_k D)^4}{4} + \ldots , \quad (A.86) \]

expanded via Taylor Series. The positive real \( A(e^{-j\omega T}) = \ln |(1 + z_k e^{-j\omega})| \) is periodic in \( \omega \) and therefore itself has a “Fourier Series” representation in terms of \( \omega \):

\[ A(e^{-j\omega T}) = \sum_{k=-\infty}^{\infty} a_k \cdot e^{-j\omega k} \quad . \quad (A.87) \]

Since \( A(e^{-j\omega T}) \) is real, then \( a_k = a_{-k}^* \) and thus

\[ A(e^{-j\omega T}) = a_0 + 2 \cdot \sum_{k=1}^{\infty} \Re \{ a_k \} \cdot e^{-j\omega k} \quad . \quad (A.88) \]

Then \( A(D) \) where \( e^{-j\omega} \rightarrow D \) is also minimum phase and clearly causal. Indeed then

\[ R(D) = e^{A(D)} \cdot e^{A^*(D^{-1})} = (1 + z_k D) \cdot (1 + z_k^* D^{-1}) \quad . \quad (A.89) \]

Because \( (A.85) \)’s sum, this process can be continued for each term, resulting \( A(D) \) be sum of such terms and thus \( e^{A(D)} \) being the product of corresponding terms. Pole terms can be handled similarly with a minus sign in front. Thus, the construction process leaves an

\[ R_{xx}(D) = \prod_{k=-L}^{L} \frac{(1 + z_k D)}{(1 + p_k D)} \cdot \frac{(1 + z_k^* D^{-1})}{(1 + p_k^* D^{-1})} \quad . \quad (A.90) \]

While the result could be produced by simple factoring, it illustrates the more general construction’s thought processes. The more general procedure essentially corresponds to \( L \rightarrow \infty \) in the example. This prevents simple factorization in some cases, even with poles. Essentially, the extended process corresponds to polynomials that cannot be factored, as the following illustrates.

The periodic function

\[ R(e^{-j\omega}) \triangleq 1 - \frac{\omega}{\pi} \quad \forall |\omega| \leq \pi \quad (A.91) \]

is clearly positive real. It also clearly satisfies PWC. It’s log-square-root function is also period and equal to

\[ A(e^{-j\omega}) = \ln \left\{ \sqrt{1 - \frac{|\omega|}{\pi}} \right\} = \frac{1}{2} \ln \left( 1 - \frac{|\omega|}{\pi} \right) \quad . \quad (A.92) \]
This function has a Fourier Series representation

\[ A(\omega) = \sum_k f_k \cdot e^{-j\omega k} \]  \hspace{1cm} (A.93)

with \( f_{-k} = f_k^* \), with

\[ f_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( 1 - \frac{|\omega|}{\pi} \right) \cdot e^{-j\omega k} \cdot d\omega , \]  \hspace{1cm} (A.94)

which simplifies to

\[ f_{k\ge0} = \frac{1}{2\pi} \int_{0}^{\pi} \left( 1 - \frac{|\omega|}{\pi} \right) \cdot \cos(\omega k) \cdot d\omega . \]  \hspace{1cm} (A.95)

Matlab can be used to integrate numerically and find the coefficients (for \( k < 0 \), take conjugate).

```matlab
>> F=zeros(1,11);
>> F(2) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(x)) ,0,pi);
>> F(3) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(2*x)) ,0,pi);
>> F(4) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(3*x)) ,0,pi);
>> F(5) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(4*x)) ,0,pi);
>> F(6) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(5*x)) ,0,pi);
>> F(7) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(6*x)) ,0,pi);
>> F(8) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(7*x)) ,0,pi);
>> F(9) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(8*x)) ,0,pi);
>> F(10) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(9*x)) ,0,pi);
>> F(11) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi).*cos(10*x)) ,0,pi);
>> F(1) = (1/(2*pi))*integral(@(x) (log(1-abs(x)/pi)) ,-pi,pi) =
            -1.0000  0.2947  -0.1129  0.0888  -0.0594  0.0520  -0.0403  0.0367
            -0.0305  0.0284  -0.0245
```

So,

\[ A(D) \approx -1 + .2947D - .1129D^2 + .0888D^3 - .0594D^4 + .052D^5 - .0403D^6 + .0367D^7 - .0305D^8 + .0284D^9 - .0245D^{10} . \]
Figure A.1: Illustration of non pole-zero spectral factorization.

Figure A.3.4.1 illustrates the positive-index coefficients for the first 10 values, which are clearly decreasing in magnitude. It is not hard to show by induction that the magnitude of the coefficients is monotonically decreasing, and of course since this function already satisfies PWC, it is absolutely convergent. The canonical factor is therefore:

\[ G(D) = \frac{1}{\sqrt{S_0}} \cdot e^{-1.2947D - 0.1129D^2 + 0.888D^3 - 0.0594D^4 + 0.052D^5 - 0.0403D^6 + 0.0367D^7 - 0.0305D^8 + 0.0284D^9 - 0.0245D^{10}}. \]  

(A.96)

Using \( e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \) in (A.96) will lead to all positive powers in \( D \) and a multi-term expression for \( \sqrt{S_0} \cdot G(D) \) that could be truncated in practice as the division by \( n! \) decreases the value of higher-order terms. \( \sqrt{S_0} \) is this expansion’s constant term.

### A.3.5 Linear Prediction

The inverse of \( R_{xx}(D) \) is also an autocorrelation function and can be factored when \( R_{xx}(D) \) also satisfies the PW criterion with finite \( \delta_{x,0} \). In this case, as with the MMSE-DFE in Section 3.6, the inverse autocorrelation factors as

\[ R_{xx}^{-1}(D) = S_{x,\theta}^{-1} \cdot \tilde{G}(D) \cdot \tilde{G}^*(D^{-*}) \]  

(A.97)

where \( \tilde{G}(D) = G_x^{-1}(D) \).

If \( A(D) \) is any causal and monic sequence, then \( 1 - A(D) \) is a strictly causal sequence that may be used as a prediction filter, and the prediction error sequence \( E(D) \) is given by

\[ E(D) = X(D) - X(D) \cdot [1 - A(D)] = X(D) \cdot A(D) \]  

(A.98)

The autocorrelation function of the prediction error sequence is

\[ R_{ee}(D) = R_{xx}(D) \cdot A(D) \cdot A^*(D^{-*}) = \frac{A(D) \cdot A^*(D^{-*})}{S_{x,\theta}^{-1} \cdot \tilde{G}(D) \cdot \tilde{G}^*(D^{-*})}, \]  

(A.99)
so its average energy satisfies $E_x = S_0 \cdot \| (1/g) \ast a \|^2 \geq S_0$ (since $A(D)/G(D)$ is monic), with equality if and only if $A(D)$ is chosen as the whitening filter $A(D) = G(D)$. The process $X(D)\cdot G(D) = \frac{X(D)}{G_x(D)}$ is often called the innovations of the process $X(D)$, which has mean square value $S_{x,0} = 1/S_0$. Thus, $S_{x,0}$ of the direct spectral factorization is the mean-square value of the innovations process or equivalent of the MMSE in linear prediction. $X(D)$ can be viewed as being generated by inputting a white innovations process $V(D) = G(D) \cdot X(D)$ with mean square value $S_{x,0}$ into a filter $G_x(D)$ so that $X(D) = G_x(D) \cdot V(D)$.

The inverse’s factorization and $G(D)$’s resultant linear-prediction-filter interpretation helps develop Section 3.6’s interesting MMSE-DFE interpretation where the MS-WMF output sequence $D$-Transform replaces $X(D)$.

### A.3.6 Cholesky Factorization - Finite-Length Spectral Factorization

Cholesky Factorization is spectral factorization’s finite-length equivalent for an $N$-dimensional symbol/packet. There are really two equivalent Cholesky Factorizations, both of which converge to the infinite-length spectral factorization when the process is stationary with successive time-indexed dimensions as $N \to \infty$.

#### A.3.6.1 Cholesky Form 1 - Forward Prediction

Cholesky factorization of a positive-definite (nonsingular)\(^{14}\) $N \times N$ matrix $R_{xx}(N)$ produces a unique upper triangular monic (ones along the diagonal) matrix $G_x(N)$ and a unique diagonal positive-definite diagonal matrix $S_x(N)$ of Cholesky factors such that\(^{15}\)

$$R_{xx}(N) = G_x(N) \cdot S_x(N) \cdot G_x^*(N) \cdot . \quad (A.100)$$

The matrix $R_{xx}(N)$ is often an autocorrelation matrix for $N$ samples of some random vector process $x_k$ with ordering

$$X_N = \begin{bmatrix} x_{N-1} \\ \vdots \\ x_0 \end{bmatrix} \quad (A.101)$$

A corresponding order of $G_x(N)$’s and $S_x(N)$’s elements is then

$$G_x(N) = \begin{bmatrix} g_{N-1} \\ g_{N-2} \\ \vdots \\ g_0 \end{bmatrix} \quad \text{and} \quad S_x(N) = \begin{bmatrix} s_{N-1} & 0 & \cdots & 0 \\ 0 & s_{N-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_0 \end{bmatrix} \quad (A.102)$$

Since $G_x(N)$ is monic, it is convenient to write $g_t = [0_{N-1} \ast 1 \ \tilde{g}_1]^T$, \quad (A.103)

where $0_j$ in general is a column vector with $j$ zeros in it, and $\tilde{g}_0 = 0$ or $g_0 = 1$. The determinant of $R_N$ is

$$S_{x,0} = |R_{xx}(N)| = \prod_{n=0}^{N-1} s_n \quad (A.104)$$

(or $\ln S_{x,0} = \ln |R_{xx}(N)|$ in the limit as $N \to \infty$). A convenient recursive description of $R_{xx}(N)$’s components is, with $r_n = E \left[ |x_n|^2 \right]$ and $r_{N-1} = E \left[ X_n X_N^* \right]$;

$$R_{xx}(N) = \begin{bmatrix} r_N & r_{N-1}^* \\ r_{N-1} & R_{xx}(N-1) \end{bmatrix} \quad (A.105)$$

\(^{14}\)This is the equivalent of $\ln |R_{xx}(N)| < \infty$ where $|R_{xx}(N)|$ is the determinant of $R_{xx}(N)$. Thus the determinant cannot be infinite nor can it be zero, eliminating singularity. The determinant is the product of the eigenvalues of $|R_{xx}|$ so the ln transforms that product into a sum of the log eigenvalues - analogous to the integral in the PWC summing the log transform values. The Fourier transform values are the eigenvalues at each frequency for infinite time extent.

