Wireless Channel Estimation With Applications to Secret Key Generation

by

Alireza Movahedian
B.Sc., Ferdowsi University of Mashhad, 1993
M.Sc., University of Tehran, 1996

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

Wireless Channel Estimation With Applications to Secret Key Generation

by

Alireza Movahedian
B.Sc., Ferdowsi University of Mashhad, 1993
M.Sc., University of Tehran, 1996

Supervisory Committee

Dr. Michael L. McGuire, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Stephen W. Neville, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yvonne Coady, Outside Member
(Department of Computer Science)

Supervisory Committee

Dr. Michael L. McGuire, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Stephen W. Neville, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yvonne Coady, Outside Member
(Department of Computer Science)

ABSTRACT

This research investigates techniques for iterative channel estimation to maximize channel capacity and communication security. The contributions of this dissertation are as follows: i) An accurate, low-complexity approach to pilot-assisted fast-fading channel estimation for single-carrier modulation with a turbo equalizer and a decoder is proposed. The channel is estimated using a Kalman filter (KF) followed by a zero-phase filter (ZPF) as a smoother. The combination of the ZPF with the KF of the channel estimator makes it possible to reduce the estimation error to near the Wiener bound. ii) A new semi-blind channel estimation technique is introduced for multiple-input-multiple-output channels. Once the channel is estimated using a few pilots, a low-order KF is employed to progressively predict the channel gains for the upcoming blocks. iii) The capacity of radio channels is investigated when iterative channel estimation, data detection, and decoding are employed. By taking the uncertainty in decoded data bits into account, the channel Linear Minimum Mean Square Error (LMMSE) estimator of an iterative receiver with a given pilot ratio is obtained. The derived error value is then used to derive a bound on capacity. It is shown that in slow fading channels, iterative processing provides only a marginal advantage over non-iterative approach to channel estimation. Knowing the capacity gain from iterative processing versus purely pilot-based channel estimation helps a designer to compare the performance of an iterative receiver against a non-iterative one and select the best balance between performance and cost. iv) A Radio channel is characterized by random parameters which can be used to generate shared secret keys by the communicating parties when the channel is estimated. This research studies upper bounds on the rate of the secret keys extractable from iteratively estimated channels. Various realistic scenarios are considered where the transmission is half-duplex and/or the channel is sampled under the Nyquist rate. The effect of channel sampling interval, fading rate and noise on the key rate is demonstrated. The results of this research can be beneficial for the design and analysis of reliable and secure mobile wireless systems.

Contents

Supervisory Committee **ii**

Abstract **iii**

Table of Contents **v**

List of Figures **viii**

Symbols and Notations **xi**

Abbreviations **xii**

Acknowledgements **xiii**

1 Introduction **1**

1.1 Motivation 1

1.2 Problem Statement 2

1.3 Contributions 4

2 Background **7**

2.1 General Characteristics of Radio Channel 7

2.2 Channel Capacity 13

2.3 Channel Estimation 16

2.3.1 Iterative Channel Estimation 18

2.3.2 System Model 21

2.3.3 Optimal Linear Channel Estimator Bound 22

2.4 Physical-layer security 22

2.4.1 Basic Concepts 23

2.4.2 Secrecy Capacity 24

2.4.3 Secret Key Generation 26

2.4.4 Channel Fading as a Source of Randomness for Key Generation 28

3 Estimation of Fast Fading SISO Channels 30

3.1 Introduction 30

3.2 System Model 31

3.2.1 Transmitter 31

3.2.2 Channel Model 32

3.2.3 Receiver 32

3.3 Channel Estimation and Smoothing 34

3.4 ZPF Design 39

3.5 Soft-in-Soft-Out Equalizer and Decoder 41

3.5.1 Generating Equalizer Extrinsic Information on Data Bits . . . 44

3.5.2 Generating Decoder Extrinsic Information 45

3.5.3 LLR to Symbol Conversion 45

3.6 Computational Cost 45

3.7 Simulation Results 47

3.8 Extrinsic Information Transfer Chart Analysis 51

3.9 Estimation of Fast Fading Radio Channels Based on AR(4) for 256-QAM 60

3.9.1 Channel Estimation 60

3.9.2 Complexity Analysis 62

3.9.3 Simulations 62

3.9.4 EXIT Chart Analysis 65

3.10 Summary 65

4 MIMO-OFDM Channel Estimation 67

4.1 Basic MIMO-OFDM Structure 68

4.2 System Model 69

4.3 Channel Estimation Algorithm 72

4.4 Symbol Detection 73

4.5 Computational Complexity 75

4.6 Simulations 76

4.7 Summary 77

5 Capacity of Iteratively Estimated Doubly-Selective Channels 80

5.1 Introduction 81

5.2 System model 83

5.3 Bound on the Mean Square Error for Channel Estimation 88
5.3.1 Worst case scenario for channel estimation error 89
5.3.2 Approximation to LMMSE estimation error 92
5.4 A Lower Bound on Capacity 94
5.5 Finite Order Modulation 97
5.6 Numerical Results 98
5.7 More bandwidth efficient pilot schemes 101
5.8 Summary 103

6 Secret Key Generation 105

6.1 Secret Key Generation In Vehicular Wireless Networks 105
6.1.1 Channel Estimation Error 106
6.1.2 Secret Key Capacity of Iterative Receivers: Full-duplex Transmission 109
6.1.3 Secret Key Capacity of Iterative Receivers: Time-division Duplexing 111
6.1.4 Analysis of Long Measurement Blocks 113
6.2 Simulations 115
6.3 Summary 120

7 Conclusions and Future Work 122

7.1 Research Results 122
7.2 Future Work 124

Bibliography 125

List of Figures

Figure 1.1 A multipath fading channel 3

Figure 2.1 Transmitter 21

Figure 3.1 Receiver structure 33
Figure 3.2 Block processing. 35
Figure 3.3 The magnitude response of an elliptic low-pass filter. 39
Figure 3.4 Soft-in-soft-out equalizer. 42
Figure 3.5 BER versus E_b/N_0 for a 4-QAM receiver. 49
Figure 3.6 BER versus E_b/N_0 for a 16-QAM receiver. 50
Figure 3.7 BER versus E_b/N_0 for a 64-QAM receiver. 51
Figure 3.8 BER versus E_b/N_0 for a 64-QAM scheme, comparing ZPF, ordinary IIR and FIR filters. 52
Figure 3.9 NMSE against the number of lags for a 64-QAM scheme assuming known symbols at the receiver. 52
Figure 3.10 NMSE versus E_b/N_0 for 64-QAM, under different channel models using a KF with four lags. 53
Figure 3.11 Performance of a 4-QAM turbo receiver with LDPC code over an eight-tap channel compared with the perfect-channel and the EKF method. 53
Figure 3.12 BER vs ZPF's passband edge f_p when the normalized Doppler frequency is fixed at f_D =0.01. 54
Figure 3.13 An EXIT chart. 55
Figure 3.14 Block diagram used to generate the EXIT charts: Decoder setup 56
Figure 3.15 Block diagram used to generate the EXIT charts: Equalizer setup 56
Figure 3.16 EXIT charts for a 4-QAM receiver under different E_b/N_0 including the average trajectory for E_b/N_0 =4dB. 57
Figure 3.17 EXIT charts for a 16-QAM receiver under different E_b/N_0 including the average trajectory for E_b/N_0 =8 dB. 58

Figure 3.18 EXIT charts for a 64-QAM receiver using CE-BEM with Q =3 and E_b/N_0 =13 dB. 58
Figure 3.19 EXIT chart for a 4-QAM receiver using LDPC code. 59
Figure 3.20 BER vs. E_b/N_0 under different modulation schemes 66
Figure 3.21 EXIT chart for a 256-QAM receiver at $E_b/N_0 = 17$dB. 66

Figure 4.1 MIMO-OFDM structure. 68
Figure 4.2 Iterative processing of the last N_a symbols of block j is performed using the last N_d symbols of block $j-1$. 71
Figure 4.3 4-QAM 4×4 MIMO-OFDM receiver with LMMSE detector compared with the statistical approach of [35] and the RLS based method of [94] at $f_D = 10^{-4}$. 78
Figure 4.4 16-QAM 4×4 MIMO-OFDM receiver with LMMSE detector compared with the RLS based method of [94] at $f_D = 10^{-4}$. . . 78

Figure 5.1 Receiver structure 84
Figure 5.2 Distribution of eigenvalues for known data symbols ($\sigma_u^2 = 1$); $L = 2$, $f_D = 0.01$, SNR=10dB 94
Figure 5.3 $\sigma^2_{\tilde{g}\tilde{g}}$(dB) versus f_D for known data symbols ($\sigma_u^2 = 0$); $L = 2$, SNR=10dB 94
Figure 5.4 Lower bound on capacity under various pilot rates, for SNR=$7dB$, $l_p = 5, f_D = 0.01, L = 2$ 101
Figure 5.5 Lower bound on capacity under various fading rates, for SNR=$7dB$, $l_p = 5, l_s = 20, L = 2$ 101
Figure 5.6 SNR penalty versus data-to-pilot ratio l_s/l_p, $f_D = 0.01, L = 2$ . . 102
Figure 5.7 Theoretical bounds on the channel capacity compared to the capacity of an iterative receiver with $l_s = 20, l_p = 5, f_D = 0.01, L = 2$102
Figure 5.8 Detector's mutual information $\mathcal{I}_{E_f}$ versus $\mathcal{I}_D$ for an iteratively estimated single-tap fading channel using 4-QAM, 16-QAM and 64-QAM modulation compared with Gaussian transmission, at SNR=7dB 103
Figure 5.9 Capacity versus $\mathcal{I}_D$ under different values of M with $L = 2, N =$ 1000, $l_s = 20$ and SNR=7 dB 104

Figure 6.1 Signal model 110
Figure 6.2 Equivalent whitening process 110

Figure 6.3 NMSE versus SNR for a Kalman filter compared to a Wiener filter 116

Figure 6.4 Key rate in bits per second versus SNR for different wireless standards. 118

Figure 6.5 Key capacity of IEEE 802.11p versus SNR for different number of paths. 118

Figure 6.6 Key capacity of IEEE 802.11p versus SNR for a 10-tap channel under different delay-power profile. 119

Figure 6.7 Key capacity versus SNR under full-duplex and time-division duplexing at $f_D = 0.01$. 119

Figure 6.8 Key capacity versus f_D under full-duplex and time-division duplexing at SNR= 10dB. 120

Symbols and Notations

$\mathbf{a}$	A vector
$\mathbf{a}_i$	i^{th} Element of vector $\mathbf{a}$
$\mathbf{A}$	A matrix
$(\mathbf{A})_{i,j}$	Element on row i column j of matrix $\mathbf{A}$
$\mathbf{A}(n,:)$	Row n of matrix $\mathbf{A}$
$\mathbf{A}(:,m)$	Column m of matrix $\mathbf{A}$
$[\cdots]_{N\times M}$	A matrix with N rows and M columns
$\lambda_i(\mathbf{A})$	i^{th} Eigenvalue of matrix $\mathbf{A}$, , where the eigenvalues are ordered so that $\lvert\lambda_1\rvert \geq \lvert\lambda_2\rvert \geq \cdots \lvert\lambda_N\rvert$ where $\mathbf{A}$ is an N by N matrix.
$\mathbf{A}^H$	Hermitian of matrix $\mathbf{A}$
$\mathbf{A}^T$	Transpose of matrix $\mathbf{A}$
$\det \mathbf{A}$	Determinant of matrix $\mathbf{A}$
$\mathcal{C}$	Channel capacity
$\operatorname{diag}(\mathbf{A})$	Diagonal of matrix $\mathbf{A}$
$\operatorname{diag}(\mathbf{a})$	Diagonal matrix with vector $\mathbf{a}$ on the main diagonal
$g(n;l)$	Channel gain of propagation tap l at time n
f_D	Normalized Doppler frequency
$H(X)$	Entropy of random variable X
$H(X\lvert Y)$	Conditional entropy of random variable X conditioned on random variable Y
$\mathcal{I}(X;Y)$	Mutual information between random variables X and Y
$\mathbf{I}_N$	Square $N\times N$ matrix
δ_{ij}	Kronecker delta function of variables i and j
$\otimes$	Matrix Kronecker product

Abbreviations

AR	Auto-regressive
AWGN	Additive White Gaussian Noise
BEM	Basis Expansion Model
BER	Bit Error Rate
BICM	Bit Interleaved Coded Modulation
CE-BEM	Complex Exponential-Basis Expansion Model
DPSS	Discrete Prolate Spheroidal Sequence
ECC	Elliptic Curve Cryptography
EKF	Extended Kalman Filter
EXIT	Extrinsic Information Transfer
FIR	Finite Impulse Response
IIR	Infinite Impulse Response
ISI	Intersymbol Interference
KF	Kalman Filter
KLT	Karhunen-Loève Transform
LDPC	Low Density Parity Check
LLR	Log Likelihood Ratio
LMMSE	Linear Minimum Mean Square Error
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
NMSE	Normalized Mean Squared Error
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
RLS	Recursive Least Squares
RV	Random Variable
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
UWB	Ultra Wide Band
ZPF	Zero Phase Filter

ACKNOWLEDGEMENTS

I would like to thank:

Professor Michael McGuire, for his invaluable support

My friends, Dr. Masoud Haji Aghajani and Dr. Ebrahim Ghafarzadeh for their generous help.

.
Alireza Movahedian

Chapter 1

Introduction

In this chapter, we outline the work motivation, problem statement, and research contributions. The organization of the dissertation comes at the end of this chapter.

1.1 Motivation

Almost three-quarters of the world's population already has access to mobile communications; yet the global mobile communications industry is expected to continue to grow rapidly for many years [91]. This growth is fueled by the proliferation of mobile applications, now penetrating every aspect of daily life. The volume of mobile data is expected to increase 13-fold in five years, whereas the connection speed will assume a 7-fold growth [1]. This enormous demand for high data rates is primarily powered by mobile video and online gaming.

One of the main areas in mobile communications concerns vehicular networks. The advancement in mobile technology enables cars to exchange real-time information with external devices, other cars, or base stations to increase the vehicle performance and improve the driving experience. Video calls, mobile video and gaming are among the emerging services which will revolutionize the way cars are used[1].

Next-generation mobile systems need new techniques to fully exploit the available wireless spectrum. Exploiting the full communications capacity of wireless fading channels is challenging [22]. New techniques are needed to allow wireless channels to provide provably secure and private communications [24]. To achieve full capacity use and/or perfect privacy over radio channels requires that the radio channel be

[1] By 2017, 60% of new cars will include connected car solutions, according to an Allied Business Intelligence (ABI)'s report, 2012.

accurately estimated by the communicating parties [131]. In light of the prospective necessities of the next-generation mobile networks, this dissertation contributes new algorithms for estimating fast-fading radio channels suitable for next-generation wireless protocols. The proposed techniques may be considered in the future standards and specifications as viable solutions to some of the above mentioned challenges. Next generation wireless networks are under development such as the WiGig and IEEE 802.11ad standards [45, 127, 133]. These standards exploit the 60 GHz spectrum to increase the data rates of existing networks. Implementing the new technology on wireless embedded systems, wireless sensor nodes, etc., requires low-complexity and efficient receiver techniques designed to handle rapidly varying channels. Potential applications include control signaling for remote-operated aerial vehicles, high-reliability communications for emergency vehicles, videoconferencing on high-speed trains, narrowband communications between cars, etc.

1.2 Problem Statement

This research concerns iterative channel estimation as well as secret key generation from the channel estimates for mobile wireless communications. An accurately estimated channel is not only important to reliable communication, it is also an abundant source of secret key bits for securing the communication.

Different types of channel impairments must be treated by the receiver to achieve accurate channel estimation. In common radio communication systems, the signal arrives at the receiver via different propagation paths, each with distinct amplitude and delay, creating so-called multipath propagation (Fig. 1.1). Different propagation delays cause different phase shifts of signal components, giving rise to constructive or destructive interference. The phase shift depends mainly on the relative location of the receiver with respect to the transmitter, as well as to interacting objects on the path. Therefore, the overall signal amplitude will change with time if any object movement is involved. As the signal is received through multiple paths, the superposition of the components coming from different directions and having different phases, induces rapid fluctuations in the signal strength [132]. This phenomenon is caused by the mobile's movements and is known as time-selectivity. Another impediment in communication systems is inter-symbol interference (ISI) caused by channel memory, creating frequency selectivity in the channel.