\(^{15}\)Dots are used in this subsection to help notation, even though they here may correspond to matrix multiplication - it just makes it easier to read here.
The submatrix $R_{xx}(N - 1)$ also has a Cholesky Factorization

$$R_{xx}(N - 1) = G_x(N - 1) \cdot S_x(N - 1) \cdot G_x^*(N - 1) \quad ,$$

(A.106)

which because of the 0 entries in the triangular and diagonal matrices shows the recursion inherent in Cholesky decomposition; the $G_x(N - 1)$ matrix is the lower right $(N - 1) \times (N - 1)$ submatrix of $G_x(N)$, which is also upper triangular. Thus, the corresponding recursion description for $G_x$ is

$$G_x(N) = \begin{bmatrix} 1 & \tilde{g}_{N-1}^* \\ 0_{N-1} & G_x(N - 1) \end{bmatrix} \quad ,$$

(A.107)

so then

$$\begin{bmatrix} r_N & r_{N-1}^* \\ r_{N-1} & R_{xx}(N - 1) \end{bmatrix} = \begin{bmatrix} 1 & \tilde{g}_{N-1}^* \\ 0_{N-1} & G_x(N - 1) \end{bmatrix} \begin{bmatrix} s_{N-1} & 0 \\ 0 & S_x(N - 1) \end{bmatrix} \begin{bmatrix} 1 & 0_{N-1}^* \\ \tilde{g}_{N-1}^* & G_x^*(N - 1) \end{bmatrix}$$

(A.108)

Equation (A.108) then admits by observation these recursions:

$$r_{N-1}^* = \tilde{g}_{N-1}^* \cdot S_x(N - 1) \cdot G_x^*(N - 1) \quad ,$$

(A.109)

or equivalently to compute $\tilde{g}_{N-1}^*$ in terms of previously known quantities

$$\tilde{g}_{N-1}^* = r_{N-1}^* \cdot G_x^*(N - 1) \cdot S_x^{-1}(N - 1) \quad ,$$

(A.110)

and

$$s_{N-1} = r_{N-1} - \tilde{g}_{N-1}^* \cdot S_x(N - 1) \cdot \tilde{g}_{N-1}^* \quad .$$

(A.111)

The inverse of $R_{xx}(N)$ has a Cholesky factorization

$$R_{xx}^{-1}(N) = G_x^{-1}(N) \cdot S_x^{-1}(N) \cdot G_x^{-1}(N) \quad ,$$

(A.112)

where $G_x^{-1}(N)$ is also upper triangular and monic with ordering

$$G_x^{-1}(N) = \begin{bmatrix} \tilde{g}_{N-1}^* \\ \tilde{g}_{N-2}^* \\ \vdots \\ \tilde{g}_0 \end{bmatrix} = \begin{bmatrix} 1 & \tilde{g}_{N-1} \cdot G_x^{-1}(N - 1) \\ 0 & G_x^{-1}(N - 1) \end{bmatrix} \quad ,$$

(A.113)

where the use of $\tilde{g}$ from $G_x(N)$’s Cholesky Factorization follows easily through multiplication of $G_x(N) \cdot G_x^{-1}(N) = I$. Also, because it is monic,

$$\tilde{g}_i = \begin{bmatrix} 0_{N-1-i}^* & 1 \tilde{g}_i \end{bmatrix} = \begin{bmatrix} 0_{N-1-i}^* & 1 \tilde{g}_{N-1-i} \cdot G_x^{-1}(N - 1 - i) \end{bmatrix} \quad .$$

(A.114)

### A.3.6.1.1 Cholesky and finite-length linear prediction

Cholesky factorization also derives from a linear-prediction interpretation. The innovations, $V_N$, of the $N$ samples of $X_N$ are defined by

$$X_N = G_x(N) \cdot V_N \quad ,$$

(A.115)

where $E[V_N V_N^*] = S_N$, and the individual innovations are independent (or uncorrelated if not Gaussian), $E[v_i \cdot v_j^*] = S_x(i) \cdot \delta_{ij}$. Then

$$V_N = \begin{bmatrix} v_{N-1} \\ \vdots \\ v_0 \end{bmatrix} \quad .$$

(A.116)
Also,
\[ V_N = G_x^{-1}(N) \cdot X_N \quad (A.117) \]

The cross-correlation between \( X_N \) and \( V_N \) is
\[ R_{yx} = S_x(N) \cdot G_x^*(N) \quad (A.118) \]

which is lower triangular. Thus,
\[ E \left[ v_k \cdot x_{n-k}^* \right] = 0 \quad \forall \ i \geq 1 \quad (A.119) \]

Since \( X_{N-1} = G_x(N-1) \cdot V_{N-1} \) shows a reversible mapping from \( V_{N-1} \) to \( X_{N-1} \), then (A.119) relates that the sequence \( v_k \) is a set of growing-order MMSE prediction errors for \( x_k \) in terms of \( x_{k-1} \ldots x_0 \) (i.e., (A.119) is the orthogonality principle for linear prediction). Thus,
\[ v_N = x_N - r_{N-1}^* \cdot R_{xx}^{-1}(N-1) \cdot X_{N-1} \quad (A.120) \]

since \( r_{N-1} \) is the cross-correlation between \( x_N \) and \( X_{N-1}^* \) in (A.105). The top row of Equation ((A.115) also says

\[
\begin{align*}
  x_N &= v_N + r_{N-1}^* \cdot R_{xx}^{-1}(N-1) \cdot X_{N-1} \\
  &= v_N + r_{N-1}^* \cdot G_x^*(N-1) \cdot S_x^{-1}(N-1) \cdot G_x^{-1}(N-1) \cdot X_{N-1} \\
  &= g_{N-1} \cdot V_N,
\end{align*}
\]

confirming that
\[ g_{N-1} = \begin{bmatrix} 1 & r_{N-1}^* \cdot \tilde{g}_{N-1}^* \cdot S_x^{-1}(N-1) \end{bmatrix} \quad (A.125) \]

Then, from (A.117), (A.120), and (A.123),
\[ \tilde{g}_{N-1} = \begin{bmatrix} 1 & -\tilde{g}_{N-1} \cdot G_x(N-1) \end{bmatrix} \quad (A.126) \]

Finally, the mean-square error recursion is
\[
\begin{align*}
  s_N &= E \left[ v_{N-1} \cdot v_{N-1}^* \right] \\
  &= E \left[ x_{N-1} \cdot x_{N-1}^* \right] - r_{N-1}^* \cdot R_{xx}^{-1}(N-1) \cdot r_{N-1} \\
  &= r_N - \tilde{g}_{N-1} \cdot S_x(N-1) \cdot \tilde{g}_{N-1}^* \\
  &= r_N - \tilde{g}_{N-1}^* \cdot S_x(N-1) \cdot \tilde{g}_{N-1}.
\end{align*}
\]

**A.3.6.1.2 Forward [Upper Triangular] Cholesky Algorithm:**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>( \hat{g}<em>n = r</em>{n-1}^* \cdot G_x^{-1}(n-1) \cdot S_x^{-1}(n-1) ).</td>
</tr>
<tr>
<td>b.</td>
<td>( G_x(n) = \begin{bmatrix} 1 &amp; \hat{g}_{n-1} \ 0 &amp; G_x(n-1) \end{bmatrix} ).</td>
</tr>
<tr>
<td>c.</td>
<td>( G_x^{-1}(n) = \begin{bmatrix} 1 &amp; -\hat{g}_{n-1} \cdot G_x^{-1}(n-1) \ 0 &amp; G_x^{-1}(n-1) \end{bmatrix} ).</td>
</tr>
<tr>
<td>d.</td>
<td>( S_x(n) = r_n - \hat{g}<em>{n-1} \cdot S_x(n-1) \cdot \hat{g}</em>{n-1}^* ).</td>
</tr>
</tbody>
</table>
A singular $R_N$ means that $s_n = 0$ for at least one index $n = i$, which is equivalent to $v_i = 0$. This means that $x_i$ can be exactly predicted from the samples $x_{i-1} \ldots x_0$ or equivalently can be exactly constructed from $v_{i-1} \ldots v_0$. Such a singular process has $\ln |R_N| = 0$ and would not as $N \to \infty$ satisfy the PWC. In the singular case, Cholesky factorization is not unique. Chapter 5 introduces a generalized Cholesky factorization for singular situations that essentially corresponds to doing Cholesky factorization for the nonsingular process, and then generating the deterministic parts that are singular and depend entirely on the nonsingular parts from those nonsingular parts. This will be found equivalent there to independent sampling of each remaining nonsingular processes.

### A.3.6.2 Backward [Lower Triangular] Cholesky Algorithm

Backward Cholesky essentially corresponds to time-order reversal for the finite group of $N$ samples (for infinite-length sequences, this corresponds to $G_x(D^{-*})$). Time reversal sets $x_N \leftarrow J_N x_N$ where $J_N$ is the $N \times N$ matrix with ones on the anti-diagonal and zeros everywhere else. Backward prediction is of $x_0$ from $x_{k=1, \ldots, n-1}$. Further, $J_N^* = J_N$, and $J_N^2 = I$. For this time reversal, the autocorrelation matrices follow

$$
\hat{R}_{xx}(N) \leftarrow J_N \cdot R_{xx}(N) \cdot J_N
$$

(A.130)

So the operation in Equation (A.130) is the autocorrelation matrix corresponding to the time reversal of $x_k$’s components. This operation reversal basically “flips” the matrix about it’s anti-diagonal\(^{16}\). For a REAL Toeplitz matrix (stationary sequence), this flipping does not change the matrix; however for a complex Toeplitz matrix, the new matrix is the conjugate of the original matrix. Further, the operation $J_N \cdot G_x(N) \cdot J_N$ converts $G_x(N)$ from upper triangular to lower triangular, with the ones down the diagonal (monic) retained. The operation

$$
J_N \cdot R_{xx}(N) \cdot J_N = \begin{bmatrix} J_N \cdot G_x(N) \cdot J_N \end{bmatrix} \begin{bmatrix} J_N \cdot S_x(N) \cdot J_N \end{bmatrix} \begin{bmatrix} J_N \cdot G_x^*(N) \cdot J_N \end{bmatrix}
$$

(A.131)

which is the desired lower-diagonal-upper or “Backward-Cholesky” factorization. Thus, the backward algorithm can start with the forward algorithm, and then just use the “tilded” quantities defined in (A.131) as the backward Cholesky factorization (including $\hat{G}_x^{-1}(N) \rightarrow J_N \cdot G_x^{-1}(N) \cdot J_N$).

### A.3.6.3 Infinite-length convergence

Extension to infinite-length stationary sequences takes the limit as $N \to \infty$ in either forward or backward Cholesky factorization. In this case, the matrix sequence $R_{xx}(N)$ (and therefore $G_x(N)$ and $S_x(N)$) must be nonsingular to satisfy the Paley-Weiner Criterion. The equivalence to spectral factorization is evident from both the finite-length and infinite length linear prediction discussions.

From a stationary perspective, forward and backward prediction are the same except that the backward predictor reverses the time index (and conjugates when complex) of the forward predictor’s coefficients (and vice versa). This is the equivalent (with re-index of time 0) of $G_x^*(D^{-*})$ being the reverse of $G_x(D)$ with conjugate coefficients.

Thus, the inverse autocorrelation function factors as

$$
R_{xx}^{-1}(D) = S_{x,0}^{-1} \cdot G_x^{-1}(D) \cdot G_x^{-*}(D^{-*})
$$

(A.132)

where $G_x^{-1}(D)$ is the forward prediction polynomial (and its time reverse specified by $G_x^*(D^{-*})$ is the backward prediction polynomial). The series $\{R_{xx}(n)\}_{n=1: \infty}$ formed from the coefficients of $R_{xx}(D)$ creates a series of linear predictors $\{G_x(N)\}_{N=1: \infty}$ with D-transforms $G_{x,N}(D)$. In the limit as $N \to \infty$ for a stationary nonsingular series,

$$
\lim_{N \to \infty} G_{x,N}(D) = G_x(D)
$$

(A.133)

Similarly,

$$
\lim_{N \to \infty} G_{x,N}^*(D) = G_x^*(D^{-*})
$$

(A.134)

\(^{16}\)“Flip” is like transpose but around the anti-diagonal.
As \( N \to \infty \), the prediction-error variances \( S_{N-1} \), will tend to a constant, namely \( S_{x,0} \). Finally, defining the geometric-average determinants as \( S_{x,0}(N) \triangleq |R_{xx}|^{1/N} \) and \( S_{x,0}^{-1}(N) = |R_{xx}^{-1}|^{1/N} \)

\[
\begin{align*}
\lim_{N \to \infty} S_{x,0}(N) &= S_{x,0} = e^{\frac{1}{2} \int_{-\pi}^{\pi} \ln(R_{xx}(e^{-j\omega})) \, d\omega} \\
\lim_{N \to \infty} S_{x,0}^{-1}(N) &= S_{x,0}^{-1} = e^{-\frac{1}{2} \int_{-\pi}^{\pi} \ln(R_{xx}(e^{-j\omega})) \, d\omega}.
\end{align*}
\]

(A.135) (A.136)

The convergence to these limits implies that the series of filters converges or that the bottom row (last column) of the Cholesky factors tends to a constant repeated row/column. Chapter 5 has examples of this effect.