Reliable communication over time-varying, frequency-selective channels calls for

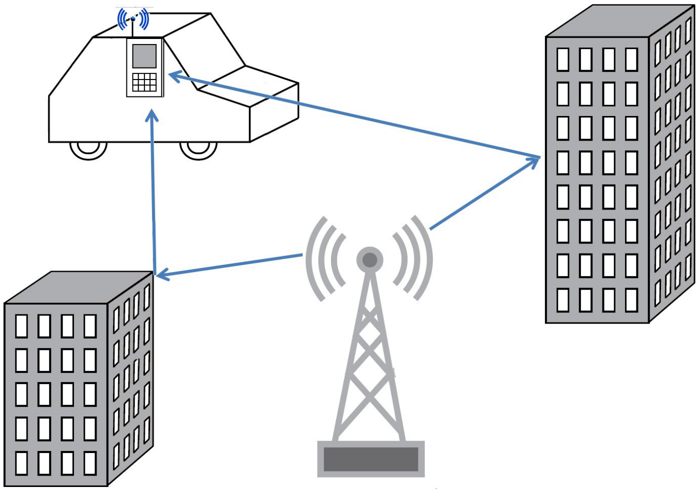

Figure 1.1: A multipath fading channel

accurate channel estimation algorithms at the receiver. For quasi-static or slow fading channels, conventional pilot-based estimation methods offer adequate performance [52,166]. For lower fading rates of up to 0.1% of the sample rate, low complexity and near optimal techniques are available for orthogonal frequency division multiplexing (OFDM) and frequency-domain equalization schemes [72,102,166]. However, these methods rely on the fundamental assumption that the channel is nearly static over the duration of long blocks of symbols, so they fail to work at the higher fading rates. Conventional frequency domain channel estimation and equalization methods exhibit irreducible error floor at high Doppler frequencies [79]. Accurate estimation of fast-fading channels with fading rates of up to 1% of the symbol rate using the previously proposed methods entails high computational cost, which are not feasible for many mobile computing applications.

This research first addresses the problem of estimating a fast-fading channel, with a normalized Doppler frequency as large as 1% of the symbol rate for higher order modulation schemes. By using a basis expansion model (BEM) to represent the channel variation over a block, the estimation problem reduces to estimating the less numerous BEM coefficients. The accuracy of the BEM depends on the block length and the fading rate. For fast-fading channels, shorter blocks may be used. Short blocks of data make it possible to perform data detection without exponential computational cost [107, p. 281]. For higher order modulations, high-precision estimation of the channel is critical since the detector is sensitive to the estimation errors.

The initial estimation of the channel in iterative processing relies on the pilot symbols. The rate of pilots must be greater than the Nyquist rate to allow channel identification unless blind or semiblind estimation is used. The Nyquist rate in this case is determined by how fast the channel varies, i.e., the fading rate. As the fading rate increases, the pilot overhead increases, leading to a reduction in the effective bandwidth of the channel [105]. The pilot overhead is more of a problem in Multiple-Input Multiple-Output (MIMO) channels where a large number of parameters corresponding to the channels from each of the different antennas must be estimated. Semiblind and blind techniques address this problem by sending fewer or no pilots, but the computational cost is a burden, because they often require the inversion of large matrices [64, p. 3]. We introduce a method for semiblind estimation of MIMO channels with near-optimal performance and reasonable computational cost.

Although doubly-selective channels pose challenging problems to reliable communication, the property that makes these channel difficult for communications, i.e., there are large number of random parameters needed to characterize the channel, can actually be beneficial for security. Two parties using a doubly-selective radio channel for two-way communication must both characterize this channel to achieve high data rates. Many of these parameters cannot be measured by any third party [96], so that the random values of these parameters can be used as shared secrets to support private communications. When channel gain estimates are used to generate secret keys, the key rate is determined by the accuracy of the channel estimates as well as the rate of acquiring independent estimates from the channel, assuming that the channel is unknown to any adversary. This fact brings about a close relationship between the reliability and security problems considered in this dissertation, in the sense that more accurate channel estimation would lead to higher channel capacity as well as higher key rates (privacy).

1.3 Contributions

This research investigates techniques for iterative channel estimation and studies the effect of using these techniques on channel capacity and security. The contributions of this dissertation may be summarized as follows.

- Introducing a novel approach to iteratively estimating single-carrier fast-fading radio channels using a smoother;

- Proposing a low-complexity and accurate channel estimation method for higher order modulation in fast-fading channels with a low pilot rate;
- Introducing a method to evaluate the capacity gain of iterative channel estimation;
- Proposing a semi-blind iterative channel estimation technique for MIMO-OFDM;
- Calculating bounds on the rate of secret keys extractable from channel estimates under realistic scenarios, where the channel is sampled under the Nyquist rate and with half-duplex transmission.

The dissertation is organized into the following chapters.

Chapter 2 gives a literature review, describing important results in the area of iterative receivers and laying the foundations of the dissertation.

Chapter 3 introduces an efficient approach to estimating fast-fading doubly selective single-input-single-output (SISO) channels using a Kalman filter (KF) and smoother. The performance of the proposed method is compared with a similar state-of-the-art method. An extrinsic information transfer (EXIT) chart analysis is performed to clarify the convergence behavior of the system for the specific parameters and code in use.

Chapter 4 introduces a low-complexity and accurate semiblind channel estimation technique for MIMO-OFDM systems.

Chapter 5 investigates the capacity of iteratively estimated doubly-selective channels when Linear Minimum Mean-Squares Error (LMMSE) estimators are used. Lower bounds on the capacity are found. These bounds are used in an EXIT analysis to predict the performance of an iterative receiver. The method can be used to design such receivers.

Chapter 6 considers the problem of generating secret keys from channel estimates and explores the secret key capacity for realistic channel measurement techniques including half-duplex transmission with long transmit blocks.

Chapter 7 concludes the dissertation by summarizing the research, the contributions and pointing to the future work.

Chapter 2

Background

2.1 General Characteristics of Radio Channel

In a wireless system, the signal may reach the destination via different propagation paths, each with distinct amplitude and delay characteristics. Different propagation delays cause different phase shifts of the signal components, creating constructive or destructive interference. For instance, at a carrier frequency of 2GHz, just a 10cm movement may turn a constructive addition to a destructive one, attenuating the signal at the receiver [115]. The phase shift depends on the relative locations of the transmitter, receiver, and any objects in the environment interacting with the radio signal. Therefore, the overall signal strength will change with time if any moving object is involved. Small-scale fading is described as the variation of signal strength due to movements of the mobile station over distances as short as a fraction of the wavelength. The movement leads to a shift in the received frequency, known as the Doppler shift [115]. The shift can be compensated in the receiver. However, the interference between the signal components creates small-scale fading. The Doppler frequency measures the rate of change of the channel and is proportional to the relative velocity of the receiver. This type of fading is captured by fading models such as Rayleigh or Rician models. The Rayleigh model suits rich scattering environments where a large number of scatterers contribute to the received signal. Rayleigh fading is created when there is no line-of-sight (LOS) propagation path and the channel gains from all directions to the antenna are identically and independently distributed complex Gaussian random variables (RV's). The channel gains for this model are complex Gaussian RV's with zero-mean, their magnitude follows a Rayleigh distribution, and their phase is uniformly distributed over $[0, 2\pi]$ [22]. If a dominant path exists, the likelihood of deep fades becomes much smaller and a Rician probability density function (PDF) is used. In this case, the impulse response has a non-zero mean component, e.g., due to the line of sight path. A model with more degrees of freedom is the Nakagami model [28]. The amplitude of the sum of multiple i.i.d. Rayleigh-fading signals follows a Nakagami distribution. This model fits best for fading channels of large delay spreads, with multiple independent clusters of reflected radio waves, such as urban radio channels [159].

In addition to small-scale fading, the amplitudes of the received signal via LOS or Non-Line-of-Sight (NLOS) paths may gradually vary over long distances (a few meters to several hundreds of meters), for example, when an obstacle creates a shadow on the path. This phenomenon is known as "shadowing", causing *large-scale fading* [115].

When several propagation paths with different delays exist between the transmitter and receiver, the duration of the radio channel's impulse response may be longer than a symbol period if the relative delay differences are larger than the symbol period. The channel impulse response in these systems is not a single impulse, but rather is spread over time [132]. As a result, the signal from one symbol affects the reception of the following symbols. This phenomenon is called *inter-symbol interference* (ISI). The existence of multiple propagation paths and signal reflections from fixed and moving objects like mountains, buildings and vehicles cause selectivity both in the time and frequency domain. Such a channel is called Doubly-selective channel (DSC). A frequency-selective channel has different gains for different frequency components, and thus, distorts the signal. A time-selective channel has different gains over different time instances. In broadband systems with high symbol rate, frequency-selectivity is mainly due to the different delays of the propagation paths, whereas time-selectivity is due to the mobile or objects moving in the propagation environment.

Delay Spread and Coherence Bandwidth

Delay spread and coherence bandwidth characterize the signal dispersion in time. Coherence bandwidth is defined as the frequency width over which the channel response is well-modeled as being constant [132, p. 164]. It is inversely proportional to the delay spread which is defined as the difference between the delay of the longest path and that of the shortest path. The *excess delay* of a path is defined as the time difference from the shortest path delay to the longest path.

A *flat-fading* channel (or *narrow-band* channel) is one where the coherence bandwidth is greater than the signal bandwidth (or equivalently, the symbol period greater than the delay spread). In frequency-selective fading channels, the signal bandwidth is larger than the coherence bandwidth.

Doppler Spread and Coherence Time

The channel fading rate is determined by the mobile station's speed and measured by Doppler spread and coherence time. When a sinusoidal signal of frequency f_0 is sent over a fading channel, the received signal will have frequency components over the range $f_0 - f_d$ to $f_0 + f_d$, where f_d denotes the Doppler shift. The amount of frequency dispersion is a function of the relative velocity and the angle of the receiver. Doppler spread B_D describes the degree by which the spectrum is broadened, while coherence time T_C represents the time interval over which the channel is considered to be unchanging. A rule of thumb relationship states that $T_C = 0.4/f_m$ [132, p. 165], where $f_m = v/\lambda$ is the maximum Doppler shift with v and λ denoting the velocity and the wavelength, respectively.

A *fast-fading* channel is identified by high Doppler spread, where the channel gains are uncorrelated after relative delays of greater than a one hundred symbol periods. In general, while efficient near-optimal estimators for slow-fading channels have already been proposed in the literature [166], the design of channel estimators for fast-fading channels has been a more challenging problem. This problem has been tackled in the literature [92,101]. However, low-cost high accuracy estimators suitable for higher-order modulation schemes used in high data-rate devices remained to be explored.

A well designed wireless system must consider the above-mentioned factors to achieve the best performance-cost compromise. In this research we explore channels where the coherence time is on the order of the symbol duration and where the coherence bandwidth is shorter than the signal's bandwidth. In the next section, we will review some physical characteristics of the most common vehicular wireless network.

Radio Channel Model: Single-Input Single-Output

A single-input single-output radio channel can be modeled as a causal linear time varying filter with input $s_c(t)$, output $y_c(t)$ and time-variant impulse response $g_c(t;\tau)$ at time t to an impulse at time $t - \tau$, related as

$$y_c(t) = \int_{-\infty}^{t} s_c(\tau) g_c(t; t-\tau) \mathrm{d}\tau + v_c(t) = \int_{0}^{\infty} s_c(t-\tau) g_c(t;\tau) \mathrm{d}\tau + v_c(t) \tag{2.1}$$

with $v_c(t)$ being the measurement noise. Typically, for fixed τ, $g(t;\tau)$ is a wide-sense stationary random process with respect to time variable t. If it is also uncorrelated with respect to delay variable τ, we have a wide-sense stationary uncorrelated scattering (WSSUS) channel. The correlation function of a WSSUS channel is invariant over time. Further, the channel gains for different propagation delays are uncorrelated. Real radio channels do not completely follow WSSUS model, as their statistics varies with time. However, the WSSUS model can still be used with acceptable precision for times periods of up to a sizable fraction of a second which is suitable for analyzing most modern wireless systems.

To apply digital signal processing techniques, the analog signals are sampled with a period of T_s at times $t = nT_s$. A doubly selective multipath channel can then be modeled as a linear time-varying FIR filter with $L + 1$ taps, where the largest delay is L sample periods. Let $g(n;l)$ denote the sampled time-varying channel's response at time n to a discrete-time impulse applied at the discrete time $n - l$. The function $s(n)$ gives the symbols transmitted at times n. The vector function $\mathbf{s}(n) = [s(n)\ s(n-1) \cdots s(n-L)]^T$ gives the $L+1$ most recent symbols at time n. The discrete-time signal at the receiver input can be expressed as,

$$y(n) = \sum_{l=0}^{L} g(n;l)s(n-l) + v(n) = \mathbf{g}^T(n)\mathbf{s}(n) + v(n) \tag{2.2}$$

for $n = 1, 2, \ldots, N$, where $v(n)$ denotes the Gaussian zero-mean complex white noise with variance σ_v^2, and the channel impulse response at time n is given by,

$$\mathbf{g}(n) := [g(n;0)\ g(n;1) \cdots g(n;L)]^T. \tag{2.3}$$

Stacking the signal samples into vectors of size N defines by$\mathbf{s} = [s(1) \ \ \cdots \ s(N)]^T$, Eq. (2.2) can alternatively be written as,

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{v} \tag{2.4}$$

where $\mathbf{H}$ is the matrix representation of the convolution operation in (2.2). The entries of $\mathbf{H}$ are either $g(n;l)$ or zero. In most cases of interest, $|g(n;l)|$ at a given instant n can be assumed to follow a Rayleigh distribution. In a rich scattering environment with the maximum Doppler frequency f_d, the correlation between the channel gains follows a model introduced by Clarke and Jake [81] given by,

$$E\left[g(n,k)g(n+m,r)^*\right] = P_l\mathcal{J}_0(2\pi f_dT_sm)\delta_{kr}, \tag{2.5}$$

where $\mathcal{J}_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind, P_l is the mean power of the l-th propagation path, and δ_{kr} is the Kronecker delta function [115], where $\delta_{kr} = 1$ for $k = r$, and is zero otherwise. The Jake's model assumes that the Doppler shift as well as the power of channel paths are constant, and a large number of interacting objects are distributed uniformly around the mobile station [115]. The channel's power spectral density (PSD) is the Fourier transform of the autocorrelation function in (2.5) and takes the form of a U-shape curve given by

$$S_{gg}(f) = \begin{cases} \frac{1}{\pi f_D\sqrt{1-(f/f_D)^2}}, & |f| < f_D; \\ 0, & \text{otherwise}, \end{cases} \tag{2.6}$$

where $f_D = f_dT_s$ is the Doppler frequency normalized to the sampling rate.

A powerful tool for analyzing and modeling band-limited channel gain processes with only a small number of parameters is the basis expansion model (BEM). BEM's are commonly used to describe the sequence of varying channel gains as the weighted sum of time-domain basis functions. Given a BEM period of T_p samples and a set of Q basis functions $b_q(n)$, $q = 1, \ldots, Q$; $n = n_0, \cdots, n_0 + T_p - 1$, the BEM representation of the channel impulse response $g(n;l)$ for a fixed delay l is described as,

$$g(n;l) = \sum_{q=1}^{Q} h_q(l)b_q(n), \tag{2.7}$$

for $n = n_0, \cdots, n_0 + T_p - 1$, where the weights $h_q(l)$ are called the BEM coefficients. The parameters T_p and Q are usually chosen as a compromise between complexity and performance. As the channel gains are usually highly correlated over time, one has $Q \ll N$; that is, the channel gain sequence can be characterized with much fewer parameters. The estimation problem is reduced to tracking the BEM coefficients over time. For channel tap l, let $\mathbf{g}_l := [g(n_0;l) \cdots g(n_0+T_p-1;l)]^T$ and $\mathbf{h}_l := [h_1(l) \cdots h_Q(l)]^T$ denote the channel gain vector and the BEM vector, respectively, when $l = 0, \cdots, L$.

Let $\mathbf{g} := [\mathbf{g}_0^T \cdots \mathbf{g}_L^T]^T$ and $\mathbf{h} := [\mathbf{h}_0^T \cdots \mathbf{h}_L^T]^T$. The matrix form of (2.7) is written as,

$$\begin{aligned} \mathbf{g}_l &= \mathbf{E}\mathbf{h}_l, & (2.8) \\ \mathbf{g} &= \mathcal{B}\mathbf{h} & (2.9) \end{aligned}$$

where $\mathbf{E}$ is the BEM matrix with entries $(\mathbf{E})_{m,q} = b_q(n_0 + m - 1)$, $\mathcal{B} := \mathbf{I}_{L+1} \otimes \mathbf{E}$, and $\otimes$ denotes the matrix Kronecker product.

Using (2.2) and (2.7), the received signal is given as,

$$y(n) = \sum_{l=0}^{L} \sum_{q=1}^{Q} h_q(n) b_q(n) s(n-l) + v(n), \tag{2.10}$$

for $n = n_0, \cdots, n_0 + T_p - 1$. Although the BEM coefficients are constant within a BEM block, they may vary between the blocks. Therefore, we sometimes use $\mathbf{h}(n)$ to denote this time dependency of the BEM vector. Using (2.2) and (2.10), the vector of channel gains for different delay taps at the discrete time n denoted with $\mathbf{g}(n) := [g(n;0) \cdots g(n;L)]^T$ can be written as,

$$\mathbf{g}(n) = \mathbf{B}(n)\mathbf{h}(n) \tag{2.11}$$

where $\mathbf{B}(n) := \mathbf{I}_{L+1} \otimes \mathbf{E}(n,:)$, with $\mathbf{E}(n,:)$ denoting row n of the BEM matrix.