Interestingly, Chapter 5’s Generalized Cholesky Factorization of a singular process exists only for finite lengths. Using the modifications to this appendix section’s Cholesky Factorization with “resampling” in each disjoint PWC-satisfying frequency band, it becomes obvious why such the original combined random process cannot converge to a constant limit. So only nonsingular processes have (infinite-length) spectral factorization. A singular process’ infinite-length factorization first separates that process into a sum of subprocesses, each of which is resampled at a new sampling rate that satisfies the PWC over each of the frequency bands associated with these processes. This is equivalent to the multiple Cholesky’s for each of the process’ nonsingular components at infinite length. For more, see Chapter 5.

### A.4 MIMO Spectral Factorization

#### A.4.1 Vector Transforms

The D-Transform’s extension to an \( L \)-dimensional vector sequence \( x_k \) is

\[
X(D) = \sum_{k=\infty}^{\infty} x_k \cdot D^k ,
\]

(A.137)

for all scalar complex \( D \) in the convergence region, \( D \in \mathcal{D}_x \). The inverse D-Transform is given by

\[
x_k = \frac{1}{2\pi j} \int_{D \in \mathcal{D}_x} X(D) \cdot D^{-k} \cdot dD.
\]

When the unit circle is in the convergence region \( |D| = 1 \in \mathcal{D}_x \), then a vector Fourier Transform exists and is

\[
X(e^{-j\omega}) = \sum_{k=\infty}^{\infty} x_k \cdot e^{-j\omega k} .
\]

(A.138)

Sufficient conditions for convergence of the Fourier Transform generalize to:

\[
\|x\|_1 = \sum_{k=\infty}^{\infty} |x_k| < \infty ,
\]

(A.139)

or

\[
\|x\|_2 = \sum_{k=\infty}^{\infty} \|x_k\| < \infty .
\]

(A.140)

Vector D-Transforms with poles on the unit circle are handled in the same limiting sense of approaching the unit circle arbitrarily closely from outside, which essentially allows generalized functions to be used in the frequency domain. The vector sequence’s Fourier Transform then also exists. The transform’s similarity to its inverse then allows equivalently that the inverse Fourier Transform:

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{-j\omega})| \cdot d\omega < \infty ,
\]

(A.141)

or if

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} \|X(e^{-j\omega})\|^2 \cdot d\omega < \infty .
\]

(A.142)

---

17 There is not a separate \( D \) value for each dimension of the vector \( x_k \) - just one scalar \( D \) value for all vector elements.
Rather than repeat all for a continuous-time vector function’s Laplace Transform, the Vector Laplace Transform for continuous-time vector process $x(t)$ is

$$X(s) = \int_{-\infty}^{\infty} x(t) \cdot e^{-st} dt,$$  \hspace{1cm} (A.143)

with convergence region $s \in S_x$. When the imaginary axis is in the convergence region $S_x$, the Fourier Transform is

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt.$$  \hspace{1cm} (A.144)

Convergence conditions on the vector continuous-time process are then

$$\int_{-\infty}^{\infty} |x(t)| \cdot dt < \infty,$$ \hspace{1cm} (A.145)

or equivalently

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| \cdot d\omega < \infty.$$  \hspace{1cm} (A.146)

This means that the vector function $x(t)$ belongs to the space of continuous functions $L^1$. Another sufficient condition is that vector function $x(t)$ belongs to $L^2$ or has finite energy according to

$$\int_{-\infty}^{\infty} \|x(t)\|^2 \cdot dt < \infty,$$ \hspace{1cm} (A.147)

or equivalently

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \|X(\omega)\|^2 \cdot d\omega < \infty.$$  \hspace{1cm} (A.148)

Similarly “generalized” functions complete the capability to handle continuous Fourier Transforms that are “on the stability margin” where values of $s$ in a region of convergence arbitrarily close to the $j\omega$ axis (left of, but not on this axis). These will converge so the criteria are considered satisfied in a limiting or generalized sense.

Similarly a matrix time series can also have a matrix $D$-Transform (Laplace Transform) that contains the $D$-Transform (Laplace Transform) of each and every time-series element. There is only one common $D$ variable for all these matrix elements. The convergence region is the intersection of the elements’ convergence regions. When that overall convergence region includes the unit circle (complex axis), the Fourier Transform also exists and the inversion formula are all the obvious extensions. For square matrices, Fourier-Transform sufficiency criteria simplifies to

$$\sum_{k=-\infty}^{\infty} |R_k| < \infty$$  \hspace{1cm} (A.149)

where the summed entities are the matrix-time-series coefficients’ determinants, so equivalently Definition A.2.3’s summed individual matrix norms. Basically the sum of the matrix-time-series coefficients’ norms for transforms (or integral for Laplace) has to be finite as a sufficient condition for any valid norm’s existence. So in this case the chosen norm is the trace, but that is equivalent to summing determinants in terms of existence. As Definition A.2.3, the square norm has essentially two equivalent forms for MMSE estimation.

The bilinear transform is unchanged with respect to scalar processes.

### A.4.2 Autocorrelation and Power Spectra for vector sequences

This subsection generalizes scalar $D$-Transforms factorization to autocorrelation functions and associated power spectra for stationary and wide-sense stationary vector processes.

**Definition A.4.1 (Autocorrelation and Power Spectrum for Vector Sequences)** If $x_k$ is any stationary complex vector sequence, its autocorrelation matrix is $r_{xx,j} =$
\[ R_{XX}(D) \triangleq E \left[ X(D) \cdot X^*(D^{-1}) \right] . \]  
(A.150)

By stationarity, \( r_{XX,j} = r_{XX,-j} \) and \( R_{XX}(D) = R_{XX}(D^*) \). The **power spectrum matrix** of a stationary vector sequence is the Fourier transform of its autocorrelation matrix function, which is

\[ R_{XX}(e^{-j\omega}) = R_{XX}(D)|_{D = e^{-j\omega}} , \quad -\pi < \omega \leq \pi , \]  
(A.151)

which is positive semi-definite\(^{19} \) for all \( \omega \), namely

\[ x^* \cdot R(e^{-j\omega}) \cdot x \geq 0 , \quad \forall x \in \mathbb{C}^{Lx} \text{ and } \forall \omega \in (-\pi, \pi) . \]  
(A.152)

Consequently the determinant and the norm are both non-negative:

\[ |R_{XX}(e^{-j\omega})| \geq 0 , \quad \forall \omega \in (-\pi, \pi) \]  
(A.153)

\[ \text{trace}(R_{XX}) \geq 0 , \quad \forall \omega \in (-\pi, \pi) . \]  
(A.154)

Conversely, any positive semi-definite function \( R_{XX}(e^{-j\omega}) \) that is real and nonnegative definite over the interval \( \{-\pi < \omega \leq \pi\} \) is a power-spectrum matrix and has corresponding matrix autocorrelation function \( R_{XX}(D) = R_{XX}(D^{-1}) \), or equivalently \( r_{XX,k} = r_{XX,-k}^\ast \).

The quantity \( E \left[ |x_k|^2 \right] \) is \( \mathcal{E}_x \), and can be determined from either the autocorrelation matrix or the power spectrum matrix as follows:

\[ \mathcal{E}_x = E \left[ |x_k|^2 \right] = \text{trace} \{ r_{XX,0} \} \]  
(A.155)

\[ = \frac{1}{2\pi} \text{trace} \left\{ \int_{-\pi}^\pi R_{XX}(e^{-j\omega})d\omega \right\} . \]  
(A.157)

If a matrix sequence \( \mathbf{R}_k \) in question is deterministic such as might be formed by inverse Fourier Transform of the filter \( R(e^{-j\omega}) = \mathbf{H}(e^{-j\omega}) \cdot \mathbf{H}^\ast(e^{j\omega}) \), then the averages are not necessary above. The power spectra matrix is then the positive semi-definite matrix \( R(e^{-j\omega}) \) for discrete time. These Fourier Transforms' magnitudes can be thought of also as power spectra matrices, and the corresponding inverse transforms as autocorrelation functions in this text.

**Definition A.4.2 (Continuous-Time Vector Autocorrelation & Power Spectra)** If \( \mathbf{x}(t) \) is any stationary complex vector function, its **autocorrelation matrix** is \( R_{\mathbf{x},\mathbf{x}}(t) = E[\mathbf{x}_c(u) \cdot \mathbf{x}_c^\ast(u-t)] \) with Laplace Transform \( R_{\mathbf{x},\mathbf{x}}(s) \); symbolically\(^{20} \)

\[ R_{\mathbf{x},\mathbf{x}}(s) \triangleq E \left[ \mathbf{X}(s) \cdot \mathbf{X}^*(-s^*) \right] . \]  
(A.158)

By stationarity, \( r_{\mathbf{x},\mathbf{x}}(t) = r_{\mathbf{x},\mathbf{x}}^\ast(-t) \) and \( R_{\mathbf{x},\mathbf{x}}(s) = R_{\mathbf{x},\mathbf{x}}^\ast(-s^*) \). The **power spectrum matrix** of a stationary continuous-time vector process is the Fourier transform of its autocorrelation matrix function, which is

\[ R_{\mathbf{x},\mathbf{x}}(\omega_c) = R_{\mathbf{x},\mathbf{x}}(s)|_{s = j\omega_c} , \quad -\infty < \omega_c \leq \infty , \]  
(A.159)

and which is real and nonnegative definite for all \( \omega_c \), namely

\[ |R_{\mathbf{x},\mathbf{x}}(\omega_c)| \geq 0 , \quad -\infty < \omega_c < \infty . \]  
(A.160)

Conversely, any function \( R(\omega_c) \) that is real and nonnegative definite over the interval \( \{-\infty < \omega_c \leq \infty\} \) is a power-spectrum matrix and has autocorrelation matrix function satisfying \( R(s) = R^\ast(-s^*) \).

\(^{18}\)The expression \( R_{XX}(D) \triangleq E \left[ X(D)X^*(D^{-1}) \right] \) is used in a symbolic sense, since the terms of \( X(D)X^*(D^{-1}) \) are of the form \( \sum_{k} a_k x_{k-j}^\ast \), implying the additional operation \( \lim_{N \to \infty} [1/(2N + 1)] \sum_{-N \leq k \leq N} \) on the sum in such terms.

\(^{19}\)And positive definite if strictly greater than 0.

\(^{20}\)The expression \( R_{\mathbf{x},\mathbf{x}}(s) \triangleq E \left[ \mathbf{X}(s)\mathbf{X}^*(-s^*) \right] \) is used in a symbolic sense, since the terms of \( \mathbf{X}(s)\mathbf{X}^*(-s^*) \) are of the form \( \int_{-T}^T E[\mathbf{x}_c(u)\mathbf{x}_c^\ast(u-t)]du \), implying the additional operation \( \lim_{T \to \infty} [1/(2T)] \int_{-T}^T \) on the integral in such terms.