A popular and analytically tractable model to describe a vector of varying channel gains is the complex-exponential basis expansion model (CE-BEM). Since the channel gain process is band-limited to $f_D \ll 1/2$, the size of the CE-BEM vector is much smaller than that of the channel gain vector. For a CE-BEM, $b_q(n) = (1/\sqrt{T_p})e^{j\omega_q n}$. The channel impulse response $g(n;l)$ can be expressed as,

$$g(n;l) = \frac{1}{\sqrt{T_p}} \sum_{q=1}^{Q} h_q(l) e^{j\omega_q n}, \tag{2.12}$$

where $\omega_q := (2\pi/T_p)[q - (Q+1)/2]$ assuming that Q is an odd integer, when the number of basis functions is bounded by $Q \geq 2\lceil f_d T_p T_s \rceil$.

Other BEM's have also been employed in the literature to describe a varying band-limited channel gain process. The discrete prolate spheroidal sequences (DPSS's) are finite sequences whose spectrum is also maximally concentrated over a limited frequency band [145]. The columns of the BEM matrix are the Q eigenvectors associated with the largest eigenvalues of the matrix $\mathbf{C}$ defined as $\mathbf{C}_{n,m} = \sin[2\pi(n -$

$m)f_D]/[\pi(n-m)]$. DPSS BEM is used by Movahedian and McGuire [118] to estimate a fast-fading radio channel. The Karhunen-Loève Transform (KLT) BEM exploits the autocorrelation function of the channel gains to create a set of uncorrelated BEM coefficients [147]. It is the optimal mapping in the sense that the mean-square error of the (truncated) BEM representation is minimized [65]. The KLT basis functions $b_q(n)$ are the Q eigenvectors of the channel autocorrelation matrix $\mathbf{R}_{gg} = E[\mathbf{g}\mathbf{g}^H]$ associated with the largest eigenvalues.

In most of this research, a CE-BEM is used to represent the channel gain process. Since the CE-BEM coefficients are the first Q coefficients of the fast Fourier transform (FFT) of the signal, they can be computed using the FFT techniques, and the theory of the FFT may conveniently be used for the analysis of the model.

After reviewing the physical properties of wireless channels, we now discuss channel capacity as an important performance measure for a communication channel.

2.2 Channel Capacity

In his seminal paper [141], Shannon established that by using infinite-length codes, a noisy channel can transfer information up to a maximum rate called *capacity*, with probability of error in receiving information approaching zero. Channel capacity is an important basic performance metric in the analysis and design of communication systems. The past few decades have witnessed the effort put into designing practical codes to approach the channel capacity. Turbo codes [21] and low-density parity check (LDPC) codes [57] are capacity-approaching iterative coding schemes widely used in mobile systems.

The definition of channel capacity uses the concepts of entropy and mutual information. Entropy measures the uncertainty of an RV. For the discrete RV X taking values in set $\mathcal{X}$, entropy is defined as [41]

$$H(X) := -\sum_{x\in\mathcal{X}} p(x)\log p(x) \tag{2.13}$$

where $p(x)$ denotes the probability density function of X. The conditional entropy of X conditioned on the discrete RV Y taking values in $\mathcal{Y}$ is defined as

$$H(X|Y) := -\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} p(x,y)\log p(x|y) \tag{2.14}$$

with $p(x|y)$ denoting the conditional probability density function of X given Y. $H(X|Y)$ represents the average uncertainty of X when Y is known. The mutual information between the RVs X and Y, denoted by $\mathcal{I}(X;Y)$, measures the amount of information in X about Y, and is defined as

$$\mathcal{I}(X;Y) := H(X) - H(X|Y) \tag{2.15}$$

Channel capacity is defined in terms of the mutual information between the channel input and output. Let RV's X and Y denote the transmitted and received symbols, respectively. The capacity of a channel with finite-dimension input process X^N denoting a sequence of N inputs to the channel, and output process Y^N defined likewise, is given by [50],

$$C_e = \lim_{N\to\infty} \sup_{p_X} \frac{1}{N}\mathcal{I}(X^N;Y^N), \tag{2.16}$$

when the limit exists, where $\mathcal{I}(X^N;Y^N)$ denotes the mutual information between random vectors X^N and Y^N, and $\sup\limits_{X^N}$ stands for the supremum of the mutual information taken over all possible choices of p_X, the probability density function of X^N.

For the discrete-time fading channel described in (2.4) with the receiver channel state information (CSI), the capacity is given by [22],

$$C_e = \lim_{N\to\infty} \sup_{P_s\le 1} \frac{1}{N}\mathrm{E}\left[\log\det\left(\mathbf{I}_N + \frac{1}{\sqrt{\sigma_v^2}}\mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H\right)\right] \tag{2.17}$$

with $\mathbf{H}$ and $\mathbf{s}$ defined as in (2.4), $\mathbf{R}_{ss}$ is the autocorrelation matrix of $\mathbf{s}$, P_s denotes the mean power of information symbols, and the expectation is taken over all possible realizations of the random channel.

It has been shown that the input signal can be extracted from the channel output even if the channel is unknown to the receiver and no pilots are sent by the transmitter [61]. In the method proposed by Godard [61], the parameters of an adaptive equalizer are iteratively updated by minimizing a special cost function using gradient method without estimating the channel. Differential modulations do not require the channel estimates, but cause elevated BER [62]. In differential modulation, the data bits are encoded into the relative phase of the consecutive symbols. Assuming that the phase of the channel gain is invariant from one symbol to the next, the phase change in the received signal is mainly due to the signal. This approach does not need the knowledge of the channel gains, and is used in non-coherent detection where the channel is not explicitly estimated by the receiver. Drawbacks with this strategy include slow convergence and the possibility of local minima resulting in detection errors [49]. In [179], the BER performance of differential detection is compared to that of coherent detection in the presence of channel estimation error. It is shown that with accurate channel estimation, the coherent technique outperforms the differential detection. To avoid these drawbacks and simplify the receiver design, coherent symbol detection is performed using the channel estimates [101]. To perform coherent detection, the receiver has to estimate the time varying channel gains through pilot-based or (semi-)blind techniques. The accuracy of the channel estimator significantly affects the capacity [105, 119]. The capacity in this case is defined as [105],

$$C_e = \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sup_{P_s \le 1} \mathcal{I}(\mathbf{y}, \hat{\mathbf{H}}; \mathbf{s})\ |\mathbf{H}\right], \tag{2.18}$$

where $\hat{\mathbf{H}}$ denotes the estimated gain matrix and the expectation is taken over all possible realizations of the random channel. Define channel estimation error as $\tilde{\mathbf{H}} := \mathbf{H} - \hat{\mathbf{H}}$. Then $\mathbf{y} = \hat{\mathbf{H}}\mathbf{s} + \tilde{\mathbf{H}}\mathbf{s} + \mathbf{v}$. The uncertainty in $\mathbf{s}$ given $\mathbf{y}$ is due to both the channel noise $\mathbf{v}$ and the term $\tilde{\mathbf{H}}\mathbf{s}$ due to the error in the knowledge of channel impulse response. Therefore, a larger channel estimation error corresponds to a higher uncertainty on $\mathbf{s}$ when $\mathbf{y}$ is known. A poorly estimated channel will reduce $\mathcal{I}(\mathbf{y}, \hat{\mathbf{H}}; \mathbf{s})$, and thus, the capacity. A closed-form expression for capacity of an estimated channel remains an open problem. The capacity of purely pilot-based estimated channels using linear minimum mean-squares error (LMMSE) channel estimators has been studied by Ma, Giannakis, and Ohno [105], where an optimal pilot scheme to maximize a lower bound on capacity was proposed. This bounds will be used in this research to evaluate the capacity of iterative receivers, where the detected data symbols are iteratively used by the channel estimator to improve the accuracy. Agarwal and Honig [4] studied the capacity of a block fading channel with partial feedback is studied in a non-iterative setting. The trade-off between transmission rate and channel estimation error is considered by W. Zhang, Vedantam, and Mitra [178]. This work demonstrated that higher transmission rates give rise to higher channel estimation errors, and establishes a relationship between the channel capacity and the maximum allowable channel estimation error (the so-called "capacity-distortion function"). The formulation and derivations however, are limited to finite-alphabet signals and non-iterative channel estimation schemes, and extending the approach to doubly-selective continuous-state channels is difficult.

2.3 Channel Estimation

Coherent detection requires that the channel be estimated. Pilot-assisted channel estimation techniques periodically insert known symbols between data symbols in a time-multiplexed fashion [31, 32, 105, 114], or superimpose the pilots on data symbols [56, 59, 158, 163]. The pilot overhead in time-multiplexed pilot-assisted channel estimation depends on the fading rate. According to the Nyquist sampling theorem, $2f_D$ pilots per channel path per data symbol are required to uniquely identify the channel impulse response. Pilot overhead of pilot based methods may lead to significant throughput loss in fast-fading environments or when a large number of antennas or channel taps is involved. For example, with the optimal pilot scheme proposed by Ma et al. [105] where a lower bound on capacity is maximized, the pilot overhead may exceed 50% of the bandwidth in the scenarios with many channel taps and high fading rates.

A formal analysis of pilot-assisted estimation was first presented by Cavers [31] where a Wiener filter was used to minimize the estimation error in a flat-fading scenario, and the trade-off between estimation accuracy and bandwidth efficiency was studied. The idea was that more pilot signaling would reduce the useful bandwidth but increases the estimation accuracy. This technique was extended to frequency-selective channels where the superiority of pilot-based schemes in terms of the BER performance, over non-coherent detection was shown [32].

Pilot design may be optimized based on various criteria, such as bounds on capacity, bit error rate (BER) or LMMSE. The optimal design determines power allocation between pilots and data symbols, pilot placement in the transmit stream and the number of pilot symbols. Crozier, Falconer, and Mahmoud [42] proposed least-squares (LS) filtering for estimating frequency-selective channels. The optimal pilot sequence is found by minimizing the LS error. The case of doubly-selective channels was considered by Ma et al. [105], where the optimal sequence was found that maximizes a lower bound on capacity. It was shown that the optimal pilot pattern would consist of a non-zero pilot, and null symbols (zeros) before and after the pilot. The high-SNR regime is studied by Kannu and Schniter [87], where a pilot scheme to maximize the spectral efficiency is proposed. Training policy in multiple-antenna communications was explored by Marzetta [109], Hassibi, and Hochwaldand [71] for BLAST (Bell

Labs Layered Space-Time). BLAST is a MIMO technology which exploits the spatial diversity for reliable communications in broadband systems. Marzetta [109] employed a method to evaluate the effect of channel estimation error on the outage capacity of a Rayleigh flat fading channel. The outage capacity was used to design the optimal pilot scheme to maximize the overall transmission rate. Hassibi and Hochwaldand [71] proposed a training policy for multiple-antenna systems to maximize a lower bound on the capacity. It was shown that for optimal pilot allocation over the pilot and data symbols, the number of pilots should equal the number of transmit antennas.

Channel estimation based on the *input-constrained* capacity maximization was considered by Baltersee, Fock, and Meyr [16] for the case of time-selective flat fading channels. The input-constrained capacity is used to refer to channel capacity when the symbol alphabet is discrete equiprobable rather than Gaussian. It was shown that the mutual information is a function of the estimation LMMSE. Also, the optimal pilot rate in the sense of channel capacity was found to always be above the Nyquist rate.

Pilot-assisted channel estimation may require a significant portion of the bandwidth to be allocated to pilots, specially in channels with high fading rates or large numbers of paths between the transmitter and the receiver. Semi-blind and blind techniques exploit the properties of the channel and input signals to reduce the necessary pilot rate below the Nyquist sampling rate of the channel gain processes and save the bandwidth. By allocating all the bandwidth to data symbols, the spectral efficiency increases, but this often translates into much higher computational cost to obtain accurate enough channel estimates to support useful data reception. The second order statistics of the received signal along with the cyclo-stationarity of the input are used by Tong, Xu and Kailath [153] to identify the channel without training. A class of blind/semi-blind techniques called *subspace* methods decompose the output auto-correlation matrix to obtain the signal or noise subspace [33,117,152,154,167]. These subspaces correspond to the largest and smallest eigenvalues of the auto-correlation matrix. The signal subspace is spanned by the channel impulse response matrix. As such, the channel matrix may be obtained up to a phase ambiguity [117]. Singular value decomposition to decompose the subspaces may be computationally inefficient due to the large dimensions of the auto-correlation matrix.

If standard channel estimation methods are used with MIMO, the pilot overhead increases with the number of transmission antennas to the point that a significant portion of the available bandwidth is consumed by pilots. Blind and semiblind techniques can significantly increase spectral efficiency compared to standard pilot-based channel estimation where the pilot rate must be above the Nyquist sampling rate of the channel gain process. The semiblind techniques introduced by Y. Chen and Song [33, 35] apply a linear precoding before transmission to create correlation between symbols, which allows for channel estimation without pilots but also makes symbol detection more difficult and prone to errors [35]. As the channel state is estimated from estimates of the covariance of the received signal, for accurate channel estimation these methods require the channel state to be static over a long period of time. This issue makes the method inapplicable to fast-fading channels where the channel coherence time is on the order of only tens of symbols. Moreover, the computational cost of these techniques is higher than the pilot-assisted methods.

A blind and semiblind technique is presented by Yu, B. Zhang and P. Chen [176] where the statistics of the blocks of the received signal are used to compute the magnitudes of the channel gain processes for different propagation delays. Sparse pilots are then used to resolve the phase ambiguity and obtain the final channel process estimates before data detection. Since the channel is assumed to be invariant over a block, channel variations within a block are not captured. Therefore, this approach works well only for very slow fading channels. However, at the fading rates encountered in mobile radio channels, an unwanted error rate floor is hit.

2.3.1 Iterative Channel Estimation

With purely pilot based estimation, the channel gains for times between the pilots are estimated using interpolation. Accurate pilot-based estimation in fast-fading channels calls for a large amount of pilots which leads to low spectral efficiency, especially in channels with a large number of taps [105]. Iterative channel estimation employs the detected data as virtual pilots to enhance the channel estimation accuracy, hence reducing the pilot overhead needed for a given accuracy. This approach to channel estimation has widely been used with *turbo* channel estimation which perform channel estimation, symbol detection and data decoding in an iterative manner. At each iteration, soft information on coded bits is exchanged between the equalizer and decoder until convergence is reached. Turbo equalization reduces the receiver complexity as compared to the optimal method of building a large trellis of the channel and decoder states and then performing a maximum *a posteriori* (MAP) sequence detection. Iterative equalization is inspired by the work of Berrou on turbo codes [21] which reduced the complexity of decoders for capacity approaching codes. It was first employed by Douillard et al. [54] to improve the performance of a coded modulation system over a *known* frequency selective channel using a MAP equalizer and a MAP decoder. The BER performance of this receiver was shown to be close to that of the optimal method at a much lower computational cost. However, the cost of a MAP equalizer could still be a burden with higher-order constellations and/or large delay spreads due to the the large number of the trellis states. Laot, Glavieux, and Labat [97] replaced the hard-decision MAP equalizer with a soft ISI canceler, to increase the accuracy. Based on the work of Douillard et al. [54], MMSE-based soft-input-soft-output equalizers were proposed by Tüchler, Singer, and Koetter [161] with significant complexity reduction.

Extending the turbo principle to channel estimation, Davis, Collings, and Hoeher [46] studied the problem of joint channel estimation and equalization using a MAP equalizer for doubly-selective channels where an expanded trellis was employed to include the extra memory required to estimate the channel. To overcome the complexity of trellis-based equalizers, the use of linear adaptive filters or variations of the Kalman filter has been considered for coded modulation systems [92, 101]. An extended Kalman filter [144] is used by Li and Wong [101] for joint channel estimation and symbol detection. The channel gain process is characterized by a first order auto-regressive (AR) model. AR models and CE-BEMs have also been employed in several other works [15, 60, 83, 156]. A downside with using the low-order AR models is the existence of an error-rate floor at high SNRs [95], due to their imperfect representation of the channel's time evolution [15].

Based on the method of Li and Wong [101], a more accurate channel model using CE-BEM was employed by H. Kim and Tugnait [92]. The superiority of block-wise CE-BEM over symbol-wise AR models in modeling and tracking fast-fading channels has been well investigated by Tugnait, He, and H. Kim [157] for different adaptive algorithms and by H. Kim and Tugnait [93] for MIMO channels. Tugnait et al. [157] explored adaptive blockwise tracking of a doubly-selective channel using a KF and recursive least squares method. The time variations of the channel over a block are captured by a CE-BEM, whereas the evolution of the BEM coefficients between the blocks are represented by an AR model. Compared to the AR channel models, this method reduces the modeling mismatch, resulting in performance improvement in fast-fading environments.

In fast-fading environments, these methods typically suffer a BER floor at higher

SNRs, which prevents their application to higher-order modulations which must operate at these SNR levels [112]. Higher order modulations are more prone to the channel estimation error, hence, more accurate channel estimators are needed. Enhanced performance requires prohibitively large KF state vectors, particularly for multipath fast-fading radio channels, posing a computational complexity problem. For slower fading channels with normalized Doppler frequencies of less than 0.001, a near optimal and efficient channel estimation approach was proposed by Wan, McGuire, and Dong [166] for OFDM, but it does not scale well to fast-fading channels since it is based on the assumption that the channel gains are constant over a period of 128 samples. Fast channel variations destroy this assumption required by the popular frequency-domain equalization schemes used by the OFDM and single-carrier methods proposed in much of the literature. The invariant gain assumption implies that the channel is almost invariant When these methods are applied to fast-fading channels, they exhibit an unacceptably high error floor [136].

Iterative channel estimation methods proposed by Movahedian, McGuire, and Wan [120, 166] for single transmit and receive antenna systems can be extended to MIMO-OFDM case. Unfortunately, the technique used by Movahedian and McGuire [120] requires signal blocks with durations many times the coherence time of the channel to guarantee good performance which leads to unacceptable latency at the normalized fading rates of 10^{-4} and lower considered in most of MIMO literature. The approach of Wan and McGuire [166] works well in SISO case, but the number of pilots required to estimate the channel grows unacceptably large for MIMO systems with many antennas. An impediment to using the aforementioned single-antenna techniques is that they require time diversity of the radio channel within the period of single processing block. Time diversity is achieved by sending different bits of the codeword at different times. As the channel varies with time, only part of the codeword is likely to be transmitted during a time period when the channel gain is low and thus corrupted by the noise and fading effects of the channel. If too many bits of the codeword are corrupted during the transmission, the decoder is not able to recover the erroneous bits. To guarantee the required time diversity, these methods require unacceptably long signal blocks incurring undesirable receiver latency. This is much less of an issue in MIMO channels where the spatial diversity may be exploited to compensate for a short processing block to obtain the same diversity as a long processing block in a SISO receiver. An iterative semiblind estimation approach is proposed by K. J. Kim, Tsiftsis, and Schober [94] for LDPC coded MIMO-OFDM,

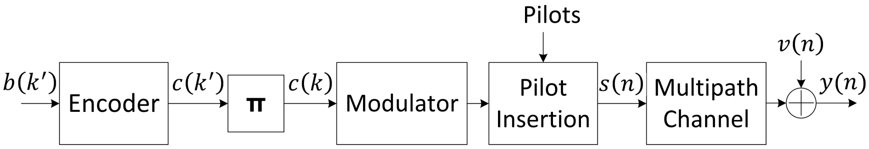

Figure 2.1: Transmitter

where a recursive-least-squares (RLS) algorithm is employed to estimate the channel gains for each fading block. While the method works well with reasonable complexity for quasi-static or very slow fading channels, its performance deteriorates in fast channel variations as the forgetting factor in the RLS algorithm cannot be tuned well for the estimator to track the channel variations [122].

2.3.2 System Model

The iterative approach to channel estimation, symbol detection and data decoding as considered in this thesis assumes the following models for transmitter and receiver.

Transmitter

A bit-interleaved coded modulation system transmitting over a time-varying fading channel is considered here as shown in Fig. 2.1. We use the notation for single-carrier signaling and samples from the work of H. Kim and Tugnait [92]. A block of independent data bits $\{b(k'), k' = 1, 2, \ldots, N_d\}$ is encoded by a convolutional or LDPC encoder with code rate R. The encoded sequence $c(k')$ goes through a bit-wise random interleaver $\Pi(\cdot)$ of length N_i, generating the interleaved coded sequence $\{c(k), k = 1, 2, \ldots, N_i\}$. The resulting interleaved data are modulated according to some constellation χ, mapping every N_{mod} bits into a constellation point.

Receiver

The receiver performs iterative channel estimation, data detection, and decoding. At each iteration, the channel estimator uses the data estimates from previous iteration to enhance the estimation accuracy. The channel estimates are then used by the equalizer to detect the data symbols. Data symbol estimates are demodulated and used by the decoder to generate soft data bits. Using soft information rather than hard information, results in a performance improvement. Soft data bits carry reliability information about the decision made by the decoder on each decoded bit. New estimates on data symbols are calculated using the soft bits and fed back to the equalizer and channel estimator.

2.3.3 Optimal Linear Channel Estimator Bound

For the continuous-time single-path channel estimation problem, assuming the signal is uncorrelated with the noise, the mean-square-error (MSE) of the optimum Wiener filter is given by

$$W := \int_{-\infty}^{\infty} \frac{S_{gg}(f) S_{vv}(f)}{S_{gg}(f) + S_{vv}(f)} \, \mathrm{d}f, \tag{2.19}$$

where $S_{gg}(f)$ and $S_{vv}(f)$ denote the PSD of the channel gains and noise, respectively [126]. Since the channel is band limited, the above bound applies also to discrete-time filters. For the case of a Rayleigh fading channel with the Jakes' model, the PSD of the channel gains is described by (2.6).

A good approximation to (2.19) at high SNRs, where $S_{gg} + S_{vv} \approx S_{gg}$, can be obtained as

$$W \approx 2 f_D \sigma_v^2 \tag{2.20}$$

where σ_v^2 is the noise variance. The error of this approximation is less than 1% for SNR's greater than 20 dB.

The study of channel estimation techniques and the capacity of estimated channels bears significant implications for communication security. As it will be shown, the ability of the legitimate parties to establish a secure communication channel depends on the channel capacity. An accurately estimated channel not only is crucial for reliable communication, but also serves as an abundant source of secret keys for secure communication. This fact motivates the study of security aspects of data communications in the physical layer in the following section.

2.4 Physical-layer security

Channel estimates can be exploited to generate secret keys used to encrypt the data transmitted over a public channel. This section lays the foundations for secret key generation from the channel impulse response as a common source of randomness between the communicating parties.

2.4.1 Basic Concepts

We study the security aspects of data communication systems exposed to adversaries with unlimited computational power. This physical-layer approach to security differs from the computational complexity approach which hinges on the assumption that performing certain computing tasks (such as prime factorization of large numbers) require much computing power or time to be feasible. The computational complexity approach employs the well-known methods of Diffie and Hellman [48] or Rivest, Shamir, and Adleman (RSA) [135]. Diffie-Hellman algorithm is employed to establish a secret key between two parties. The secret key is then used to secure the communication. With RSA, key distribution is performed by a trusted third-party. Overall, RSA is computationally expensive compared to the methods discussed here, making it less viable for mobile devices. Elliptic Curve Cryptography (ECC) is computationally faster, but it may still require accelerator hardware to run on small devices [63]. The key agreement mechanism in ECC is similar to Diffie-Hellman. For ECC, the key size to provide a certain level of security is smaller than that of an RSA system. In information-theoretic physical-layer security the computational complexity is avoided. Rather, the security is based on the solid frame of information theory and the security results are mathematically provable [111]. Information-theoretic security is concerned with unconditional security.

One drawback with information theoretic security comes from the assumption made about the noise levels in the system which may lead to either over-optimistically high or extremely low secrecy capacity [24]. This is the case when, for example, the adversary's observation of the signals in the communication channel is not as contaminated with noise as it was incorrectly assumed to be. If a security protocol relies on the assumption of a lower noise level in the signals received by the legitimate communicating parties, the communication may not be secure. When secret keys are generated using the channel estimates, the key rate depends super-linearly on the fading rate (see Chapter 6), which may be too low for many applications.

An unconditionally secure system was first introduced by Shannon [140] and involves the concept of *perfect secrecy*. Consider a message M encoded to a codeword X by a transmitter Alice, received as Y by a legitimate receiver Bob, and intercepted as Z by an eavesdropper Eve, where Z may be different from X due to reception errors on Eve's part. Perfect secrecy refers to the condition where Eve is not able to extract any information from Z regarding M, that is $H(M|Z) = H(M)$, where

$H(M)$ is the entropy of the message and $H(M|Z)$ denotes the conditional entropy of M conditioned on Z. The conditional entropy $H(M|Z)$ is called the eavesdropper's equivocation, representing the Eve's uncertainty about the message after observing Z. For perfect secrecy, the mutual information between M and Z defined as $\mathcal{I}(M;Z) := H(M) - H(M|Z)$ is zero [41]. The entropy of the message measures the information content of the message, whereas the mutual information between M and Z is a measure of the amount of information about M contained in Z. Under perfect secrecy, the codeword is statistically independent of the message given Eve's observation, implying that, knowing Z will not increase Eve's information about the message. The transmitted codeword X is computed by a function of the message M and a shared secret key K which is independent of the message M which is shared by Alice and Bob, the knowledge of which suffices to recover the message by the other party. Shannon assumed that Eve and Bob receive an exact version of the codeword, i.e., $Y = Z = X$, and showed that for perfect secrecy, the secret key must contain as many bits as the secret message which implies that the secret key rate must be equal to or greater than the message's data rate. For shorter keys, the Eve's equivocation is at most $H(K)$ and she will be able to extract some information from the codeword in the sense that $H(M|Z) < H(M)$; Observing Z decreases the Eve's uncertainty about what the message could be by an amount of $\mathcal{I}(M;Z) = H(M) - H(M|Z)$. We will show that iterative channel estimation has significant implications regarding the derivable key rate and the security of the system.

2.4.2 Secrecy Capacity

Shannon's description of perfect secrecy assumes that Bob and Eve receive the same codeword, without any communication error. A more practical conception of secrecy quantifies the maximum rate at which a reliable and secure communication over a broadcast noisy channel is possible. This maximum rate is referred to as the *channel secrecy capacity*. The concept of secrecy capacity was originally introduced by Wyner [170] for a special type of channel, called the degraded wiretap channel (DWTC). Consider a message M coded by Alice to codeword X^n and sent through a discrete memoryless channel to Bob, who receives Y^n. This channel is described by some conditional probability function $p_{Y|X}$ denoting the probability function of the RV Y conditioned on the RV X. Message M is drawn from 2^{nR_1} possible messages. The wiretapper, Eve, receives Z^n through a "degraded channel" described by some conditional probability function $p_{X|Z}$, in the sense that X^n, Y^n and Z^n form a Markov chain [1], denoted as $X^n \rightarrow Y^n \rightarrow Z^n$ [103]. This case where Eve receives a degraded version of the signal received by Bob, does not represent a practical channel, but it greatly simplifies the description of the secrecy capacity in the following. In a wiretapped channel, a rate R is called achievable if there exists a channel code with sufficiently long codewords, that can transmit R bits of message information with vanishingly small probability of error, while maintaining the Eve's equivocation $(1/n)H(M|Z^n)$ at a minimum of R bits. The secrecy capacity is the maximum achievable R. As long as $(1/n)H(M) = R_1 < R$, in the limit as $n \rightarrow \infty$, that is, if the transmission rate is below the secrecy capacity, then there exist *wiretap codes* which ensure that the information leakage rate to Eve represented by $(1/n)\mathcal{I}(M;Z^n)$, goes to zero as the codeword length n goes to infinity. The secrecy capacity C_s for a DWTC is [170]

$$C_s^{DWTC} = \max_{p_X}\{\mathcal{I}(X;Y) - \mathcal{I}(X;Z)\} \tag{2.21}$$

One interesting form of a wiretap channel is that of a fading wireless channel in which instantaneous SNR may change due to channel gain variations. In this case, the secrecy capacity will depend on the fading characteristics such as the channel coherence time, as well as whether the full CSI is available to the transmitter. Most optimistically, when the full CSI of the main channel and Eve's channel is known to the legitimate communicating parties, the capacity can be strictly positive even if the main channel is noisier than the eavesdropper's channel. The key to this remarkable result is that the legitimate receivers can cooperate while Eve cannot get their assistance. The transmitter can adjust its power to the instantaneous SNR of the legitimate receiver with respect to that of the eavesdropper. This strategy for example may only transmit data when Eve's channel is in a deep fade while Bob's channel is not. By modulating the transmit power in accordance with the relative SNR of the main channel with respect to Eve's channel, the capacity of the main channel would exceed that of the Eve's channel [17]. However, this result is more of a theoretical interest than a practical one, because the Eve's channel SNR may not be available to the transmitter. The above-mentioned bounds on secrecy capacity are based on the assumption of one-way communication from Bob to Alice. Maurer showed [111] that if there exists some external common source of randomness, a non-zero secrecy capacity is achievable for channels which would otherwise have a null

[1] RVs X, Y, and Z form a Markov chain if given Y, then X and Z are statistically independent, i.e., $\mathcal{I}(X;Z|Y) = 0$.

capacity.

A more practical wiretap channel than the DWTC where the Eve's received signal may not be a degraded version of Bob's signal, was studied by Csiszár and Körner [43]. Let U denote an auxiliary RV used by the encoder as an additional randomization factor, so that $U \to X \to YZ$. The secrecy capacity of a wiretap channel (WTC) is given by

$$C_s^{WTC} = \max_{p_{UX}} \{\mathcal{I}(U;Y) - \mathcal{I}(U;Z)\}^+ \geq C_s^{DWTC} \tag{2.22}$$

where p_{UX} denotes the joint probability function of U and X, and $\{\cdots\}^+$ indicates that only non-negative values are acceptable. By incorporating U, the channel between U and Z effectively turns into a degraded version of the channel between U and Y. The secrecy capacity is positive if for some U, $\mathcal{I}(U;Y) > \mathcal{I}(U;Z)$, in which case, Eve's channel is said to be *noisier* than Bob's channel. Note that if $X \to Y \to Z$, then Eve's channel is noisier than Bob's channel, but not vice versa. Therefore, the DWTC is a special case of WTC.

2.4.3 Secret Key Generation

Rather than constructing wiretap codes, another strategy to securing a communication is having the communicating parties to generate a shared secret key which is then used to encrypt the data. In this model, Alice and Bob can both measure some common source of information, such as the wireless channel itself. Eve may also observe this source of randomness, but her measurements are inferior to both Alice and Bob. Alice and Bob can publicly discuss their measurements using key agreement protocols to agree on a common key without revealing this key to Eve. The secret key capacity is defined as the maximum rate key bits that Alice and Bob can generate, while keeping Eve almost ignorant about the key. Secret key agreement over a public, noiseless and authenticated channel between Alice and Bob was theorized by Maurer [111], Ahlswede and Csiszar [5]. The key agreement process consists of four phases [24] as follows.

Common randomness establishment: Correlated RVs are observed by Alice, Bob, and Eve. The correlation may be characterized either by a *source model*, where an external source of randomness generates X^n, Y^n, Z^n with joint PDF $p(x^n, y^n, z^n)$, or by a *channel model*, where the channel delivers a noisy version of the signal produced by Alice to Eve and Bob. This model is described with $p(y^n, z^n|x^n)$.

Advantage distillation: In this phase, Alice and Bob gain some advantage, in terms of mutual information about the measured signal, over Eve. Distillation through two-way public transactions between Alice and Bob was first described by Maurer, Ahlswede, and Csiszar [5,111], where upper and lower bound on key rate and secrecy capacity were derived. To see how advantage distillation works, consider the above-mentioned source model and suppose that the mutual information between Alice's observations and Eve's observations is greater than the mutual information between Bob's and Alice's observations, giving no advantage to Alice and Bob over Eve. In this case, any attempt to derive a key solely based on common randomness would fail, as there is more "commonality" between Alice and Eve. However, it can be shown that through public discussion, Alice and Bob are still able to agree on a secret key. The idea is that Alice and Bob pick out only those realizations of (X^n, Y^n) that are highly correlated. Other less-correlated observations are discarded. The decision as to which realization be chosen is communicated between Alice and Bob through public channel.

Information reconciliation: At this stage, error-correction techniques such as low-density parity check (LDPC) codes are employed to provide Bob with an almost error-free bit string about which Eve has only a partial knowledge [19,27,55].

A reconciliation protocol identifies the discrepancies between the Alice's and Bob's observation by exchanging parity checks over the public channel. A simple method is described in [172], where Alice and Bob observe jointly Gaussian random variables X^n and Y^n. Alice first quantizes X^n and then converts the quantized vector to the bit string X_a. The parity check vector for X_a, called the syndrome, is calculated using some LDPC code. The syndrome is sent to Bob over an error-free public channel. Using the syndrome and Y^n, Bob decodes X_a using an error correcting algorithm.

Careful selection of the error-correcting code is crucial for this method to work. If the code is too powerful, it will correct not only Bob's, but Eve's errors, and the key will be revealed to Eve.

Privacy amplification: After reconciliation, Alice and Bob have some information in common about which Eve may still have nonzero knowledge. For a perfect key, the common information between Alice and Bob must be mutually independent of the Eve's knowledge. Privacy amplification establishes a *secret* key between Alice and Bob by extracting a shorter key from the common reconciled key, about which, the eavesdropper would have no information. A common approach to privacy amplification uses one-directional hash functions [18]. Assume that Eve has estimated a secret key $\tilde{S}$ with some bits identical to those of Alice and Bob. Eve, however, would not know the positions of the common bits. Therefore, if a deterministic function, involving shuffling and combining the bits of the key, is applied to the secret keys, Eve cannot determine $f(\tilde{S})$. It has been shown that there exist functions the output of which are equally likely for Eve.