377
The quantity $E[|x_c(t)|^2]$ is the power $P_x$ and can be determined from either the autocorrelation function or the power spectrum as follows:

$$P_x = E[|x_c(t)|^2] = \text{trace}\{r_{x_c,x_c}(0)\} = \frac{1}{2\pi}\text{trace}\left\{\int_{-\infty}^{\infty} R_{x_c,x_c}(\omega) \cdot d\omega\right\}. \quad (A.161)$$

If a matrix function $r(t)$ in question is deterministic such as might be formed by inverse Fourier Transform of the filter $R(j\omega) = H(j\omega) \cdot H^*(j\omega)$, then the averages are not necessary above. The power spectra matrix is then the positive semi-definite matrix $R(j\omega)$ for continuous time. These Fourier Transforms' magnitudes can be thought of also as power spectra matrices, and the corresponding inverse transforms as autocorrelation functions in this text.

### A.4.3 Factorizability for MIMO Processes

Vector sequences have two types of dimensions, discrete (or continuous) time with infinite index span $k \in \mathbb{Z}$ (or $t \in \mathcal{R}$) and space with finite index $l = 1, \ldots, L$. The infinite time index corresponds to a continuous transform variable $D$ for discrete time and $s$ for continuous time, while the finite space index leads to Subsection A.2’s vector and matrix generalizations. Sometimes this text will group time-domain samples into finite symbols with $k \in [0, \ldots, N - 1]$ and the corresponding infinite transforms ($D$ and $s$) sampled (on the unit circle or imaginary axis respectively) also to a discrete frequency index $n$. In this case, the matrix indexing, $k$ or $n$, corresponds to a SISO system, and $L = 1$. Chapter 3 uses discretization in Section 3.8’s FIR Equalizers to compute all equalizers (linear or decision-feedback). Chapters 4 and 5 show how each sampled frequency may be viewed in MIMO as a separate MIMO system, so there will be two levels of matrix application: (1) time-frequency and (2) space. Those designs are finite-length implementable subsets of the infinite-length (time-frequency) MIMO processes appearing in Section 3.10.

Matrix “factorization” has some definition flexibility; this text will combine Cholesky Factorization and infinite-length factorization to create a MIMO factorization theory that this text uses, particularly Section 3.10. One contributor to non-unique factorization is the order of the spatial dimensions. Each order produces a different factorization, but they all clearly from a high-level intuitive perspective must be the same. The spatial indexing is arbitrary from that viewpoint. However, this order leads to different MMSE’s on each of the spatial dimensions, which favors certain dimensions. Such favored dimensions will have application in Chapter 5’s multi-user MMSE theory.

At time of writing, the author is aware of no treatment anywhere that approaches this appendix, including the classic text by [2], which calls this area intractable. However, the short paper by [3] was helpful in developing the methods of this subsection, which the author hopes readers find useful for the general MIMO theory of equalization and MMSE in general.

The **spectral factorization** of a discrete-time vector autocorrelation function’s D-Transform is:

**Definition A.4.3 (Factorizability for Vector Sequences)** An autocorrelation function $R_{xx}(D)$, or equivalently any positive semi-definite real function $R_{xx}(e^{-j\omega})$ at all $\omega \in (-\pi, \pi)$ and corresponding inverse transform $r_{xx,k} = r_{x,c,-k}$, will be called **factorizable** if it can be written in the form

$$R_{xx}(D) = G_{xx}(D) \cdot S_{x,0} \cdot G_{xx}^*(D^{-*}), \quad (A.162)$$

where $S_{x,0}$ is a constant positive-real diagonal matrix and where $G_{xx}(D)$ is a canonical matrix filter response. The canonical matrix-filter factor $G_{xx}(D)$ for a factorizable $R_{xx}(D)$ must also be causal $G_{xx}(D) = \sum_{k=0}^{\infty} G_{x,k}$ or $G_{x,k} = 0 \forall k < 0$, have **upper triangular** $G_x(0)$ that is also monic $\text{Diag}\{G_x(0)\} = I$, and **minimum phase** (all of its poles and zeros are outside or on the unit circle). If $G_{xx}(D)$ is canonical, then $G_{xx}^*(D^{-*})$ is ant canonical: i.e., anticausal, monic lower triangular $G_{x,vec}(\infty)$, and maximum-phase (all poles and zeros inside or on the unit circle).

The convergence region for a factorizable $R_{xx}(D)$ clearly includes the unit circle, as do the regions for both $G_{xx}(D)$ and $G_{xx}^*(D^{-*})$. If canonical $G_{xx}(D) = G_o$, a constant, then $|G_{xx}(D)| = 1$ and $\text{trace}\{G_{xx}(D)\} = L_x$, which are clearly their minimum norm values for the more general canonical
correlation functions is a ratio of finite-degree polynomials in all entries is factorizable, then the corresponding $R_x x_c(s) = R_xx (\frac{1-s}{1+s})$ is also factorizable into

$$R_x x_c(s) = G_x c(s) \cdot S_{x} \cdot G_x c(-s^*) .$$

**Definition A.4.4 (Factorizability for Continuous Vector Functions)** An autocorrelation matrix $R_x x_c(s)$, or equivalently any positive semi-definite $R_x x_c(\omega)$ at all $\omega \in (-\infty, \infty)$ and corresponding inverse transform $r_x x_c(t) = r_x x_c(-t)$, will be called **factorizable** if it can be written in the form

$$R_x x_c(s) = G_x c(s) \cdot S_{x} \cdot G_x c(-s^*),$$

where $S_{x} \cdot$ is a finite positive real diagonal matrix and $G_x c(s)$ is a canonical filter response. The canonical matrix-filter response $G_x c(s)$ for a factorizable $R_x x_c(s)$ must also be **causal** ($g_{x c}(t) = 0$ for $t < 0$), have **monic** upper triangular $G_x c(s) = 0$ so that (Diag $\{G_x c(0)\} = I$), and **minimum-phase** (all of its poles and zeros are in the left half plane). If $G_x c(s)$ is canonical, then $G_x c(-s^*)$ is **anticanonical**; i.e., anticausal, monic, lower triangular $G_x c(0)$, and maximum-phase (all poles and zeros inside in the right half plane). Further, (note use of bold and plain fonts on “$\times$”)

$$S_{x,0} \triangleq |S_{x}| .$$

The region of convergence for factorizable $R_x x_c(s)$ clearly includes the $\omega_c$ axis, as do the regions for both $G_x c$ and $G_x c(-s^*)$. If $G_x c(s)$ is a canonical response, then $\|g_{x c}\|^2 \triangleq \text{trace} \left\{ \int_{-\infty}^{\infty} |g_{x c}(t)|^2 \right\} \geq L_x$, with equality if $G_x c(s) = I$, since $G_x c(s)$ is monic. Further, the inverse also factorizes similarly into

$$R_x^{-1} x_c(s) = G_x^{-1} c(s) \cdot S_{x,0} \cdot G_x^{-1} c(-s^*) .$$

The determinant of $|R_x x_c(s)|$ will capture all poles and zeros in any and all terms of the factorization (and hidden cancellations will not be important in practice).

### A.4.4 Finite-Degree MIMO polynomial factorization

Any matrix autocorrelation function that is a ratio of finite-degree polynomials in all entries is factorizable. This subsection provides a direct calculation of the factors by expanding on a method suggested by [3]. The square autocorrelation matrix of polynomials is

$$R_x x(D) = \begin{bmatrix} R_{x x, l_x, l_x}(D) & \ldots & R_{x x, l_x, 1}(D) \\ \vdots & \ddots & \vdots \\ R_{x x, 1, l_x}(D) & \ldots & R_{x x, 1, 1}(D) \end{bmatrix}$$

[21]Hidden cancellations will not be important in practice; In the non-polynomial-ratio case, the number of zeros may be infinite.
Each element of \( R_{\mathbf{x}}(D) \), here simplified to \( R(D) \) has the form

\[
R_{i,j}(D) = \frac{K_{i,j} \cdot \left( \prod_{q=1}^{Q} (1 - z_q D) \right) \cdot \left( \prod_{q=1}^{Q} (1 - z_q D^{-1}) \right)}{Z_{i,j} \cdot \left( \prod_{m=1}^{M_p} (1 - p_m D) \right) \cdot \left( \prod_{m=1}^{M_p} (1 - p_m D^{-1}) \right)},
\]

(A.169)

with \( z_q, q = 1, \ldots, Q \) with \( |z_q| < 1 \) are the max-phase zeros inside the unit circle and \( 1/z_q \) are the minimum-phase zeros outside the unit circle. Complex zeros occur in conjugate pairs. Zeros on the unit circle will also occur in pairs, where one in each unit-magnitude pair is minimum phase and the other is maximum phase. The poles \( p_m, m = 1, \ldots, M_p \) with \( |p_m| < 1 \) are the max-phase poles inside the unit circle, and \( 1/p_m \) are the minimum-phase poles outside the unit circle. Complex poles occur in conjugate pairs. Poles on the unit circle will also occur in pairs, where one in each unit-magnitude pair is minimum phase and the other is maximum phase. \( Q_2 < \infty \) is the maximum degree of the min- (or max-) phase numerator component, while \( M_p < \infty \) is the maximum degree of the min- (or max-) phase denominator component. These maximums can be taken over all elements in the matrix and simple \( z_i = 0 \) or \( p_i = 0 \) terms used for those terms of lower degree where the pole zero terms drop. The constant \( K_{i,j} \) will be the ratio of the highest-degree non-zero coefficient in \( R_{i,j}(D) \)'s numerator to the corresponding highest-degree non-zero coefficient in its denominator.

Prescripts will be used to denote generations of quantities within the algorithm that follows, as will become evident.

A.4.4.1 STEP ZERO - INITIALIZATION

A initial step computes the least common multiple of all \( R_{i,j}(D) \) elements' pole factors \( \Delta(D) \cdot \Delta^*(D^{-*}) \), and saves them, replacing \( R(D) \) then with this new polynomial of finite maximum degree \( 2Q_2 \). The pole factors return in a final step because they easily separate into minimum and maximum phase and then will multiply the algorithm’s corresponding \( G(D) \) output factors that arise. The factorization then proceeds of the remaining all-zero polynomial \( R(D) \).

\[
aR(D) = \Delta(D) \cdot R(D) \cdot \Delta(D).
\]

(A.170)

A.4.4.2 STEP ONE - Polynomial Cholesky Factorization

Cholesky Factorization applies also to transform polynomials even though presented earlier for constant matrices. The Cholesky factorization of a symmetric positive definite (semi-definite in case of polynomials at any zeros) matrix \( R(D) \) for the \( 2 \times 2 \) case is:

\[
R(D) = \begin{bmatrix}
a(D) & b^*(D^{-*}) \\
0 & g(D)
\end{bmatrix} \cdot \begin{bmatrix}
a^*(D^{-*}) \\
b(D) & g^*(D^{-*})
\end{bmatrix} = \begin{bmatrix}
a(D) \cdot a^*(D^{-*}) + b^*(D^{-*}) \cdot b(D) & b^*(D^{-*}) \cdot g^*(D^{-*}) \\
g(D) \cdot c \cdot b(D) & g(D) \cdot g^*(D^{-*})
\end{bmatrix}.
\]

(A.171)

(A.172)

In the \( 2 \times 2 \) case for \( aR(D) \), Cholesky factors are easily determined by factoring the lower right-side term (half the 11 term) in \( aR(D) \)

\[
g(D) = \sqrt{K_{1,1}} \cdot \prod_{q=1}^{Q_2} (1 - z_{q11} \cdot D).
\]

(A.173)

and then

\[
b(D) = \frac{aR_{1,2}(D)}{g(D)}
\]

(A.174)

and finally

\[
a(D) = \sqrt{aR_{2,2}(D) - b^*(D^{-*}) \cdot b(D)}.
\]

(A.175)

The square root in (A.175) is chosen through this all-zero polynomial’s factorization terms \((1 - z \cdot D)\) with \( |z| \leq 1 \) and their complimentary terms \((1 - z \cdot D^{-1})\). The algorithm retains the former, along with
the square root of a highest-degree-D coefficient. Equation (A.175)’s \( a(D) \) then is the minimum-phase square root.

Once the above \( 2 \times 2 \) recursion’s steps are understood, progress to the \( L \times L \) general version for \( R(D) \) follows as the STEP ONE recursion that takes advantage of the \( G(D) \) “upper triangular matrix” growing in size at each iteration:

\[
R_l(D) = \begin{bmatrix}
    r_{l,l}(D) & r_l^*(D^{-*}) \\
    r_l(D) & R_{l-1}(D)
\end{bmatrix}
\]

\( a_l(D) \) \( b_l^*(D^{-*}) \) \( G_{l-1}(D) \)

\[
= \begin{bmatrix}
    a_l(D) & b_l^*(D^{-*}) \\
    0 & G_{l-1}(D)
\end{bmatrix} \cdot \begin{bmatrix}
    a_l^*(D^{-*}) & 0 \\
    b_l(D) & G_{l-1}^*(D^{-*})
\end{bmatrix}
\]

(\( A.176 \))

with iteration \( l \) solution for \( l = 2, ..., L \) as\(^{22}\)

\[
b(D) = G_{j-1}^{-1} \cdot r_l(D)
\]

(\( A.179 \))

then minimum-phase square root for \( a(D) \) as

\[
a(D) = \sqrt{r_{l,l}(D) - b_l^*(D^{-*}) \cdot b_l(D)}
\]

(\( A.180 \))

and finally \( G(D) \) is then given in \( (A.178) \). When \( l = L \), a valid square-root, but not necessarily minimum phase, has been found for the original matrix \( R(D) \). The remaining steps then drive to a canonical factor from this square root. Thus,

\[
1G(D) \cdot 1G^*(D^{-*}) = oR(D)
\]

(\( A.181 \))

\subsection{A.4.4.3 STEP TWO: Left Extraction of Poles}

Cholesky factorization will leave the upper left entry of \( G(D) \) as the ratio of the determinants \( |R_L(D)| \) to \( |R_{L-1}(D)| \), and similarly for \( l = L - 1, ..., 2 \). Because of the minimum-phase square root choices made in STEP ONE, the denominators of these determinants will have poles outside the unit circle. The factor \( 1G(D) \) thus may now have poles that will be stable (outside or on the unit circle), which STEP TWO extracts to the left through a diagonal-matrix multiple removal so that

\[
2G(D) = \begin{bmatrix}
    \prod_{m=1}^{\infty} (1-p_{m,j,l,D}) & 0 & \cdots & 0 \\
    0 & \prod_{m=1}^{\infty} (1-p_{m,j,l-1,D}) & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & \cdots & 0 & 1
\end{bmatrix} \cdot 1G_l(D)
\]

(\( A.182 \))

Poles at infinity, \( p_{m,j,j} = 0 \) are simply the factor \( 1/D \) that can be ignored without altering the square root (because an \( D \) factor on the right side for \( G^*(D^{-*}) \) eventually cancels them or equivalently and advance of a stationary sequence does not change its autocorrelation). Pole extraction is not necessary but simplifies the remaining steps. These poles will necessarily cancel with zeros in the final \( G(D) \) because the determinant is \( oR(D) \) an all-zero polynomial. That determinant is maintained in the product of the square root with its paraunitary factor \( oR^{1/2} \cdot oR^*/2(D^{-*}) = oR(D) \).

\subsection{A.4.4.4 STEP THREE: Compute and factor the determinant of \( G \)}

The current square root \( 2G(D) \) may have zeros inside the unit circle and is thus not yet minimum phase. These zeros are found by computing the determinant \( |2G(D)| \) and factoring this scalar polynomial:

\[
|2G(D)| = k_{q2} \cdot \prod_{q=1}^{Q_2} (1 - z_{q2} \cdot D)
\]

(\( A.183 \))

\(^{22}\)The inverse \( G_{l-1}^{-1}(D) \) is determined easily by adding a new row at the top of the value for \( G_{l-1}^{-1}(D) \), and there is essentially only \( 1 \) polynomial divide operation per iteration step.
where \( z_{q2} \) are the zeros and \( k_{q2} \) is a constant equal to the coefficient of lowest power of \( D \) (typically \( D^0 \) at this point) in the original determinant. The zeros inside the unit circle where \(|z_{q2}| > 1\) will need to be removed in the next step, and STEP THREE identifies them.

### A.4.4.5 STEP FOUR: Remove the maximum-phase zeros from \( G(D) \)

Any determinant zero causes \( 2G(D) = 0 \) and thus means that its columns are linearly dependent at this value of \( D = z_{q2} \). There consequently exists a unitary constant transformation \( U_3 \) (with \( U_3 U_3^* = I = U_3^* U_3 \) for each such zero that rotates \( 3G(D) = 2G(D) \cdot U_3 \) to zero any selected column. Such a zeroed column has a common zero factor at \( z_{q2} \) for all elements. Such a transform preserves \( 3G(D) \) as a square root. Usually finding \( U_3 \) is trivial (as a later example shows), but generally it is found from the constant matrix \( 2G(z_{q2})'s \) null space (which is of dimension at least one because this matrix is singular) and placing a normalized null-space basis vector in the last column of \( U_3 \) so that the last column of the product \( 2G(z_{q2}) \cdot U_3 \) is zero, which also means that column has a common zero at \( z_{q2} \) in all its elements.

STEP FOUR extracts this common zero to the right through the multiplication

\[
4G(D) = 2G(D) \cdot U_3. 
\]

The matrix \( \tilde{U}_3(D) \) is paraunitary and will cancel with its opposite phase equivalent in the parallel formation of \( 3G^*(D^-*) \). The zero will be cancelled and replaced by a minimum phase zero at \( 1/z_{q2} \). This fourth step repeats until there are no maximum-phase zeros remaining in \( 4G(D) \). \( 4G(D) \) is clearly a valid minimum-phase square root.

### A.4.4.6 STEP FIVE - remove constants to the center

The term \( 4G(0) \) is not necessarily upper triangular, nor monic. STEP FIVE continues with matrix scaling so that \( 5G(0) = I \) and thus

\[
5G(D) = 4G(D) \cdot 4G^{-1}(0). \tag{A.185}
\]

Then

\[
1R(D) = 5G(D) \cdot 5R_{4G(0) \cdot 4G^*(0)} \tag{A.186}
\]

The middle matrix is not yet diagonal, but \( 5G(0) \) is now monic.

### A.4.4.7 STEP SIX - Constant-Matrix Cholesky Factorization and adjustment

STEP SIX factors the constant middle matrix through normal Cholesky Factorization as

\[
5R = 6G(0) \cdot S \cdot 6G^*(0). \tag{A.187}
\]

and so now

\[
6G(D) \overset{\Delta}{=} 5G(D) \cdot 6G(0). \tag{A.188}
\]

\( 6G(D) \) is a valid square root and canonical for \( 1R(D) \).

### A.4.4.8 STEP SEVEN - Restore the original poles

Step 0 removed poles from \( R(D) \), and this STEP SEVEN now restores them.

\[
G(D) = \frac{1}{\Delta(D)} \cdot 6G(D), \tag{A.189}
\]

and finally the desired canonical factorization is

\[
R(D) = G(D) \cdot S \cdot G^*(D^-*) \tag{A.190}
\]

382
A.4.4.9 Swenson’s Example

Dr. Norm Swenson, a former student and as a visiting Scholar in 2020, suggested the following $L = 2$ MIMO channel for factorization. He encountered it in an optical mixed-mode situation and challenged the author to factor:

The matrix to be factored is:

$$R(D) = \begin{bmatrix} 8D^{-1} + 23 + 8D & 7D^{-1} + 7 - D \\ -D^{-1} + 7 + D & -6D^{-1} + 18 - 6D \end{bmatrix} \quad (A.191)$$

The following matrix commands for the matrix and find its roots, determinant, and STEP ONE Cholesky Factor (there are no poles):

```matlab
>> R11=[-6 18 -6];
>> R22=[8 23 8];
>> R12=[-1 7 7 ];
>> R21=[7 7 -1];
>> roots(R11)
ans = 2.6180 
0.3820
>> z11 = 0.3820;
>> sr11 = sqrt(R11(1)/-z11) = 3.9634
>> roots(R22)
ans = -2.4702
-0.4048
>> z22 = -0.4048
>> sr22 = sqrt(R22(1)/-z22) = 4.4454
>> detR= conv(R22,R11)-conv(R21,R12)
detR = -41 -36 219 -36 -41
>> roots(detR)
-2.8306 
1.7267
0.5791
-0.3533
>> z1= -0.3533
>> z2 = 0.5791
>> sdetR=sqrt(detR(1)/(-z1*z2)) = 14.1559
>> GA22=sdetR*conv([1 -z1],[1 -z2]) = 14.1559 -3.1973 -2.8963
>> GA11=sr11*[-z11 1] = -1.5139 3.9634
>> GA21 = [ 7 7 -1 ];
>> GA12 = [ 0 0 ];
```

So the original matrix can also be written with factorization of its diagonal components as:

$$R(D) = \begin{bmatrix} 4.45^2 \cdot (1 + 0.405D)(1 + 0.405D^{-1}) & 7D^{-1} + 7 - D \\ -D^{-1} + 7 + D & 3.96^2 \cdot (1 - 0.382D)(1 - 0.382D^{-1}) \end{bmatrix} \quad (A.192)$$

and

$$|R(D)| = -41D^{-2} - 36D^{-1} + 219 - 36D - 41D^{2} = 14.16^2 (1-0.579D)(1+0.353D)(1+0.353D^{-1})(1.579D^{-1}) \quad (A.193)$$

The matlab code uses the label GA, GB, etc. instead of prescripts.