2.4.4 Channel Fading as a Source of Randomness for Key Generation

The principle of channel reciprocity guarantees that Alice and Bob would experience identical channel response when they measure the wireless channel simultaneously or with sufficiently small delays between their measurement times [128]. Furthermore, any adversary located farther than half the wavelength from Alice and Bob, will observe independent fading, meaning that, the mutual information between the channel measured by Eve and that measured by Alice or Bob is limited [5], assuming that the channel response is not predictable due to effects such as shadow fading. Using channel reciprocity to generate secret keys was proposed by Hershey, Hassan, and Yarlagadda [74]. This approach has since gained attention from researchers. Key generation from the phase of a Rayleigh fading channel was investigated by Hassan et al. [70]. Ye, Reznik, and Shah [172] have proposed a method for key generation from the gains of a Rayleigh flat fading channel. The gains as well as the estimation errors are assumed to be i.i.d. with no time correlation. The two parties quantize their continuous gain measurements to generate bit strings, which are encoded with an error-correcting code to manage gain measurement and quantization errors (refer to Section 2.4.3). Afterwards, a hash function is used to generate a key which is completely unknown to Eve. These techniques have been extended to frequency-selective multipath cellular channels by Ye et al. [173]. Assuming i.i.d. channel samples the secret key capacity C_k is evaluated based on an upper bound given as

$$C_k \le \sum_{i=1}^{L} \log\left(1 + \frac{\mathrm{SNR}_l}{2 + 1/\mathrm{SNR}_l}\right) \tag{2.23}$$

where $\mathrm{SNR}_l := P_l/\sigma_e^2$ with P_l and σ_e^2 denoting the power of path l and the power of estimation error, respectively. The above inequality does not take the temporal correlation of the channel samples into account. This temporal correlation reduces the key rate. Further, if the channel impulse responses on different paths are correlated, the key rate decreases, as the overall entropy of the channel gain process is reduced.

This simple inequality also refers to the basic limitation imposed by the estimation error on the secret key capacity of an estimated wireless channel.

If the gain samples are correlated, an orthogonal decomposition algorithm is used to extract uncorrelated samples from correlated gain measurements [138]. If the time interval between observations is shorter than the coherence time of the channel, there would be some correlation between the key strings extracted from successive observations, which is not addressed in the work of Ye et al. [173]. Moreover, in iterative receivers the channel is estimated at much shorter intervals than the coherence time, limiting the application of the method even further. The case of correlated eavesdropper is addressed by Chou, Draper, and Sayeed [37].

In the research by Sayeed, and Perrig [137], the phase of channel gains is estimated by an MMSE estimator and quantized to obtain the secret key. The quantization strategy and transmit power are optimized for minimum energy consumption. Key generation based on the location and duration of channel deep fades is studied by Azimi-Sadjadi et al. [14]. This technique does not require identical channel measurements at the two ends; only matching deep fades would suffice. This method does not exploit the full secret key capacity of the channel as it discards part of the signal envelope which is not in a deep fade. UWB channel pulse response is used by Wilson, Tse, and Scholtz [169] to generate secret keys. The channel gains are assumed to be i.i.d. and unknown and estimated by Alice and Bob. Research by Liu, Draper, and Sayeed [104] showed that the use of channel coefficients in IEEE802.11a OFDM as the key source is feasible.

The radio channel considered in all the proposed techniques is either with i.i.d. gains, or quasi-static. A quasi-static fading channel has constant gains over the whole codeword, in contrast to the slow fading type where the channel gains are allowed to change every block of N symbols [24, p. 194].

Chapter 3

Estimation of Fast Fading SISO Channels

Accurate channel estimation is crucial in higher order modulation systems due to the sensitivity of the demodulator to estimation errors. The achievable capacity of the channel is influenced by the estimation error. Also, when the channel gain estimates are used to generate secret keys by the communicating parties, the key rate is determined by the estimation accuracy.

In this chapter we propose an efficient, low-complexity approach to pilot-assisted fast-fading channel estimation for single-carrier modulation with a turbo equalizer and a decoder. The error performance of the proposed method is close to the the ideal case where the channel is known to the receiver.

3.1 Introduction

We address the problem of estimating a fast-fading channel, with normalized Doppler frequencies as large as 1% of the symbol rate, capable of supporting single-carrier 16-quadrature-amplitude modulation (QAM) and 64-QAM modulation schemes. For lower fading rates, bandwidth-efficient and computationally cheap OFDM channel estimators and frequency-domain equalizers are available. These techniques assume that the channel is quasi-static, i.e., the channel gains remain nearly constant over time periods of about 100 symbols. These methods fail to work at the higher fading rates considered by this research. It is well known that the performance of an OFDM receiver is very sensitive to the Doppler spread with OFDM systems experiencing undesirable error floors at high SNRs when the channel gains evolve significantly over the period of a single OFDM symbol [79].

For higher order modulations, high-precision estimation of the channel is critical since the detector is sensitive to the estimation errors. To keep the receiver computational cost to reasonable levels, the channel estimation and signal equalization tasks are performed separately, using two cascaded low-order KFs. Although the channel gain processes are band limited to the Doppler frequency, the channel estimator's KF output is contaminated with an estimation error, which is not strictly band limited to the Doppler frequency. This chapter demonstrates the use of a zero-phase filter (ZPF) as a smoother to suppress the out-of-band estimation error. The combination of the ZPF with the KF of the channel estimator makes it possible to reduce the estimation error to near the Wiener bound. To get similar performance using just a KF would require the KF to have a state vector of a large order which incurs a significant computational cost. The KF provides an independent estimate of the channel gains for each propagation path, which are smoothed independently. Since the smoothing function is applied to each channel path independently, the method efficiently fits on multicore processors. Standard sequential methods such as Extended Kalman Filter (EKF) implementations (e.g., in the reference [92]), do not map onto multicore processing platforms efficiently.

The rest of the chapter is organized as follows: Section 3.2 describes the transmitter, receiver, and channel models. In Section 3.3, the fixed-lag approach to channel estimation is presented. It also outlines a technique for processing the symbol blocks to improve the performance of the smoother. Section 3.4 introduces a method for designing the smoother. The soft-in-soft-out equalizer and decoder are briefly described in Section 3.5. In Section 3.6, the computational complexity of the proposed method is evaluated and compared. Simulation examples and results are given in Section 3.7. The EXIT analysis is presented in Section 3.8.

3.2 System Model

3.2.1 Transmitter

The transmitter follows the model described in Section 2.3.2. Following the time-multiplexed training scheme proposed by Ma et al. [105], a sequence of l_p pilot symbols is periodically inserted per l_s data symbols to form the transmit sequence $\{s(n), n =$

$1, 2, \ldots, N\}$. Each symbol $s(n)$ represents N_{mod} bits of the coded sequence which are denoted $\mathbf{c}_d(n), d = 1, \cdots, N_{\text{mod}}$. Each pilot sequence is comprised of an impulse of magnitude $\sqrt{l_p}$ guarded by $(l_p - 1)/2$ zeros on each side. The selection of the frequency of pilots is a trade-off between spectral efficiency and estimation accuracy. The sampling period of the signal T_s is identical to the symbol period. The symbol sequence $s(n)$ is assumed to be zero mean and have unit mean power.

3.2.2 Channel Model

The channel is a multipath radio channel with each path being subject to Rayleigh fading and additive white Gaussian noise. The channel is assumed to be wide-sense stationary uncorrelated scattering (WSSUS), following the Jakes' model with the maximum Doppler frequency f_d described in (3.31). The channel's input-output equation is given by (2.2). A CE-BEM with $Q = 1$ basis function is used for 4-QAM and 16-QAM. The long-time autocorrelation properties of the radio channel are imperfectly modeled with AR(1) models of the channel coefficients, leading to elevated estimation errors. This additional error is particularly problematic for detection of higher order modulation, which requires a high SNR to achieve low BERs. Therefore, more accurate channel models are required to avoid an unacceptably high BER floor. To provide this accuracy, a higher order CE-BEM channel model is proposed for the 64-QAM receiver. A first order AR model will then be used to track the coefficients of the CE-BEM, as in the method of H. Kim and Tugnait [92].

3.2.3 Receiver

The iterative (turbo) receiver (see Fig 3.1) is comprised of the channel estimator, equalizer and decoder modules, each exchanging soft information with each other as described in Section 2.3.2. Soft information on a data bit measures the level of confidence in the decision for the bit being 0 or 1. Channel estimation is assisted with pilots. Pilots contain no useful information for the detector/decoder and are removed from the equalizer output before it is forwarded to the decoder. The soft information communicated with the decoder is in the form of log-likelihood ratio (LLR) on data bits given by $L(c(k)) := \log\{P(c(k) = 1)/P(c(k) = 0)\}$. A large absolute value for LLR implies a high degree of certainty in the knowledge of $c(k)$. A zero LLR corresponds to no knowledge of $c(k)$. The received signal and the soft decisions on data symbols from the previous iteration along with the reinserted pilots

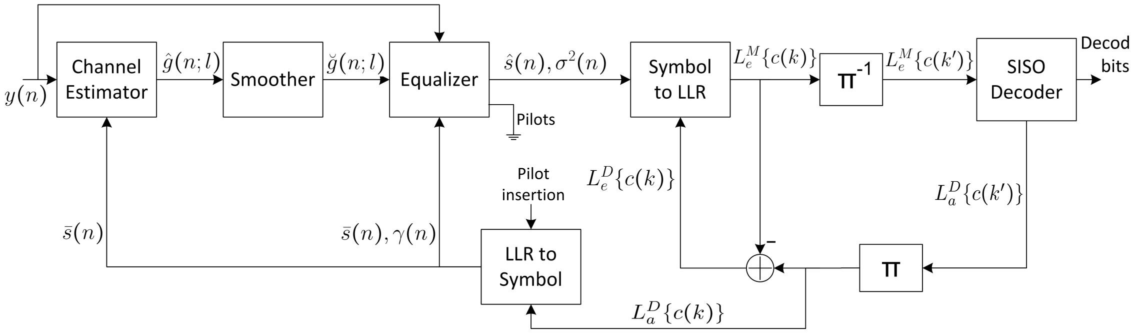

Figure 3.1: Receiver structure

are input to the channel estimator followed by a smoother. Smoothed channel gain estimates are fed to the equalizer module.

Data symbols are first estimated with the equalizer. The output of the equalizer including the data symbol estimates $\hat{s}(n)$ and their estimated variance $\sigma^2(n)$ is put through a soft-in-soft-out demapper to generate the extrinsic LLRs for the coded bits $L_e^M\{c(k)\}$. This extrinsic information consists of only the additional information on the transmitted data sequence generated by the decoder excluding the input information from the detector. The extrinsic information on a given bit is obtained by subtracting the input LLR from the output LLR. For stable iterative loop operation, the input to a given block at each iteration must be nearly independent of its output of the previous iteration, so as to prevent unwanted positive feedback where a component (channel estimator, detector, or decoder) is directly fed its own output. In every iteration, the extrinsic information from the detector/channel estimation block are fed to the decoder, whereas the information from the decoder generated in the previous iteration is fed to the channel estimation/detector block. In the decoder, $L_e^M\{c(k)\}$ is deinterleaved to provide the soft input to the soft-in-soft-out convolutional decoder. The soft-in-soft-out decoder produces LLR information on coded bits, denoted with $L_a^D\{c(k)\}$. This LLR information is then used to generate updated symbol estimates $\bar{s}(n)$ and their variance $\gamma(n)$, which are used by the channel estimator and the equalizer. At the same time, the extrinsic information $L_e^D\{c(k)\}$ is extracted to be fed back to the soft-in-soft-out demapper to further improve the next-round decisions on data symbols.

3.3 Channel Estimation and Smoothing

A fixed-lag KF smoother is used to estimate the channel at sample time n using measurements taken up to time $n + D$ where D is the smoothing lag interval of the filter. The KF is followed by a ZPF smoother. Since the channel gains are correlated, the future samples of channel output can be used by a fixed-lag KF to improve the estimation accuracy of the current sample [144]. While the ZPF can remove the portion of the estimation error process only in the band $|f| > f_D$, the fixed-lag KF is able to reduce the estimation error in the frequency band $|f| < f_D$.

In the initial iteration of the receiver algorithm, the channel is estimated using only the pilot symbols. In the second and following iterations, the pilot symbols are augmented with the data symbols detected at successively higher confidence levels with each iteration to improve the channel estimates. The smoother takes the estimated gains $\hat{g}(n;l)$ from the KF and generates the *smoothed* gain estimates $\breve{g}(n;l)$. The long memory of the smoother turns out to be beneficial in reducing the estimation error, while keeping the processing cost low.

The smoother is a high-selectivity low-pass filter. This filter must have a linear phase response to avoid phase distortion [10]. Phase distortion occurs when different frequencies in the signal are not delayed equally. As a result the signal is spread over time and distorted. Phase distortion is undesirable in communication systems. While finite-impulse-response (FIR) filters with linear phase response can be designed to have a high selectivity, these FIR filters require a large number of taps. On the other hand, infinite-impulse-response (IIR) filters can offer high selectivity with much less memory and computational complexity, but they require phase equalization to provide the linear phase response needed for detection of digital signals. If block processing is possible, the use of ZPFs can be considered. A ZPF has high selectivity at low computational cost with ideal phase response [10]. A ZPF with a component IIR filter using an elliptical approximation is employed. Given an IIR filter, a ZPF is obtained by passing a block of data through the filter in the forward and backward (reverse) directions. A ZPF has a zero phase response. The ZPF is applied to the channel gain samples for each propagation path separately.

The ZPF introduces unwanted transients at the beginning and end of the filtered output signal. Adding extra samples to the current block from the preceding and the succeeding sample blocks before applying the ZPF mitigates these transient effects. After the ZPF is applied, the unwanted transient effects will only affect the extra

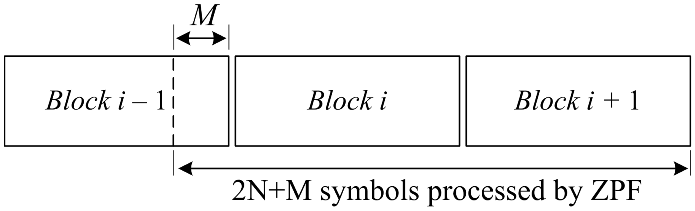

Figure 3.2: Block processing.

appended samples. The extra samples are then discarded before the smoothed channel gains are sent to the symbol detection system.

The received signal is partitioned into blocks of N samples. Each block contains symbols for N_i bits, where N_i is the length of the interleaver. At each time step, the channel gains for two blocks of N samples are being processed by the receiver as shown in Fig. 3.2, with an additional M samples taken from a previously processed block, i.e., $2N + M$ symbols are processed. Each block is processed in two stages. At time step i, the channel gains and data symbols for blocks i and $i + 1$ are estimated by the receiver using the iterative algorithm. However, only the symbol decisions for block i are forwarded to the receiver output as the results for block $i + 1$ are likely to be contaminated by the transient effects of the ZPF. At the end of each step, the final M samples of the estimated channel gains for each propagation path of block i are stored to provide the initial samples for the ZPF in the next time step.

Processing block i twice, in time steps $i - 1$, as well as time step i, increases the computational cost of channel filtering by a factor of two, but greatly reduces the final channel estimation error at the start and the end of each block. With sufficiently large M, the transient effects of the ZPF are nearly completely absent, and the remaining error floor effect is moved to cases of $E_b/N_0 > 20$ dB, which allows for acceptably low error performance (BER $\ll 10^{-6}$ with 64-QAM) for most applications. Without the two stage processing, an error floor was observed of about 10^{-5} for 64-QAM, which is too high for most applications of interest.

This procedure is similar to the overlap-and-save method used for filtering of long signals [10]. The method used in this research is integrated into the iterative processing. It must consider the interaction of detection/decoding with signal filtering, which is not considered in the prior art of overlap-and-save signal processing.

The channel estimation problem is posed as the estimation of CE-BEM coefficients given by (2.8). The complex exponential functions capture fast variations of the channel, and the evolution of the CE-BEM coefficients are assumed to obey a first-order AR model given by [92],

$$\mathbf{h}(n) = \alpha\mathbf{h}(n-1) + \mathbf{w}(n), \tag{3.1}$$

where α represents the AR model coefficient and the driving noise $\mathbf{w}(n)$ is a $Q(L{+}1){\times}1$ vector and assumed to be zero-mean white Gaussian process with the autocorrelation matrix $\mathbf{Q} := E\left[\mathbf{w}(n)\mathbf{w}(n)^H\right] = \sigma_w^2\mathbf{I}_{Q(L+1)}$ and $\sigma_w^2 = \left(1-|\alpha|^2\right)/Q(L+1)$. Each channel taps is assumed to have equal mean gain power of $P_l = 1/(L+1)$. [1]

The dynamic model for the fixed-lag KF is based on the evolution of a state vector defined as

$$\mathbf{h}_a(n) := \left[\mathbf{h}_{a,0}^T(n)\ \mathbf{h}_{a,1}^T(n)\dots\mathbf{h}_{a,D}^T(n)\right]^T,$$

where $\mathbf{h}_{a,d}(n) := \mathbf{h}(n-d)$, and D is the number of lags for the filter. Using (2.2), the state and measurement equations for the KF are written as [144]

$$\begin{aligned} \mathbf{h}_a(n) &= \mathbf{F}_a(n)\mathbf{h}_a(n-1) + \mathbf{G}_a\mathbf{w}(n), &(3.2)\\ y(n) &= \mathbf{E}_a(n)\mathbf{h}_a(n) + v(n), &(3.3)\end{aligned}$$

for n =1 , 2, ..., $N_E := 2N + M_E$, where M_E symbols from the previous block are used for the KF training. In addition, we have

$$\mathbf{F}_a := \begin{bmatrix} \alpha\mathbf{I}_{Q(L+1)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{I}_{Q(L+1)} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{Q(L+1)} & \mathbf{0} \end{bmatrix}, \tag{3.4}$$

$\mathbf{G}_a := \left[\mathbf{I}_{Q(L+1)}\ \ \mathbf{0}_{Q(L+1)\times DQ(L+1)}\right]^T$, $\mathbf{E}_a(n) := \left[\bar{\mathbf{s}}^T(n)\mathbf{B}(n)\ \mathbf{0}_{DQ(L+1)\times 1}\right]$, $E\left[v(k)v^*(j)\right] = \sigma_v^2\delta_{kj}$, and $\bar{\mathbf{s}}(n) := \left[\bar{s}(n)\ \bar{s}(n-1)\dots\bar{s}(n-L)\right]^T$ as the updated symbol estimates provided by the "LLR to symbol" block.