So far then, the factoring has produced (since the Cholesky Factor 22 entry is the ratio of the overall square-root determinant to the corresponding square root of 11 entry), extracting the common top-row’s $1/(3.96(1 - 0.382D))$ divide-by-g pole term to the left.

$$G_A(D) = \begin{bmatrix} \frac{1}{3.96(1-0.382D)} & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 14.15(1 + 0.353D)(1 - 0.579D) & 7 + 7D - D^2 \\ 0 & 3.96(-0.382 + D) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D^{-1} \end{bmatrix} \quad (A.194)$$
STEP THREE and STEP FOUR identify the zero at .382 and remove it from last column.

\[
G_A(0.382) = \begin{bmatrix} 12.512 & 9.5279 \\ 0 & 0 \end{bmatrix},
\]

(A.195)

which makes it obvious that the unitary matrix to zero is a simple two-dimensional rotation with a cosine and sine computed as:

\[
\begin{align*}
\text{normA} &= \sqrt{(GA22atz11)^2 + (GA21atz11)^2} = 15.7268 \\
cA &= \frac{GA22atz11}{\text{normA}} = 0.7956 \\
sA &= \frac{GA21atz11}{\text{normA}} = 0.6058 \\
GA21tilde &= GA22*(-sA) + GA21*cA = -3.0070 7.5062 0.9591 \\
\text{roots}(GA21tilde) &= 2.6180 \quad -0.1218 \\
GA22tilde &= GA22*cA + GA21*sA = 15.5031 1.6971 -2.9101 \\
GA12tilde &= GA12*cA + GA11*sA = -0.9172 2.4011 \\
\text{roots}(GA12tilde) &= \text{roots}\left(\begin{bmatrix} -0.9172 & 2.4011 \end{bmatrix}\right) = 2.6181 \\
\text{checks both 11 term and 21 term in last column now have the max-phase zero at 1/2.618, which can now be factored out.}
\end{align*}
\]

\[
\begin{align*}
\text{detGAtilde} &= \text{conv}(GA22tilde,GA11tilde) - \text{conv}(GA21tilde,GA12tilde) = -21.4301 60.9450 -8.2873 -11.4792 \\
\text{roots}(\text{detGAtilde}) &= [2.6180 \quad -0.1218] \\
\text{GB21} &= GA21tilde(1)*-X1*[1 -X2] = 7.8725 0.9591 \\
\text{GB22} &= GA22tilde = 15.5031 1.6971 -2.9101 \\
\text{GB12} &= GA12tilde = -0.9172 2.4011 \\
\text{GB11} &= GA11tilde(1)*-X1 = 3.1532 \\
\text{roots}(\text{GB22}) &= -0.4914 \quad 0.3820 \\
\text{zB22} &= -0.4914 \\
\text{zB21} &= -0.1218 \\
\text{To do a check, see the 1,1 term}
\end{align*}
\]

We left the [1 -z11] zero in the RHS factor (but used its pole to cancel zeros in last column of GAtilde in forming GB)

To check, we need to put it back in.

\[
\begin{align*}
\text{GB11ch} &= \text{conv}(GB11, [1 -z11]) = 3.1532 \quad -1.2044
\end{align*}
\]
\[
G_B(D) = \tilde{G}_A(D) \cdot \begin{bmatrix}
1 & 0 \\
0 & \frac{1}{D(1 - 0.382 D)}
\end{bmatrix}
\]

(A.197)

\[
= \begin{bmatrix}
15.5031 + 1.6971D - 2.9101D^2 & 7.8725 + .9591D \\
-0.9172 + 2.4011D & 3.1532
\end{bmatrix}
\]  

(A.198)

with determinant

\[
|G_B(D)| = 66.2215 \cdot (1 + .2651D) \cdot (1 - .8378D),
\]

(A.199)

which has all roots outside unit circle. Matlab commands above checked the 22 term also. Continuing by reabsorbing the pole on the left:

\[
\text{GC22} = (GB22(1)/sr110) \cdot (1 - zB22) = 3.9116 \quad 1.9223
\]
\[
\text{GC11} = GB11 \cdot (1 - z11) = 3.1532 \quad -1.2044
\]
\[
\text{GC12} = GB1 = -0.9172 \quad 2.4011
\]
\[
\text{GC21} = (GB21(1)/sr110) \cdot (1 - zB21) = 1.9863 \quad 0.2420
\]

\[
\text{conv(GC22,GC22(2:-1:1))} + \text{conv(GC21,GC21(2:-1:1))} = 8.0000 \quad 23.0000 \quad 8.0000
\]
\[
\text{conv(GC12,GC22(2:-1:1))} + \text{conv(GC11,GC21(2:-1:1))} = -1.0000 \quad 7.0000 \quad 7.0000
\]
\[
\text{conv(GC22,GC12(2:-1:1))} + \text{conv(GC21,GC11(2:-1:1))} = 7.0000 \quad 7.0000 \quad -1.0000
\]
\[
\text{conv(GC12,GC12(2:-1:1))} + \text{conv(GC11,GC11(2:-1:1))} = -6.0000 \quad 18.0000 \quad -6.0000
\]

Checks.

Now the conversion to monic in STEPS FIVE and SIX:

\[
\text{GC0} = [\text{GC22(1)} \quad \text{GC21(1)} \\
\text{GC12(1)} \quad \text{GC11(1)}] = 3.9116 \quad 1.9863
\]
GC0*GC0' = 19.2462 2.6757
2.6757 10.7839

J = [ 0 1
1 0 ];

R = J*chol(J*GC0*GC0'*J)*J =
4.3107 0
0.8148 3.2839

R = R'
4.3107 0.8148
0 3.2839

R*R' = 19.2462 2.6757
2.6757 10.7839. checks.

S = diag(diag(R)) diag(diag(R)). =
18.5823 0
0 10.7839

G = R*inv(diag(diag(R))) =
1.0000 0.1890
0 1.0000

G*S*G' = 19.2462 2.6757
2.6757 10.7839. (checks)

GC1 = [GC22(2) GC21(2)
GC12(2) GC11(2)] =
1.9223 0.2420
2.4011 -1.2044

tempG1 = GC1*inv(GC0) =
0.4439 -0.2029
0.4568 -0.6697

tempG0 = GC0*inv(GC0) =
1.0000 0.0000
0 1.0000 (checks)

inv(S) =
0.0538 0
0 0.0927

S = 18.5823 0
0 10.7839

repeated, these are MMSE prediction error variances based on past sequence values and spatial estimation of upper mode 2 from lower mode 1.

G22 = [G0(1,1) G1(1,1)]
G21 = [G0(1,2) G1(1,2)]
G12 = [G0(2,1) G1(2,1)]
G11 = [G0(2,2) G1(2,2)]

G22 = 1.0000 0.4439
G21 = 0.2481 -0.0927
G12 = 0 0.4568
G11 = 1.0000 -0.5564

S(1,1)*conv(G22,G22(2:-1:1))+S(2,2)*conv(G21,G21(2:-1:1)) = 8.0000 23.0000 8.0000
S(1,1)*conv(G22,G12(2:-1:1))+S(2,2)*conv(G21,G11(2:-1:1)) = 7.0000 7.0000 -1.0000
S(1,1)*conv(G12,G22(2:-1:1))+S(2,2)*conv(G11,G21(2:-1:1)) =-1.0000 7.0000 7.0000
Dr. Swenson’s desired factorization is
\[
\begin{bmatrix}
8D^{-1} + 23 + 8D & 7D^{-1} + 7 - D \\
-D^{-1} + 7 + D & -6D^{-1} + 18 - 6D
\end{bmatrix} = \begin{bmatrix}
1 + .4439D & .2481 - .0927D \\
.4568D & 1 - .5564D
\end{bmatrix} \cdot \begin{bmatrix}
18.5823 & 0 \\
0 & 10.7839
\end{bmatrix} \cdot \begin{bmatrix}
1 + .4439D^{-1} & .4568D^{-1} \\
.2481 - .0927D^{-1} & 1 - .5564D^{-1}
\end{bmatrix}
\]

A.4.5 MIMO Paley Wiener Criterion and Matrix Filter Realization

Minimum-phase vector signals and matrix filters are of interest in data transmission not only because they are causal and admit causal invertible inverses (one of the reasons for their study more broadly in digital signal processing) but because they allow best results with MIMO Decision Feedback as in Section 3.10. These minimum-phase matrix filters are also useful in noise whitening. While Section A.4 provided a direct construction of \(G_x(D)\) when \(R_{xx}(D)\) has fractional fraction polynomials, this subsection provides a more general construction for any more general function that satisfies the MIMO Paley Wiener criterion provided within.

A.4.5.1 Analytic Functions of Matrices

Analytic functions, like \(\ln(x)\) and \(e^x\) as used here, have convergent power-series representations like
\[
e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + ... \quad (A.201)
\]
\[
\sum_{m=0}^{\infty} \frac{x^m}{m} \quad (A.202)
\]
\[
\ln(x) = \frac{(x - 1)}{2} - \frac{(x - 1)^2}{2} + \frac{(x - 1)^3}{3} + ... \quad (A.203)
\]
\[
\sum_{m=1}^{\infty} \frac{(-1)^{m-1} \cdot (x - 1)^m}{m} \quad (A.204)
\]

for all values of \(x\). When the argument of the function is a square matrix \(R\), the value of the corresponding output matrix of the same dimension can be found by insertion of this matrix into the power series, so\(^{23}\)
\[
e^R = \sum_{m=0}^{\infty} \frac{R^m}{m} \quad (A.205)
\]
\[
\ln(R) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1} \cdot (R - I)^m}{m} \quad (A.206)
\]

With some care on aversion of commuting matrices, the following will hold:
\[
\ln(R_1 \cdot R_2) = \ln(R_1) + \ln(R_2) \quad (A.207)
\]
\[
e^{R_1 + R_2} = e^{R_1} \cdot e^{R_2} \quad (A.208)
\]

A.4.5.2 Necessity and the Sum-Rate Equivalent

Calculation of \(S_{x,0}\) and \(S_{x,0} = |S_{x,0}|\) for a factorizable D-Transform autocorrelation matrix generalizes Equations (3.232) to (3.233) in Section 3.6 as the matrix generalizations:
\[
S_{x,0} = e^{\frac{-\pi}{T} \int_0^T \ln[R_x(e^{i\omega})^2] d\omega} \quad (A.209)
\]

\(^{23}\)For non-square matrices, it is usually possible to achieve desired results by forming a square matrix \(RR^*\) and applying the power series of the function to that “squared” matrix, and then finding the positive square root through Cholesky Factorization on the result and absorbing the square root of the diagonal matrix of Cholesky factors into each of the triangular matrices.
\[
\ln (S_{x,0}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |R_{xx}(e^{i\omega})| \cdot d\omega \tag{A.210}
\]
\[
S_{x,0} = e^{\frac{1}{\pi} \int_{-\pi}^{\pi} \ln |R_{xx}(e^{i\omega})| \cdot d\omega} \tag{A.211}
\]
\[
\ln (S_{x,0}) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |R_{xx}(e^{i\omega})| \cdot d\omega \tag{A.212}
\]

which have components corresponding to the first and last terms of the factorization in (A.162) integrating to zero because they are periodic (see A.203) for \(\ln(x+1)\) and being integrated over one period of the fundamental frequency, except for the constant term that is monic and then also integrates to zero because \(\ln(1) = 0\). The integral in (A.212) must be finite for the exponent in (A.211) to be finite (also, any exponent of a real function is a real positive number, so \(S_{x,0} > 0\) and real, consistent with the power spectral density). That integral of the natural log of a power-spectral density is fundamental in filter realization and in the MIMO Paley Wiener Criterion. Again, \(P_x\) is the same for \(R_{xx}(D)\) as for \(\ln |R_{xx}(D)|\) and includes the unit circle; this also means the convergence region for \(\ln |G_x(D)|\) also is the same as for \(G_x(D)\) and includes the unit circle. Further the convergence region for \(G_x^{-1}(D)\) also includes the unit circle and is the same as for \(\ln |G_x^{-1}(D)|\).

The calculation of \(S_{x,0}^{-1}\) has a very similar form to that of \(S_{x,0}\):

\[
S_{x,0}^{-1} = e^{-\frac{1}{\pi} \int_{-\pi}^{\pi} \ln |R_{xx}(e^{i\omega})| \cdot d\omega} \tag{A.213}
\]
\[
\ln (S_{x,0}^{-1}) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |R_{xx}(e^{i\omega})| \cdot d\omega \tag{A.214}
\]

Because (A.212) and (A.214) are similar, just differing in sign, and because any functions of \(G_x\) (including in particular \(\ln(\bullet)\)) are all periodic in \(\omega\), factorizability also implies

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \{|R_{xx}(e^{-i\omega})|\} \cdot d\omega < \infty \tag{A.215}
\]

(or the function \(\ln \{|R_{xx}(e^{-i\omega})|\}\) exists because this log-autocorrelation’s determinant is absolutely integrable). Essentially the finite nature of this integral corresponding to factorizable \(R_{xx}(D)\) means that the sequence’s Fourier Transform has no frequencies (except point frequencies of non-zero measure) at which it can be either zero or infinite. The non-infinite nature is consistent with the basic criterion for the norm to be absolutely integrable, but the the non-zero portion corresponds intuitively to saying any non-satisfying matrix filter has some singular “null components.” Any energy in these null components would be linear combinations of energy of other components. Null components thus carry no new information transmitted in that corresponding null space, making reliable data-detection/communication impossible (kind of a MIMO inverse to the reversibility concept and theorem in Chapter 1). Such a filter would not be reversible (causally or otherwise). Both infinite size and singular situations should be avoided. Chapter 5 will deal with such singularity and null spaces far more precisely.