The autocorrelation matrix, $\mathbf{Q}_a$ of the driving noise is given by

$$\mathbf{Q}_a = E\left[\mathbf{G}_a\mathbf{w}(n)\mathbf{w}(n)^H\mathbf{G}_a^H\right] = \mathbf{G}_a\mathbf{Q}\mathbf{G}_a^H. \tag{3.5}$$

[1] This algorithm can also be used with non-uniform delay-power profiles. The autocorrelation matrix is computed as $\mathbf{Q} = \mathrm{diag}([P_0 \cdots P_L]) \otimes \mathbf{Q}_0$, where $\mathbf{Q}_0 := \frac{(1-|\alpha|^2)}{Q}\mathbf{I}_Q$.

The KF is used to compute the estimates of $\mathbf{h}_a(n)$, which is denoted by $\hat{\mathbf{h}}_a(n)$, from which a sequence of fixed-lag estimates of the CE-BEM coefficients $\{\hat{\mathbf{h}}(n) := E\left[\mathbf{h}(n)|y(1),\ldots,y(n+D)\right]; n = 1,\ 2,\ \ldots,\ N_E\}$ is obtained as

$$\hat{\mathbf{h}}(n) = \begin{cases} \hat{\mathbf{h}}_{a,D}(n+D), & n = 1,2,\ldots,N_E - D \\ \hat{\mathbf{h}}_{a,N_E-n}(N_E), & n = N_E - D + 1,\ldots,N_E \end{cases} \tag{3.6}$$

where $\hat{\mathbf{h}}_{a,D}(n+D)$ denotes the last $Q(L+1)$ elements of $\hat{\mathbf{h}}_a(n)$ [cf. (3.2)], and $\hat{\mathbf{h}}_{a,N_E-n}(N_E)$ is used to compute the last $D-1$ estimates, for which *fewer* than D future samples of $y(n)$ are available. From (2.11), the channel gains are computed as $\hat{\mathbf{g}}(n) = \mathbf{B}(n)\hat{\mathbf{h}}(n)$. The measurements up to the time $n+D$ contribute to the computation of the channel gains at time n, i.e.,

$$\hat{\mathbf{g}}(n) = \begin{cases} E\left[\mathbf{g}(n)|y(1),\ldots,y(n+D)\right], & n=1,2,\ldots,N_E-D; \\ E\left[\mathbf{g}(n)|y(1),\ldots,y(N_E)\right], & n=N_E-D+1,\ldots,N_E. \end{cases} \tag{3.7}$$

To take the uncertainty in the detected symbols into account, the variance $\gamma(n)$ of the symbol estimates is exploited to modify the measurement noise. This variance is calculated by the "LLR to symbol" block and reflects the uncertainty in the detected symbols. The detected symbols are modeled as being contaminated with error as $\mathbf{s}(n) = \bar{\mathbf{s}}(n) + \mathbf{u}(n)$, where $\mathbf{u}(n) := \left[u(n)\ u(n-1)\ldots u(n-L)\right]^T$ with the variance of $\mathbf{\Gamma}(n) := \left[\gamma(n)\ \gamma(n-1)\ldots\gamma(n-L)\right]^T$. Thus, (2.2) can be written as

$$\begin{aligned} y(n) &= \left(\bar{\mathbf{s}}(n) + \mathbf{u}(n)\right)^T \mathbf{g}(n) + v(n) \\ &= \bar{\mathbf{s}}(n)^T \mathbf{g}(n) + v'(n) \end{aligned}$$

where $v'(n) = \mathbf{g}^T(n)\mathbf{u}(n) + v(n)$. Assuming that the gain and symbol error vectors are independent and $E\left[u(n-l)u^*(n-l')\right] = \gamma(n-l)\delta_{ll'}$, the variance of $v'(n)$ is given as

$$\sigma_{v'}^2 = \sigma_v^2(n) + \sum_{l=0}^{L} p_g(n;l)\gamma(n-l), \tag{3.8}$$

where $p_g(n;l) = E\left[|g(n;l)|^2\right]$. $p_g(n;l)$ is approximated by

$$p_g(n;l) \approx \left|E[g(n;l)]\right|^2 + E\left[\left|g(n;l) - E[g(n;l)]\right|^2\right] \tag{3.9}$$

where all expectations are calculated using all measurements made up to time $n-$

1. The first term of the right-hand side of (3.9) is estimated to be the element l of $|\mathbf{B}(n)\hat{\mathbf{h}}_a^{1:Q(L+1)}(n|n-1)|^2$, and the second term is entry (l,l) of $\hat{\mathbf{R}}_{gg}(n|n-1) := \mathbf{B}(n)\mathbf{P}^{1:Q(L+1)}(n|n-1)\mathbf{B}^T(n)$.

The calculations of the fixed-lag KF are described in Algorithm 1.

Algorithm 1 Channel estimation fixed-lag KF

Inputs:

- Matrix $\mathbf{E}_a(n)$ for $n = 1, \cdots, N_E$
- Signal measurements: $y(n)$ for $n = 1, \cdots, N_E$
- Symbol detection error variances: $\gamma(n)$ for $n = 1, \cdots, N_E$

Output:

- Channel gain estimates: $\hat{\mathbf{g}}(n)$ for $n = 1, \cdots, N_E$

Working variables:

- CE-BEM vector estimates: $\hat{\mathbf{h}}_a(n|n-1)$, $\hat{\mathbf{h}}_a(n-1|n-1)$
- Covariance matrices for the estimated CE-BEM vector: $\mathbf{P}(n|n-1)$, $\mathbf{P}(n-1|n-1)$
- KF gain: $\mathbf{K}(n)$

```
$\hat{\mathbf{h}}_a(0|0) \leftarrow E[\hat{\mathbf{h}}_a(0)] = \mathbf{0}_{(D+1)(L+1)Q\times 1}$
$\mathbf{P}(0|0) \leftarrow (1/(L+1)Q)\mathbf{I}_{(D+1)(L+1)Q}$
for $n = 1, 2, \ldots, N_E$ do
    $\hat{\mathbf{h}}_a(n|n-1) \leftarrow \mathbf{F}_a\hat{\mathbf{h}}_a(n-1|n-1)$
    $\mathbf{P}(n|n-1) \leftarrow \mathbf{F}_a\mathbf{P}(n-1|n-1)\mathbf{F}_a^T + \mathbf{Q}_a$
    $\hat{\mathbf{R}}_{gg}(n|n-1) \leftarrow \mathbf{B}(n)\mathbf{P}^{1:Q(L+1)}(n|n-1)\mathbf{B}^T(n)$
    $\mathbf{\Gamma}(n) \leftarrow [\gamma(n)\ \gamma(n-1) \ldots \gamma(n-L)]^T$
    $\sigma_{v'}^2(n) \leftarrow \sigma_v^2(n) + [|\mathbf{B}(n)\hat{\mathbf{h}}_a^{1:Q(L+1)}(n|n-1)|^2 + \mathrm{diag}(\hat{\mathbf{R}}_{gg}(n|n-1))]^T\mathbf{\Gamma}(n)$
    $\mathbf{K}(n) \leftarrow \mathbf{P}(n|n-1)\mathbf{E}_a^H(n)[\mathbf{E}_a(n)\mathbf{P}(n|n-1)\mathbf{E}_a^H(n) + \sigma_{v'}^2]^{-1}$
    $\hat{\mathbf{h}}_a(n|n) \leftarrow \hat{\mathbf{h}}_a(n|n-1) + \mathbf{K}(n)[y(n) - \mathbf{E}_a(n)\hat{\mathbf{h}}_a(n|n-1)]$
    $\mathbf{P}(n|n) \leftarrow [\mathbf{I}_{Q(L+1)(D+1)} - \mathbf{K}(n)\mathbf{E}_a(n)]\mathbf{P}(n|n-1)$
    if $n > D$ then
        $\hat{\mathbf{g}}(n-D) = \mathbf{B}(n-D)\hat{\mathbf{h}}_a^{DQ(L+1)+1:(D+1)Q(L+1)}(n)$
    end if
end for
for $k = 1, 2, \ldots, D$ do
    $\hat{\mathbf{g}}(N_E - D + k) =$
    $\mathbf{B}(N_E - D + k)\hat{\mathbf{h}}_a^{Q(L+1)(D-k)+1:Q(L+1)(D-k+1)}(n)$
end for
```

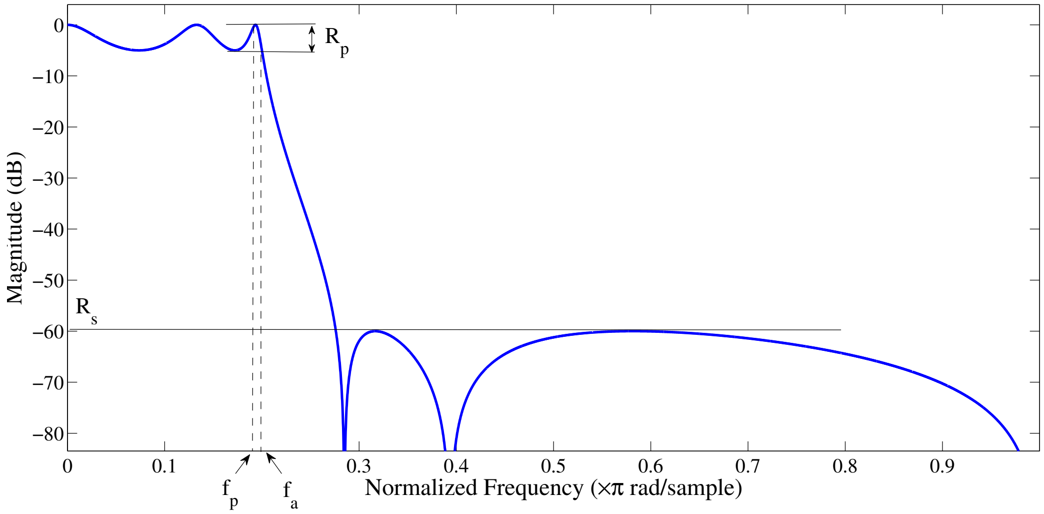

Figure 3.3: The magnitude response of an elliptic low-pass filter.

3.4 ZPF Design

The component IIR filter of the ZPF must be designed to match the characteristics of the fading process without being of unnecessary high order. The IIR filter parameters to be determined are the passband edge frequency f_p, passband ripple R_p, stopband attenuation R_s, and cut-off edge frequency f_a, as shown in Fig. 3.3. The channel gain variations are band limited to Doppler frequency f_D. Thus, one sets $f_p = f_D$. The Doppler frequency may be either estimated using methods described in [51, 77, 78] or the maximum value encountered in the application can be used. The impact of this overestimation on the performance will be demonstrated through simulation.

The other filter parameters are selected to keep the error of the ZPF output as close as possible to the minimum error calculated using the Wiener bound. In the following, a method to select these parameters is described. The input to the ZPF is the estimated channel gain $\hat{g}(n,l)$ from the KF, which is given by $\hat{g}(n;l) = g(n;l) + e(n;l)$, where $e(n;l)$ denotes the estimation error of the KF for path l. The estimation error $e(n;l)$ is assumed to be uncorrelated with $\hat{g}(n,l)$ and have constant PSD, i.e., $S_{ee}(f;l) = \sigma_v^2$, uniformly distributed over the normalized frequency range of [-1/2, 1/2]. At each iteration, the estimated channel gains from the KF are passed through the ZPF with the forward (and backward) magnitude response $A(f)$. The PSD of the output of the ZPF for propagation path l, i.e., $S_{\breve{g}\breve{g}}(f;l)$, can be written as

$$S_{\breve{g}\breve{g}}\left(f;l\right) = A^4\left(f\right)\left[S_{gg}\left(f;l\right) + S_{ee}\left(f;l\right)\right] \tag{3.10}$$

where $S_{gg}\left(f;l\right)$ is the PSD of $g\left(n;l\right)$, band limited to $\left[-f_D,\ f_D\right]$. The power of $e(n;l)$ is given by

$$P_e = P_r + P_t + P_s \tag{3.11}$$

where P_r ,P_t and P_s denote the power of the error over the passband, transition band, and stopband of the filter, respectively. The passband component P_r is given by

$$P_r = \int_{-f_p}^{f_p} S_{gg}(f;l)\left(1 - A^2(f)\right)^2 \mathrm{d}f + 2\sigma_v^2 f_p \tag{3.12}$$

where the first term of the right-hand side represents the power of the estimation error introduced by the ZPF's passband ripple, and the second term is the power of the estimation error of the KF within the passband of the ZPF. For an elliptic filter with the average magnitude of unity over the passband, if $1 - \Delta \leq A^2(f) \leq 1 + \Delta$ for $|f| < f_p$, then $\left(1 - A^2(f)\right)^2 \leq \Delta^2$. Defining $\Delta = R_p/2$, an upper bound on P_r is found to be

$$P_r^+ := \frac{R_p^2}{4}\int\limits_{-f_p}^{f_p} S_{gg}(f;l)\mathrm{d}f + 2\sigma_v^2 f_p = \frac{R_p^2}{4}P_l + 2\sigma_v^2 f_p \tag{3.13}$$

where P_l denotes the power of path l.

Similar bounds on the power of the error residing in the stopband and the transition band are $P_s^+ = 2\sigma_v^2 R_s^2(1/2 - f_a)$ and $P_t^+ = 2\sigma_v^2(f_a - f_p)$, respectively. Substituting these results in (3.11), and using $1/2 - f_a \approx 1/2$ and $2\sigma_v^2 f_p \approx W$, the upper bound on P_e is found to be

$$P_e^+ := \frac{R_p^2}{4}P_l + \sigma_v^2 R_s^2 + 2\sigma_v^2(f_a - f_p) + W. \tag{3.14}$$

Note that $P_e^+ \geq W$. Therefore, the filter is designed so that

$$P_e^+ - W = \frac{R_p^2}{4}P_l + \sigma_v^2 R_s^2 + 2\sigma_v^2(f_a - f_p) \ll W. \tag{3.15}$$

The trivial minimizer ($R_p = R_s = f_a - f_p = 0$) of the left-hand side of (3.15) is infeasible. A strategy for quickly calculating good values for the filter parameters is to select the values so that each of the values of P_r, P_t and P_s are a fraction of the Weiner bound. We select $P_r^+ - W = P_t^+ = P_s^+ = 0.1W$. Alternatively, one may fix the filter order

N_f and then try to minimize the left-hand side for the *best* filter parameters. Our simulations confirm that any selection of parameters satisfying (3.15) works almost as well for BER performance as the optimal values.

Example: For a 64-QAM receiver with $L+1=3$ equi-power propagation paths ($P_l = 1/3$), $f_D = 0.01$, and $E_b/N_0 = 13$dB, we have $\sigma_v^2 \approx 0.0167$, and $W \approx 3.3 \times 10^{-4}$. The passband edge frequency is obtained as $f_p = f_D = 0.01$. Having each component of the left-hand side of (3.15) to be $0.1W$, we get $R_p = \sqrt{4 \times 0.1 \times W/P_l} = 0.02$ or R_p(in decibels) $= -10\log_{10}(1 - R_p) \approx 0.09$dB, $R_s = \sqrt{0.1 \times W/\sigma_v^2} \approx 0.045$ or R_s(in dB) $= -10\log_{10} R_s \approx 14$dB, $f_a = f_p + 0.1 \times W/(2\sigma_v^2) = 0.011$, with $P_e^+ = 1.3 \times W$, which are realizable using an Elliptical approximation filter of order $N_f = 5$.

3.5 Soft-in-Soft-Out Equalizer and Decoder

The soft-in-soft-out equalizer was proposed in [101] to detect data symbols and produce the extrinsic information required by decoder blocks. For this purpose, a KF with a ladder-type structure from [101] is employed, as shown in Fig. 3.4. Note that at time n, the output of the ladder $\hat{s}(n)$ is independent from the input $\bar{s}(n)$, so the output $\hat{s}(n)$ is extrinsic to $\bar{s}(n)$, but does use information for other time periods. The inputs to the equalizer are the channel gain estimates $\breve{\mathbf{g}}(n)$ along with the received signal, as well as the symbol estimates and their variances from the previous iteration.

The system equations used in the equalizer KF are obtained as follows. Given the equalization delay $\delta \geq L$, we follow [101] to define a vector of δ samples as

$$\mathbf{x}_s(n) = [s(n)\ s(n-1) \dots s(n-\delta)]^T \tag{3.16}$$

where $s(n)$ is either a data symbol or a pilot symbol. Using (2.2), the system can be characterized as follows

$$\begin{aligned} \mathbf{x}_s(n+1) &= \mathbf{F}_s(n)\mathbf{x}_s(n) + \mathbf{e}_0 s(n) + \mathbf{w}_s(n) && (3.17)\\ y(n) &= \mathbf{H}_s(n)\mathbf{x}_s(n) + v(n) && (3.18) \end{aligned}$$

where $\mathbf{e}_0 := [1\ \ \mathbf{0}_{1\times\delta}]^T, \mathbf{H}_s(n) := [\mathbf{g}(n)\ \mathbf{0}_{1\times\delta-L}]$ and

$$\mathbf{F}_s := \begin{bmatrix} \mathbf{0}_{1\times\delta} & \mathbf{0}_{1\times1} \\ \mathbf{I}_\delta & \mathbf{0}_{\delta\times1} \end{bmatrix}. \tag{3.19}$$

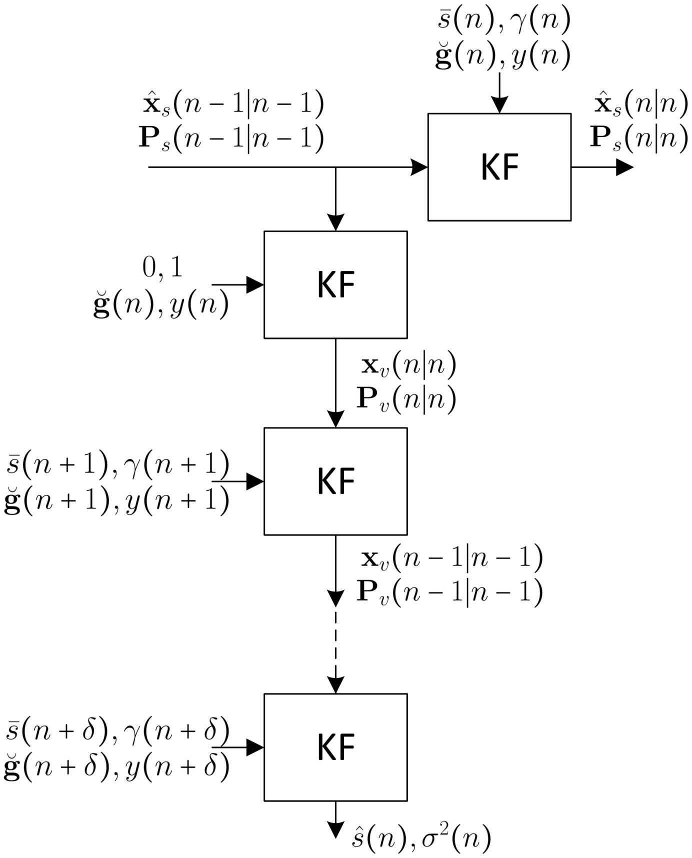
$\bar{s}(n), \gamma(n)$
$\breve{\mathbf{g}}(n), y(n)$
$\hat{\mathbf{x}}_s(n-1|n-1)$
$\mathbf{P}_s(n-1|n-1)$
KF
$\hat{\mathbf{x}}_s(n|n)$
$\mathbf{P}_s(n|n)$
$0, 1$
$\breve{\mathbf{g}}(n), y(n)$
KF
$\mathbf{x}_v(n|n)$
$\mathbf{P}_v(n|n)$
$\bar{s}(n+1), \gamma(n+1)$
$\breve{\mathbf{g}}(n+1), y(n+1)$
KF
$\mathbf{x}_v(n-1|n-1)$
$\mathbf{P}_v(n-1|n-1)$
$\bar{s}(n+\delta), \gamma(n+\delta)$
$\breve{\mathbf{g}}(n+\delta), y(n+\delta)$
KF
$\hat{s}(n), \sigma^2(n)$

Figure 3.4: Soft-in-soft-out equalizer.

In (3.18), $\mathbf{g}(n)$ is replaced with the *smoothed* estimated gains $\breve{\mathbf{g}}(n)$ and $s(n)$ with the *a priori* data symbol estimates $\bar{s}(n)$ generated by the "LLR to symbol" module from the previous iteration. At each iteration, a more accurate estimate of the data symbols is obtained. The data symbol estimates, and the pilots are fed to the channel estimator in the next iteration to refine the channel estimation. The updated channel estimate is used anew by the equalizer and the decoder to improve the estimation accuracy of the data symbols in the subsequent iteration. The noise process $\mathbf{w}_s(n)$ represents the *error* in $\bar{s}(n)$, has a variance of $\gamma(n)$, and is uncorrelated with the measurement noise $v(n)$. So, the variance of $\bar{s}(n)$ is also $\gamma(n)$. For pilot symbols, $\gamma(n) = 0$.

The soft-in-soft-out equalizer embodies two branches, as shown in Fig. 3.4. The vertical branch is initialized by the horizontal branch to generate extrinsic information for the soft-in-soft-out decoder by fixed-lag Kalman filtering. The calculations of the equalizer KF are shown in Algorithm 2.

Algorithm 2 Equalizer KF

Inputs:

- Smoothed channel gains $\breve{\mathbf{g}}(n)$ for $n = 1, \cdots, 2N$
- Signal measurements: $y(n)$ for $n = 1, \cdots, 2N$
- Symbol estimates based on the decoded bits: $\bar{\mathbf{s}}(n)$ for $n = 1, \cdots, 2N$
- Symbol detection error variances: $\gamma(n)$ for $n = 1, \cdots, 2N$

Output:

- Symbol estimates and their variances: $\hat{s}(n)$, $\sigma^2(n)$ for $n = 1, \cdots, 2N$

Working variables:

- Covariance matrices for the estimated symbol vector: $\mathbf{P}_s(n|n-1)$, $\mathbf{P}_s(n-1|n-1)$
- Covariance matrices for the estimated symbol vector on the ladder: $\mathbf{P}_v(n|n-1)$, $\mathbf{P}_v(n-1|n-1)$
- KF gain matrices: $\mathbf{K}(n)$, $\mathbf{K}_v(n)$
- Channel gain matrix: $\hat{\mathbf{H}}_s(n)$

```
$\hat{\mathbf{x}}_s(0|0) \leftarrow \mathbf{0}$
$\mathbf{P}_s(0|0) \leftarrow \mathbf{I}_{\delta+1}$
for $n = 1, 2, \ldots, 2N$ do
    $\hat{\mathbf{H}}_s(n) \leftarrow [\breve{\mathbf{g}}(n) \;\; \mathbf{0}_{1\times\delta-L}]$
    if $s(n)$ is a data symbol, then
        $\mathbf{x}_v(n-1|n-1) \leftarrow \hat{\mathbf{x}}_s(n-1|n-1)$
        $\mathbf{P}_v(n-1|n-1) \leftarrow \mathbf{P}_s(n-1|n-1)$
        $\mathbf{x}_v(n|n-1) \leftarrow \mathbf{F}_s\mathbf{x}_v(n-1|n-1)$
        $\mathbf{P}_v(n|n-1) \leftarrow \mathbf{F}_s\mathbf{P}_v(n-1|n-1)\mathbf{F}_s^T + \mathbf{e}_0\mathbf{e}_0^T$
        $\mathbf{K}_v(n) \leftarrow \mathbf{P}_v(n|n-1)\hat{\mathbf{H}}_s^H(n)[\hat{\mathbf{H}}_s(n)\mathbf{P}_v(n|n-1)\hat{\mathbf{H}}_s^H(n) + \sigma_v^2]^{-1}$
        $\mathbf{x}_v(n|n) \leftarrow \mathbf{x}_v(n|n-1) + \mathbf{K}_v(n)[y(n) - \hat{\mathbf{H}}_s\mathbf{x}_v(n|n-1)]$
        $\mathbf{P}_v(n|n) \leftarrow \left[\mathbf{I}_{\delta+1} - \mathbf{K}_v(n)\hat{\mathbf{H}}_s(n)\right]\mathbf{P}_v(n|n-1)$
        for $m = n+1 : \min(n+\delta, 2N)$ do
            $\hat{\mathbf{H}}_v(m) \leftarrow [\breve{\mathbf{g}}(m) \;\; \mathbf{0}_{1\times\delta-L}]$
            $\mathbf{x}_v(m|m-1) \leftarrow \mathbf{F}_s\mathbf{x}_v(m-1|m-1) + \bar{s}(m)\mathbf{e}_0$
            $\mathbf{P}_v(m|m-1) \leftarrow \mathbf{F}_s\mathbf{P}_v(m-1|m-1)\mathbf{F}_s^T + \gamma(m)\mathbf{e}_0\mathbf{e}_0^T$
            $\mathbf{K}_v(m) \leftarrow \mathbf{P}_v(m|m-1)\hat{\mathbf{H}}_v^H(m)[\hat{\mathbf{H}}_v(m)\mathbf{P}_v(m|m-1)\hat{\mathbf{H}}_v^H(m) + \sigma_v^2]^{-1}$
            $\mathbf{x}_v(m|m) \leftarrow \mathbf{x}_v(m|m-1) + \mathbf{K}_v\left[y(m) - \hat{\mathbf{H}}_v\mathbf{x}_v(m|m-1)\right]$
            $\mathbf{P}_v(m|m) \leftarrow [\mathbf{I}_{\delta+1} - \mathbf{K}_v(m)\hat{\mathbf{H}}_v(m)]\mathbf{P}_v(m|m-1)$
        end for
```

Continued on the next page

Algorithm 2 Equalizer KF (Continued)

```
        $\hat{s}(n) = (\delta+1)$th component of $\mathbf{x}_v(n+\delta|n+\delta)$,
        $\sigma^2(n) = (\delta+1,\delta+1)$th entry of $\mathbf{P}_v(n+\delta,n+\delta)$.
    end if
    $\hat{\mathbf{x}}_s(n|n-1) \leftarrow \mathbf{F}_s\hat{\mathbf{x}}_s(n-1|n-1) + \bar{s}(n)\mathbf{e}_0$
    $\mathbf{P}_s(n|n-1) \leftarrow \mathbf{F}_s\mathbf{P}_s(n-1|n-1)\mathbf{F}_s^T + \gamma(n)\mathbf{e}_0\mathbf{e}_0^T$
    $\mathbf{K}(n) \leftarrow \mathbf{P}_s(n|n-1)\hat{\mathbf{H}}_s^H(n)[\hat{\mathbf{H}}_s(n)\mathbf{P}_s(n|n-1)\hat{\mathbf{H}}_s^H(n) + \sigma_v^2]^{-1}$
    $\hat{\mathbf{x}}_s(n|n) \leftarrow \hat{\mathbf{x}}_s(n|n-1) + \mathbf{K}(n)\left[y(n) - \hat{\mathbf{H}}_s(n)\hat{\mathbf{x}}_s(n|n-1)\right]$
    $\mathbf{P}_s(n|n) \leftarrow \left[\mathbf{I}_{\delta+1} - \mathbf{K}(n)\hat{\mathbf{H}}_s(n)\right]\mathbf{P}_s(n|n-1)$
end for
```

3.5.1 Generating Equalizer Extrinsic Information on Data Bits

The equalizer outputs are mapped into extrinsic information on data bits as follows. The "Symbol to LLR" block takes the symbol estimates $\hat{s}(n)$ and their variances $\sigma^2(n)$ as well as the fed back decoder extrinsic information and generates the equalizer extrinsic information defined as

$$L_e^M(\mathbf{c}_d(n)) := \log\frac{P(\mathbf{c}_d(n) = 1\,|\mathbf{y})}{P(\mathbf{c}_d(n) = 0\,|\mathbf{y})} - \log\frac{P(\mathbf{c}_d(n) = 1)}{P(\mathbf{c}_d(n) = 0)} \tag{3.20}$$

for $d = 1, \cdots, N_{\text{mod}}$, where the second term represents *a priori* LLR and $\mathbf{y} := [y(1), \cdots, y(N)]$ [92]. Approximating $P(\mathbf{c}_d(n) = b\,|\mathbf{y}) \approx P(\mathbf{c}_d(n) = b\,|\hat{s}(n)), b \in \{0,1\}$ and using $P(\mathbf{c}_d(n) = b\,|\hat{s}(n)) = P(\hat{s}(n)\,|\mathbf{c}_d(n) = b)/P(\hat{s}(n))$ one has

$$L_e^M(\mathbf{c}_d(n)) := \log\frac{P(\hat{s}(n)\,|\mathbf{c}_d(n) = 1)}{P(\hat{s}(n)|\mathbf{c}_d(n) = 0)} \tag{3.21}$$

Assuming that the symbol estimation error has a Gaussian distribution with variance $\sigma^2(n)$, it can be shown that [92]

$$L_e^M(\mathbf{c}_d(n)) := \log\frac{\sum\limits_{x\in\chi(d,1)} \exp(-|\hat{s}(n) - x|^2/\sigma^2(n)) \prod\limits_{\substack{j=1\\ j\neq d}}^{N_{\text{mod}}} P(\mathbf{c}_j(n) = \mathbf{c}_j^x(n))}{\sum\limits_{x\in\chi(d,0)} \exp(-|\hat{s}(n) - x|^2/\sigma^2(n)) \prod\limits_{\substack{j=1\\ j\neq d}}^{N_{\text{mod}}} P(\mathbf{c}_j(n) = \mathbf{c}_j^x(n))} \tag{3.22}$$

where $\chi(d, b)$ is the set of all bit sequences in the constellation χ with dth bit equal to b and $\mathbf{c}_j^x(n)$ denotes the jth bit of the bit string mapped to symbol $x \in \chi$.

3.5.2 Generating Decoder Extrinsic Information

The equalizer extrinsic LLR's are de-interleaved to $L_e^M(\mathbf{c}_d(n'))$ and go to the soft-in-soft-out decoder, which computes the decoder extrinsic LLR as

$$L_e^D(\mathbf{c}_d(n')) := \log \frac{P(\mathbf{c}_d(n') = 1\,|L_e^M(\underline{\mathbf{c}}_d)}{P(\mathbf{c}_d(n') = 0\,|L_e^M(\underline{\mathbf{c}}_d)} - L_e^M(\mathbf{c}_d(n')) \tag{3.23}$$

where $L_e^M(\underline{\mathbf{c}}_d)$ is the equalizer extrinsic LLR on all bits of the codeword. The first term at the right hand side of the above equation is the decoder intrinsic LLR, $L_a^D(\mathbf{c}_d(n'))$. At the first iteration, there is no *a priori* information for the equalizer and thus, $L_e^D(\mathbf{c}_d(n')) = 0$.

3.5.3 LLR to Symbol Conversion

The LLR information on data bits from the decoder are converted to data symbol estimates used by the equalizer and channel estimator KF. The intrinsic LLR from the decoder is converted to the mean and variance of data symbols as follows.

$$\bar{s}(n) = \sum_{x \in \chi} x \prod_{d=1}^{N_{\text{mod}}} P(\mathbf{c}_j(n) = \mathbf{c}_j^x(n)) \tag{3.24}$$

$$\gamma(n) = \sum_{x \in \chi} |x - \bar{s}(n)|^2 \prod_{d=1}^{N_{\text{mod}}} P(\mathbf{c}_j(n) = \mathbf{c}_j^x(n)) \tag{3.25}$$

where

$$P(\mathbf{c}_j(n) = 1) = [1 + \exp(-L_a^D(\mathbf{c}_j(n)))]^{-1} \tag{3.26}$$

$$P(\mathbf{c}_j(n) = 0) = [1 + \exp(L_a^D(\mathbf{c}_j(n)))]^{-1} \tag{3.27}$$

3.6 Computational Cost

In this section, the computational cost of the proposed method is compared with that of the EKF-based method. The computational complexity of the EKF method of H.

Kim and Tugnait [92] is $\mathcal{O}((\delta_E + 2)[\delta_E + 1 + Q_E(L+1)]^2)$ complex multiplications per symbol per iteration, where δ_E and Q_E denote the memory of the EKF and the number of basis functions, respectively. The computational complexity of the proposed method is $\mathcal{O}(2(\delta+2)(\delta+1)^2 + 2[Q(L+1)(D+1)]^2 + 4(L+1)N_f)$ per symbol, per iteration, where D represents the memory of the channel estimation KF in samples. Keeping the dominant terms, the relative complexity of the proposed method compared with the EKF method is given by

$$\text{Relative complexity} \approx \frac{2\left(\delta^3 + [Q(L+1)D]^2\right)}{\delta_E\left[\delta_E + Q_E(L+1)\right]^2}. \tag{3.28}$$

Thanks to the smoother at the output of the channel estimator, the order of the KF in the proposed method can be much smaller than that of the EKF to achieve the same performance, or equivalently , $Q \ll Q_E$. The cost of using a lower order EKF is, of course, an inferior performance. First, let us consider the case of a three-tap channel $L = 2$ as described in the work of H. Kim and Tugnait [92], where $\delta_E = 5$ and $Q_E = 5$. In the simulations, we picked $N_f = 5, D = 4$ and $\delta = 5$. For the 4-QAM and 16-QAM schemes, the value of Q =1 was selected which corresponds to a simple temporal AR(1) model of the channel coefficients. Comparing with [92], the complexity of our method would be three times lower than the EKF. For the 64-QAM case, the CE-BEM-based channel estimator with $Q = 3$ slightly outperforms $Q = 1$; therefore, we select $Q = 3$. On the other hand, for the EKF method to have the same performance, one needs to choose Q_E =15 and δ_E =9, and its complexity is about ten times that of the proposed method.