The following MIMO Paley-Wiener Theorem from discrete-time spectral factorization theory formalizes when an autocorrelation function is “factorizable.” The ensuing development will essentially prove the theorem while theoretically developing a way to produce generally the factors \(G_x(D)\) and thus \(G_x^{-1}(D^{-*})\) of the previous subsection. This development also finds a useful way to handle the continuous-time case, which can be useful in noise-whitening. \textit{The reader is again reminded that any positive-semidefinite matrix and corresponding inverse transform is a candidate for spectral factorization.}

**Theorem A.4.1 (Paley Wiener Criterion)** If \(R_{xx}(e^{-i\omega})\) is any power spectrum such that both \(|R_{xx}(e^{-i\omega})|\) and thus \(\ln \{|R_{xx}(e^{-i\omega})|\}\) are absolutely integrable over \(-\pi < \omega \leq \pi\), and \(R_{xx}(D)\) is the corresponding autocorrelation matrix, then there exists a canonical discrete-time response \(G_x(D)\) that satisfies the equation

\[
R_{xx}(D) = G_x(D) \cdot S_{x,0} \cdot G_x^*(D^{-*}), \tag{A.216}
\]

388
where \( G_x(D) \) is canonical, and where the diagonal matrix of all positive elements \( S_{x,0} \) is given by

\[
\ln |S_{x,0}| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln R_{xx}(e^{-j\omega}) d\omega .
\]  
(A.217)

\[
\ln |S_{x,0}| = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\ln R_{xx}(e^{-j\omega})| d\omega .
\]  
(A.218)

For \( S_{x,0} \) to be finite, \( R_{xx}(e^{-j\omega}) \) must satisfy the discrete-time MIMO Paley-Wiener Criterion (PWC)

\[
\ln |S_{x,0}| = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\ln R_{xx}(e^{-j\omega})| d\omega < \infty .
\]  
(A.219)

The continuous-time equivalent of this MIMO PWC is that the Fourier Transform of the continuous-time autocorrelation function is factorizable

\[
R_{x,c,x,c}(s) = G_{x,c}(s) \cdot S_{x,c,0} \cdot G_{x,c}^*(-s^*) ,
\]  
(A.220)

where \( G_{x,c}(s) \) is minimum phase (all poles and zeros in the left half plane or on axis in limiting sense), with upper triangular monic \( |G_{x,c}(0)| \), whenever

\[
\frac{1}{2\pi} \int_{-\infty}^{\infty} |\ln R_{x,c,x,c}(\omega)| \frac{1 + \omega^2}{\omega^2} d\omega < \infty .
\]  
(A.221)

Constructive Proof: The equivalence of the two PW criteria in (A.219) and (A.221) (discrete- and continuous-time) follows directly from Equations (A.54) to (A.57). However, it remains to show that the condition is necessary and sufficient for the factorization to exist. The necessity of the criterion followed previously when it was shown that factorizability lead to the PWC being satisfied. The sufficiency proof will be constructive from the criterion itself.

The desired positive-definite matrix is any square root \( R_{xx}^{1/2}(e^{-j\omega}) \) that can, for instance, be found by Cholesky Factorization (and absorbing positive diagonal square-root equally into the two upper and lower factors), and this function in turn has a natural-log real matrix

\[
A(e^{-j\omega}) \triangleq \ln \left[ R_{xx}^{1/2}(e^{-j\omega}) \right] .
\]  
(A.222)

\( A(e^{-j\omega}) \) itself is periodic and by the MIMO PWC integral equation is absolutely integrable and so has a corresponding Fourier representation

\[
A(e^{-j\omega}) = \sum_{k=-\infty}^{\infty} A_k \cdot e^{-j\omega k} .
\]  
(A.223)

\[
A_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(e^{-j\omega}) \cdot e^{j\omega k} d\omega .
\]  
(A.224)

Because the Fourier Transform \( A(e^{-j\omega}) \) is positive real, then \( A_k = A_k^* \), and the D-Transform simplifies to

\[
A(D) = A_0 + \sum_{k=1}^{\infty} A_k \cdot D^k + \sum_{l=-1}^{\infty} A_l \cdot D^l ,
\]  
(A.225)

and then by letting \( k = -l \) in the second sum,

\[
A(D) = A_0 + \sum_{k=1}^{\infty} A_k \cdot D^k + \sum_{k=1}^{\infty} A_{-k} \cdot D^k
\]  
(A.226)
\[ A_k + \sum_{k=1}^{\infty} [A_k + A_{-k}] \cdot D^k \]

\[ a_0 + 2 \cdot \sum_{k=1}^{\infty} \Re [A_k] \cdot D^k , \]  

(A.227)

which defines a causal sequence \( A_k \) that corresponds to \( \ln \left[ R_{xx}^{1/2}(D) \right] \). So,

\[ R_{xx}(D) = e^{A(D)} \cdot e^{A^*(D^*)} . \]  

(A.229)

Then, the canonical factorization’s MIMO components are:

\[ S_{x,0} = e^{\frac{1}{2} \pi} \int_{-\infty}^{\infty} \ln[R_{xx}(e^{-j\omega})] \ d\omega \]  

(A.230)

\[ G_x(D) = e^{A(D)} \cdot S_{x,0}^{-1/2} . \]  

(A.231)

There is a second step that recognizes that \( S_{x,0} \) is not diagonal, just positive definite constant. This matrix itself is factored as a constant-matrix Cholesky factorization to

\[ S_{x,0} = G_{x,0} \cdot S_{x,0}^* \cdot G_{x,0}^* . \]  

(A.232)

with \( G_{x,0} \) as monic upper triangular and \( S_{x,0} \) positive definite diagonal. Then the canonical factor is

\[ G_x(D) = e^{A(D)} \cdot G_{x,0}^{-1} \cdot S_{x,0}^{-1/2} . \]  

(A.233)

The corresponding continuous-time spectrum factorization then would be found with \( R_{xx}(s) = R_{xx} \left( \frac{s^2}{1+\frac{s^2}{\omega_c^2}} \right) \) and thus \( A_c(s) = A \left( \frac{s^2}{1+\frac{s^2}{\omega_c^2}} \right) \). Then, with \( s \to j\omega_c \)

\[ S_{x_c,0} = e^{\frac{1}{2} \pi} \int_{-\infty}^{\infty} \ln[R_{x_c x_c}(\omega_c)] \ d\omega_c . \]  

(A.234)

\[ G_{x_c}(s) = e^{A_c(s)} \cdot S_{x_c,0}^{-1/2} . \]  

(A.235)

The continuous-time second step recognizes that \( S_{x_c,0} \) is not diagonal, just positive definite constant. This matrix itself is factored as a constant-matrix Cholesky factorization to

\[ S_{x_c,0} = G_{x_c,0} \cdot S_{x_c,0}^* \cdot G_{x_c,0}^* . \]  

(A.236)

with \( G_{x_c,0} \) as monic upper triangular and \( S_{x_c,0} \) positive definite diagonal. Then the canonical factor is

\[ G_{x_c}(D) = e^{A_c(D)} \cdot G_{x_c,0}^{-1} \cdot S_{x_c,0}^{-1/2} . \]  

(A.237)

If the original desired spectra were defined in continuous time, then it could be mapped into discrete time through \( \omega_c \to \tan(\theta_c) \) and then proceeding with that discrete-time mapped equivalent through the process above, ultimately leading to Equations (A.234) and (A.235). Sufficiency has thus been established in both discrete- and continuous-time. QED.

**A.5 The Matrix Inversion Lemma**

The matrix inversion lemma:

\[ [A + BCD]^{-1} = A^{-1} - A^{-1} B [C^{-1} + DA^{-1}B]^{-1} DA^{-1} , \]  

(A.238)

is often useful in simplifying matrix expressions.

This can be proofed through straightforward matrix multiplication.
Appendix B

Equalization for Partial Response

Just as practical channels are often not free of intersymbol interference, such channels are rarely equal to some desirable partial-response (or even controlled-ISI) polynomial. Thus, an equalizer may be used to adjust the channel’s shape to that of the desired partial-response polynomial. This equalizer then enables the use of a partial-response sequence detector or a symbol-by-symbol detector at its output (with partial-response precoder). The performance of either of these types of detectors is particularly sensitive to errors in estimating the channel, and so the equalizer can be crucial to achieving the highest levels of performance in the partial-response communication channel. There are a variety of ways in which equalization can be used in conjunction with either sequence detection and/or symbol-by-symbol detection. This section introduces some equalization methods for partial-response signaling.

Chapter 9 studies sequence detection so terms related to it like MLSD (maximum likelihood sequence detection) and Viterbi detectors/algorithm are pertinent to readers already familiar with Chapter 9 contents and can be viewed simply as finite-real-time-complexity methods to implement a full maximum-likelihood detector for an ISI channel by other readers of this Appendix.

B.1 Controlled ISI with the DFE

Section 3.8.2 showed that a minimum-phase channel polynomial $H(D)$ can be derived from the feedback section of a DFE. This polynomial is often a good controlled intersymbol interference model of the channel when $H(D)$ has finite degree $\nu$. When $H(D)$ is of larger degree or infinite degree, the first $\nu$ coefficients of $H(D)$ form a controlled intersymbol interference channel. Thus, a detector on the output of the feedforward filter of a DFE can be designed using the controlled ISI polynomial $H(D)$ or approximations to it.

B.1.1 ZF-DFE and the Optimum Sequence Detector

Section 3.1 showed that the sampled outputs, $y_k$, of the receiver matched filter form a sufficient statistic for the underlying symbol sequence. Thus a maximum likelihood (or MAP) detector can be designed that uses the sequence $y_k$ to estimate the input symbol sequence $x_k$ without performance loss. The feedforward filter $1/(\eta_0||p||H^*(D^{-1}))$ of the ZF-DFE is invertible when $Q(D)$ is factorizable. The reversibility theorem of Chapter 1 then states that a maximum likelihood detector that observes the output of this invertible ZF-DFE-feedforward filter to estimate $x_k$ also has no performance loss with respect to optimum. The feedforward filter output has $D$-transform

$$Z(D) = X(D) \cdot H(D) + N'(D).$$

(B.1)

The noise sequence $n'_k$ is exactly white Gaussian for the ZF-DFE, so the ZF-DFE produces an equivalent channel $H(D)$. If $H(D)$ is of finite degree $\nu$, then a maximum-likelihood (full sequence of transmitted signals is considered to be a one-shot transmission, which would nominally require infinite delay and complexity for an infinite-length input, but Chapter 9 shows how to implement such a detector recursively
with finite delay and finite-real-time complexity) detector can be designed based on the controlled-ISI polynomial \( H(D) \) and this detector has minimum probability of error. Figure B.1 shows such an optimum receiver.

Figure B.1: The partial response ZF equalizer, decomposed as the cascade of the WMF (the feedforward section of the ZF-DFE) and the desired partial response channel \( B(D) \).

When \( H(D) \) has larger degree than some desired \( \nu \) determined by complexity constraints, then the first \( \nu \) feedback taps of \( H(D) \) determine

\[
H'(D) = 1 + h_1 D^1 + \ldots + h_\nu D^\nu, \tag{B.2}
\]
a controlled-ISI channel. Figure

Figure B.2: Limiting the number of states with ZF-DFE and MLSD.

B.2 illustrates the use of the ZF-DFE and a sequence detector. The second feedback section contains all the channel coefficients that are not used by the ML (sequence, see Chapter 9) detector. These coefficients have delay greater than \( \nu \). When this second feedback section has zero coefficients, then the configuration shown in Figure B.2 is an optimum detector. When the additional feedback section is not zero, then this structure is intermediate in performance between optimum and the ZF-DFE with symbol-by-symbol detection. The inside feedback section is replaced by a modulo symbol-by-symbol detector when precoding is used.

Increase of \( \nu \) causes the minimum distance to increase, or at worst, remain the same. Thus, the ZF-DFE with sequence detector in Figure B.2 defines a series of increasingly complex receivers whose performance approach optimum as \( \nu \to \infty \). A property of a minimum-phase \( H(D) \) is that

\[
\sum_{i=0}^{\nu'} |h_i|^2 = \| h' \|^2 \tag{B.3}
\]
is maximum for all $\nu' \geq 0$. No other polynomial (that also preserves the AWGN at the feedforward filter output) can have greater energy. Thus the SNR of the signal entering the Viterbi Detector in Figure B.2, $\bar{E}_{\nu'2} \frac{||h'||^2}{N_0}$, also increases (nondecreasing) with $\nu$. This SNR must be less than or equal to $\text{SNR}_{MFB}$.

### B.1.2 MMSE-DFE and sequence detection

Symbol-by-symbol detection’s objective is to maximize SNR in an unbiased detector, and so SNR maximization was applied in Chapter 3 to the DFE to obtain the MMSE-DFE. A bias in symbol-by-symbol detector was removed to minimize probability of error. The monic, causal, minimum-phase, and unbiased feedback polynomial was denoted $G_U(D)$ in Section 3.6. A sequence detector can use the same structures as shown in Figure B.1 and B.2 with $H(D)$ replaced by $G_u(D)$. For instance, Figure B.4 is the same as Figure B.2, with an unbiased MMSE-DFE’s MS-WMF replacing the WMF of the ZF-DFE. A truncated version of $G_U(D)$ corresponding to $H'(D)$ is denoted $G'_U(D)$. The error sequence associated with the unbiased MMSE-DFE is not quite white, nor is it Gaussian. So, a sequence detector based on squared distance is not quite optimum, but it is nevertheless commonly used because the exact optimum detector could be much more complex. As $\nu$ increases, the probability of error decreases from the level of the unbiased MMSE-DFE, $\text{SNR}_{MMSE-DFE,U}$ when $\nu = 0$, to that of the optimum detector when $\nu \to \infty$. The matched filter bound, as always, remains unchanged and is not necessarily obtained. However, minimum distance does increase with $\nu$ in the sequence detectors based on a increasing-degree series of $G_U(D)$.

### B.2 Equalization with Fixed Partial Response $B(D)$

The derivations of Section 3.6 on the MMSE-DFE included the case where $B(D) \neq G'(D)$, which this section reuses.

#### B.2.1 The Partial Response Linear Equalization Case

In the linear equalizer case, the equalization error sequence becomes

$$E_{pr}(D) = B(D) \cdot X(D) - W(D) \cdot Y(D).$$

(B.4)

Section 3.6 minimized MSE for any $B(D)$ over the coefficients in $W(D)$. The solution was found by setting $E[ E_{pr}(D)y^*(D^{-1}) ] = 0$, to obtain

$$W(D) = B(D) \frac{R_{xy}(D)}{R_{yy}(D)} = \frac{B(D)}{||p|| (Q(D) + 1/\text{SNR}_{MFB})},$$

(B.5)

which is just the MMSE-LE cascaded with $B(D)$. Figure

![Diagram of partial-response linear equalization](image)

Figure B.3: Partial-response linear equalization.
B.3, shows the MMSE-PREQ (MMSE - “Partial Response Equalizer”). The designer need only realize the MMSE-LE of Section 3.4 and follow it by a filter of the desired partial-response (or controlled-ISI) polynomial $B(D)$.\footnote{This also follows from the linearity of the MMSE estimator.} For this choice of $W(D)$, the error sequence is

$$E_{pr}(D) = B(D)X(D) - B(D)Z(D) = B(D)[E(D)] \quad \text{(B.6)}$$

where $E(D)$ is the error sequence associated with the MMSE-LE. From (B.6),

$$\tilde{R}_{e_{pr}, e_{pr}}(D) = B(D)\tilde{R}_{ee}(D)B^{-1}(D^{-1}) = \frac{B(D)\mathcal{N}_0 B^*(D^{-1})}{\|p\|^2 (Q(D) + 1/\text{SNR}_{MFB})} \quad \text{(B.7)}$$

Thus, the MMSE for the PREQ can be computed as

$$\sigma_{MMSE-PREQ}^2 = \frac{T}{2\pi} \int_{-\pi T}^{\pi T} \frac{|B(e^{-j\omega T})|^2 \mathcal{N}_0}{\|p\|^2 (Q(e^{-j\omega T}) + 1/\text{SNR}_{MFB})} d\omega \quad \text{(B.8)}$$

The symbol-by-symbol detector is equivalent to subtracting $(B(D) - 1)X(D)$ before detection based on decision regions determined by $x_k$. The SNR$_{MMSE-PREQ}$ becomes

$$\text{SNR}_{MMSE-PREQ} = \frac{\tilde{e}_x}{\sigma_{MMSE-PREQ}^2} \quad \text{(B.9)}$$

This performance can be better or worse than the MMSE-LE, depending on the choice of $B(D)$; the designer usually selects $B(D)$ so that $\text{SNR}_{MMSE-PREQ} > \text{SNR}_{MMSE-LE}$. This receiver also has a bias, but it is usually ignored because of the integer coefficients in $B(D)$ – any bias removal could cause the coefficients to be noninteger.

While the MMSE-PREQ should be used for the case where symbol-by-symbol and precoding are being used, the error sequence $E_{pr} = B(D)E(D)$ is not a white noise sequence (nor is $E(D)$ for the MMSE-LE), so that a Viterbi Detector designed for AWGN on the channel $B(D)$ would not be the optimum detector for our MMSE-PREQ (with scaling to remove bias). In this case, the ZF-PREQ, obtained by setting $\text{SNR}_{MFB} \to \infty$ in the above formulae, would also not have a white error sequence. Thus a linear equalizer for a partial response channel $B(D)$ that is followed by a Viterbi Detector designed for AWGN may not be very close to an optimum detection combination, unless the channel pulse response were already very close to $B(D)$, so that equalization was not initially necessary. While this is a seemingly simple observation made here, there are a number of systems proposed for use in disk-storage detection that overlook this basic observation, and do equalize to partial response, “color” the noise spectrum, and then use a WGN Viterbi Detector. The means by which to correct this situation is the PR-DFE of the next subsection.

**B.2.2 The Partial-Response Decision Feedback Equalizer**

If $B(D) \neq G(D)$ and the design of the detector mandates a partial-response channel with polynomial $B(D)$, then the optimal MMSE-PRDFE is shown in Figure
Again, using our earlier result that for any feedback section $G_U(D) = B(D) + \hat{B}(D)$. The error sequence is the same as that for the MMSE-DFE, and is therefore a white sequence. The signal between the two feedback sections in Figure B.4 is input to the sequence or symbol-by-symbol detector. This signal can be processed on a symbol-by-symbol basis if precoding is used (and also scaling is used to remove the bias - the scaling is again the same scaling as used in the MMSE-DFE), and $\hat{B}_U(D) = G_U(D) - B_U(D)$, where

$$B_U(D) = \frac{\text{SNR}_{\text{MMSE-DFE}}}{\text{SNR}_{\text{MMSE-DFE},U}} \left[ B(D) - \frac{1}{\text{SNR}_{\text{MMSE-DFE}}} \right].$$  \hspace{1cm} (B.10)

However, since the MMSE-PRDFE error sequence is white, and because the bias is usually small so that the error sequence in the unbiased case is also almost white, the designer can reasonably use an ML (Viterbi Sequence Detector) designed for $B(D)$ with white noise.

If the bias is negligible, then a ZF-PRDFE should be used, which is illustrated in Figure B.2, and the filter settings are obtained by setting $\text{SNR}_{MFB} \rightarrow \infty$ in the above formula.
Bibliography


Index

alternate mark inversion, 263
AMI, 263
autocorrelation, 362
matrix, 231, 355

band limited
channel, 163
bandlimited channel, 166
bias
receiver, 179
removal, 209
bilinear transform, 363

canonical
channel, 296
factor, 221
performance, 296
response, 224
carrier-less amplitude phase, 217
CDEF, 316
Coded-OFDM, 301
cross correlation
vector, 231
cross-correlation
matrix, 355
crosstalk, 281
data rate, 164
dead bands, 312
DFE, 219
color, 245
feedback, 221
M4, 289
Matlab program, 245
MIMO, 285
MMSE, 219
optimimum, 312
performance, 221
zero-forcing, 223
Discrete Multi-Tone, 301
diversity, 272
cchannel, 272
receiver, 273
dubinary, 261
duobinary
modified, 263
encoder
differential, 265
entropy, 291
conditional, 291
equalization, 162
equalizer
decision feedback, 219
delay, 231
linear
MMSE, 206
M4, 287, 290
MIMO, 281
excess bandwidth, 186
eye diagram, 176
Gaussian
multivariate, 356
gaussian
multivariate, 356
Hilbert Space, 358
innovations, 299
intersymbol interference, 164
ISI, 162
ISI-Channel Model, 171
linear prediction, 292
Lorentzian Pulse, 176
Matched Filter Bound, 181
mean-square distortion, 175
MIMO
channel, 281
DFE, 285
equalizer, 281
precoder, 290
space-time, 281
ZFE, 290
MMSE, 354
DFE, 219
estimate, 354
whitened matched filter, 221
modified duobinary, 169
modulo, 265
Multi-Tone, 301
mutual information, 293

nodulo
  operator, 252
noise
  worst case, 320
noise-equivalent response, 170
norm
  euclidean, 358
  matrix, 358
Nyquist Criterion, 184

one shot, 181
one-shot
  MIMO, 282
orthogonality principle, 357

Paley Wiener Criterion, 298, 303, 365, 366
partial response, 258
  extended, 263
  quadrature, 269
peak distortion, 174
power spectrum, 362
precoder, 251, 264
  flexible, 254
  general, 268
  Laroia, 254
  MIMO, 290
  Tomlinson, 251
pulse response, 167
  normalized, 167
pythagorean theorem
  generalized, 359

raised cosine
  square root, 191
raised cosine pulse, 188
RAKE, 273
receiver
  SNR, 178
  unbiased, 180
  unconstrained, 180

sinc function, 186
slush packing, 322
SNNR
  biased, 180
SNR
  matrix, 283
  MFB, 181, 222
  unbiased, 180
spectral factorization, 360, 364
symbol
  period, 164
  rate, 164
symbol-by-symbol
  detection, 162, 164

transmitter
  optimum, 320
VDSL
  Olympics, 316
vector
  norm, 358
  vector space, 357
  normed, 358
vestigial symmetry, 185
water-filling, 303
zero-forcing equalizer, 194
  loss, 196
  performance, 194