As another scenario, consider a typical third-generation (3G) wideband code-division multiple-access (CDMA) channel with L =7 [155]. For the proposed method, we select $Q = 1, D = 8$ and $\delta = 8$, whereas for the EKF method, one would need to take $Q_E = 15$ and $\delta_E = 8$, to provide comparable error performance.

As mentioned earlier, the reduction in computational cost is due to the smoother, which nearly removes the out-of-band estimation error signal, the removal of which would otherwise require a high order Kalman filtering. Based on the argument made in Section 3.4, this out-of-band estimation error is proportional to $1 - 2f_D$, the width of the band. Since $1 - 2f_D \ll 1$ for most channels of interest, the out-of-band error constitutes a major portion of the overall estimation error. Therefore, the smoother is responsible for most of the error reduction in all cases of interest. Consequently, the order of magnitude reduction in the computational cost when a smoother is used, is likely to be the case for all channels of interest.

3.7 Simulation Results

The performance of the proposed method is demonstrated by evaluating the normalized MSE (NMSE) and the BER versus SNR for 4-QAM, 16-QAM, and 64-QAM constellations. The sensitivity of the system to the ZPF design is also shown. A doubly selective Rayleigh channel with three taps (L =2) is considered. For each tap, the time-variant channel impulse response $g(n;l)$ is a complex Gaussian process with zero-mean and variance P_l =1/3, independent from other taps impulse responses. The Rayleigh channel was simulated based on the method used in [175].

The sampling interval was T_s = 25 μs. The Doppler spread was f_d = 400 Hz, equivalent to a normalized Doppler spread $f_D = f_d T_s = 0.01$, which matches a fading process for a radio signal with a carrier frequency of 2 GHz, to communicate with a vehicle moving at 216 km/h. Our channel parameters are taken from the work of H.Kim and Tugnait [92] to provide a fair comparison with prior art.

Data blocks were coded by a nonsystematic convolutional code of rate 1/2 with an octal generator of (133,171). The performance of the technique with an LDPC code is also examined.

Coded bits were interleaved and Gray-mapped onto either a 4-QAM, 16-QAM, or 64-QAM constellation with unit mean power to form transmit blocks. Every l_s = 20 data symbols were multiplexed with l_p = 5 pilots including an impulse of magnitude $\sqrt{l_p}$ guarded by zeros (as in [157] and [105]), resulting in a pilot overhead of 20%. The pilot cost of 0.97 dB is not included in the E_b/N_0 values used in the x-axis of the following performance plots. The length of the transmit blocks was 10^4 symbols. Each block was prefixed with M = 2000 samples from the previous block, as described in Section 3.3. A number of M_E = 500 symbols from the previous block were used for the channel estimation KF training (see Section 3.3).

The channel estimator and equalizer were fixed-lag KFs with delays of D = 4 and δ = 5, respectively. The IIR component of the ZPF was a fifth-order elliptic filter with a normalized passband edge frequency of $f_p = f_D = 0.01$ and designed based on (3.15) for a 64-QAM modulation type at E_b/N_0 = 13 dB. The filter parameters were calculated as R_p = 0.09 dB, R_s = 14 dB and f_a = 0.011 (see Example (1) of Section 3.4). We used a Q =1 model for the 4-QAM and 16-QAM simulations. For the 64-QAM, a more accurate model, i.e., a CE-BEM with Q = 3 and T_p = 100 was used and compared to the $Q = 1$ case. The length of the channel estimation KF's state vector was 15 and 45, for the $Q = 1$ and $Q = 3$ models, respectively. The time correlation of the CE-BEM coefficients for all the three schemes was set to $\alpha = 0.98$.

The BER versus E_b/N_0 for the 4-QAM and 16-QAM schemes, based on a $Q = 1$ model for channel variations, are depicted in Figs. 3.5 and 3.6. The performance of the proposed method is compared with the EKF method of H. Kim and Tugnait [92] and the perfect CSI performance. For the EKF method, a CE-BEM with $Q = 5$ was employed with $\alpha = 0.996$. In Fig. 3.7, the case of a 64-QAM scheme is considered, where the $Q = 1$ and $Q = 3$ models' performances are compared with the perfect CSI receiver. The EKF method did not reliably converge for the 64-QAM scheme; therefore the corresponding results are not included in this figure. It can be seen that the $Q = 1$ model performs almost as well as the CE-BEM for BER $< 10^{-6}$, but starts to deteriorate afterward. With the proposed method, all BER curves are within 0.3 dB of the perfect-channel receiver. In addition, the system convergence is fast. The average number of iterations it takes for system to converge was approximately three iterations per trial for the case of the 64-QAM scheme.

The 64-QAM setup was also used to contrast the performance of ZPF with that of FIR and IIR filters in Fig 3.8, to justify the use of a ZPF. The IIR filter is designed using the same specifications as the IIR component of the ZPF, except for the passband ripple and stopband attenuation in decibels being doubled (since the magnitude response of a ZPF is equivalent to two cascaded component IIR filters). The FIR filter was designed using the least squares method, where the parameters were selected to be the same as the ZPF designed previously. It can be seen that it requires 2000 taps for an FIR filter to achieve the same performance as a ZPF with only a fifth-order component IIR filter; an obvious cost savings. Moreover, the IIR filter introduces a phase distortion that significantly degrades the performance of the receiver.

The performance of the system was also evaluated in terms of the NMSE of the channel estimator. Fig. 3.9 illustrates how the NMSE improves as the number of lags is increased under different channel models. The normalized Wiener filter bound, approximated as $2(L+1)f_D\sigma_v^2$, is also plotted (see Eq. (2.20)). We set $E_b/N_0 = 13$ dB, and used known symbols in the receiver. A significant improvement is observed by increasing the lag of the KF to four and there is a fairly small difference between the $Q = 1$ and $Q = 3$ cases. However it is obvious from this plot that no noticeable improvement should be expected by increasing the order of the CE-BEM beyond $Q =$

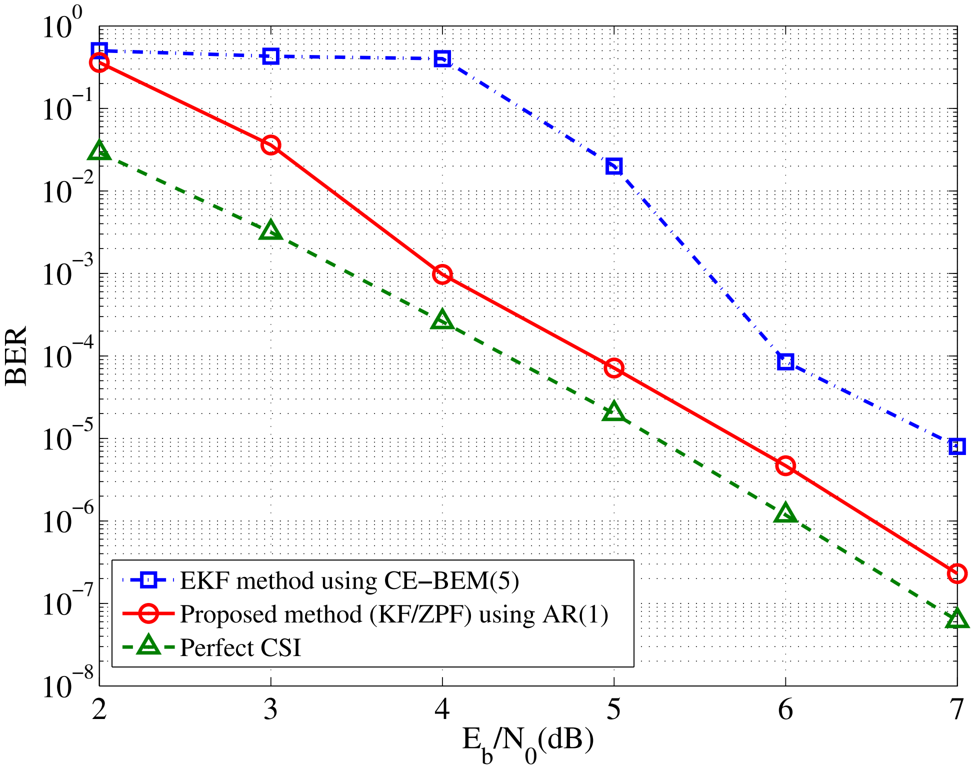

Figure 3.5: BER versus E_b/N_0 for a 4-QAM receiver.

3.

In Fig. 3.10, the NMSE versus E_b/N_0 is plotted and compared with the normalized Wiener bound. A four-lag KF filter was used ($D = 4$). A consistent and linear decrease in the NMSE with increasing E_b/N_0 is observed, and no error floor is evident.

LDPC and turbo codes have attracted huge interest in recent years. The performance of the proposed estimator when used with an LDPC code is also examined. The LDPC code is the rate- 1/2 code from DVB-S.2 standard for the satellite transmission of digital television. Fig. 3.11 shows the BER performance of a 4-QAM scheme, transmitting over an eight-tap ($L = 7$) radio channel with a power delay profile of [0 -2.4 -6.5 -9.4 -12.7 -13.3 -15.4 -25.4] dB, normalized to a total power of unity. This profile is typical of a vehicular 3G wideband CDMA system moving at 120 km/h [155]. For this simulation, we select $Q = 1$. Based on the research by Ma et al. [105], the optimal number of pilots per pilot segment is $2L$+1. The bandwidth efficiency of the optimal BER arrangement is reduced to $l_s/(l_p + l_s) \approx 57\%$ for this particular channel with $l_p = 15$ and $l_s = 20$. This efficiency is considered unacceptably low; therefore we maintained the 80% efficient pilot scheme with $l_p = 5$ and $l_s = 20$ for

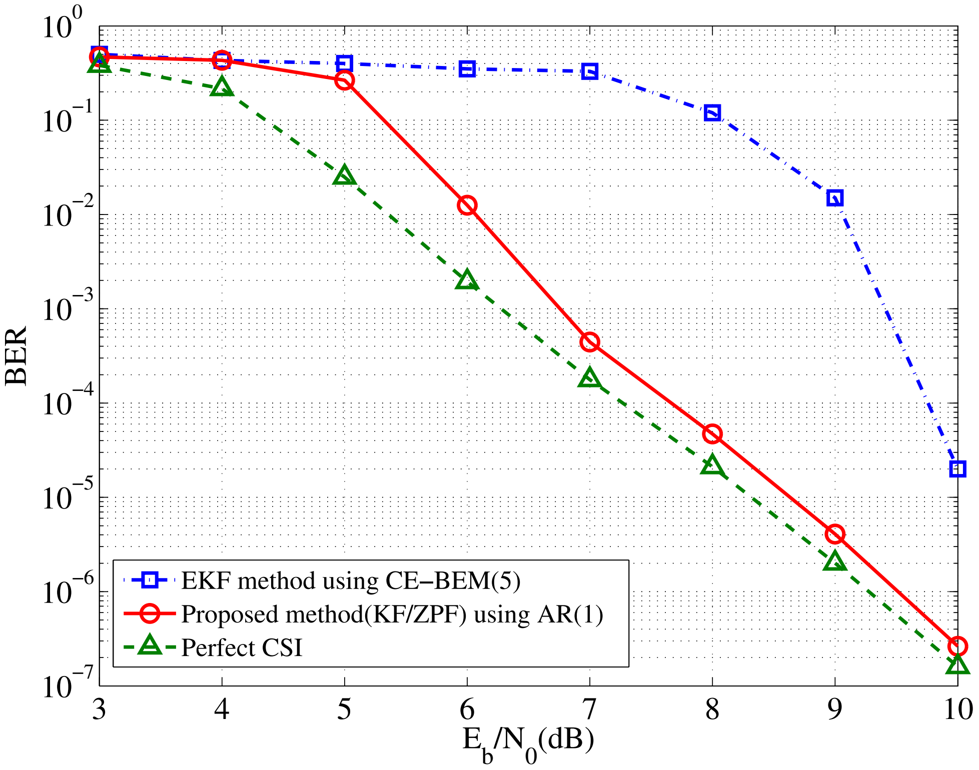

Figure 3.6: BER versus E_b/N_0 for a 16-QAM receiver.

the LDPC simulations. The first iteration only estimates the gains for the first three propagation paths, while the following iterations estimate all propagation path gains. It should be noted, however, that for our sampling period of $T_s = 25\ \mu s$, a delay of $L = 7$ samples corresponds to an extra propagation distance of over 52 km, which is unlikely to be encountered in a well designed deployed system.

It is seen that the perfect-CSI BER curve drops off at SNR = 2.2 dB. The knee point of the BER graph for our method is 5 dB, whereas that of the EKF method is 6.4 dB. Therefore, the proposed method would start to converge to low BER at a 1.4 dB less SNR, compared to the EKF. Yet, the complexity of our method is an order of magnitude less. At SNR = 5 dB, we did not observe any error in 2×10^8 data symbols, which is indicated by a downward arrow in the figure. The performance can be improved by increasing the number of pilots to $l_p = 15$ from $l_p = 5$, but the reduction of signaling efficiency is significant.

As shown, the proposed system provides excellent performance at the SNR values required for low error reception. Real wireless systems employ power control to keep the mean received SNR within a specified range. However, the sensitivity of the

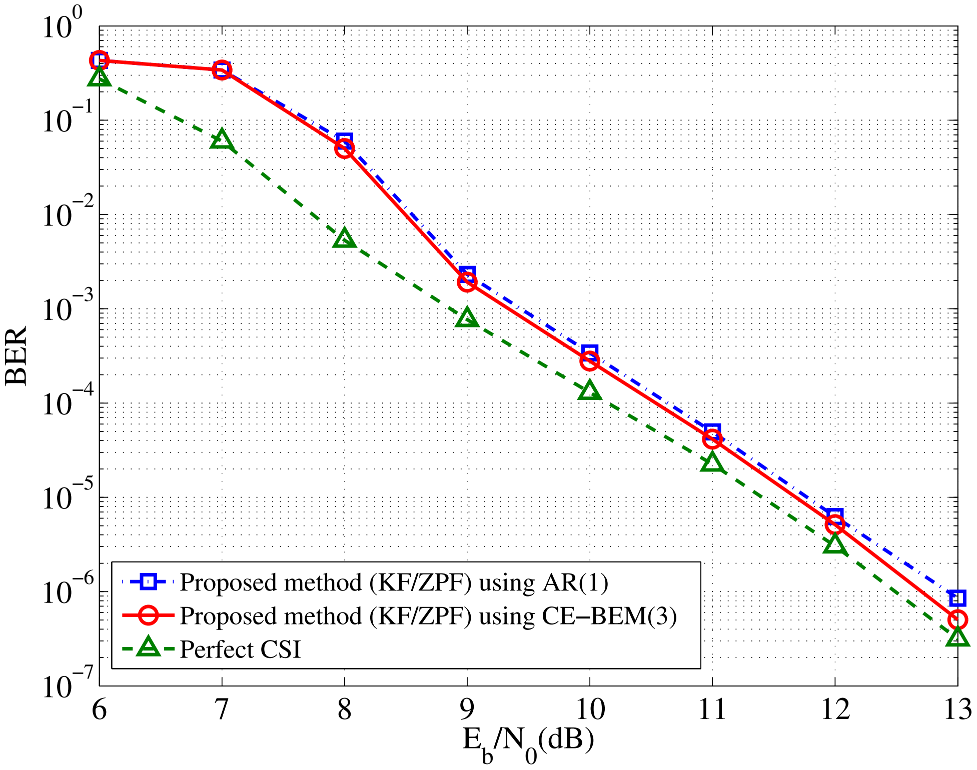

Figure 3.7: BER versus E_b/N_0 for a 64-QAM receiver.

ZPF design to the *assumed* value of the Doppler frequency within the receiver is a legitimate concern. Here, the most influential parameter is the ZPF's passband edge frequency f_p. Fig. 3.12 depicts the BER of a 64-QAM receiver versus f_p at E_b/N_0 =13 dB when the Doppler frequency is fixed at f_D =0.01. It is clear that as long as f_D is overestimated, i.e., $f_p > f_D$, the BER remains within the close vicinity of the perfect-channel case over a reasonable range of f_p. The price of using too high a value of f_p is, of course, an increased computational cost. In the next section, we will perform an EXIT analysis to infer the superior BER performance of the proposed method.

3.8 Extrinsic Information Transfer Chart Analysis

An EXIT chart is a powerful tool to analyze and predict the performance of iterative receivers by depicting the exchange of information between the decoder and equalizer in a diagram [29, 92]. It helps select the combination of error correcting code, modulation and SNR needed for convergence to low BER state for a given

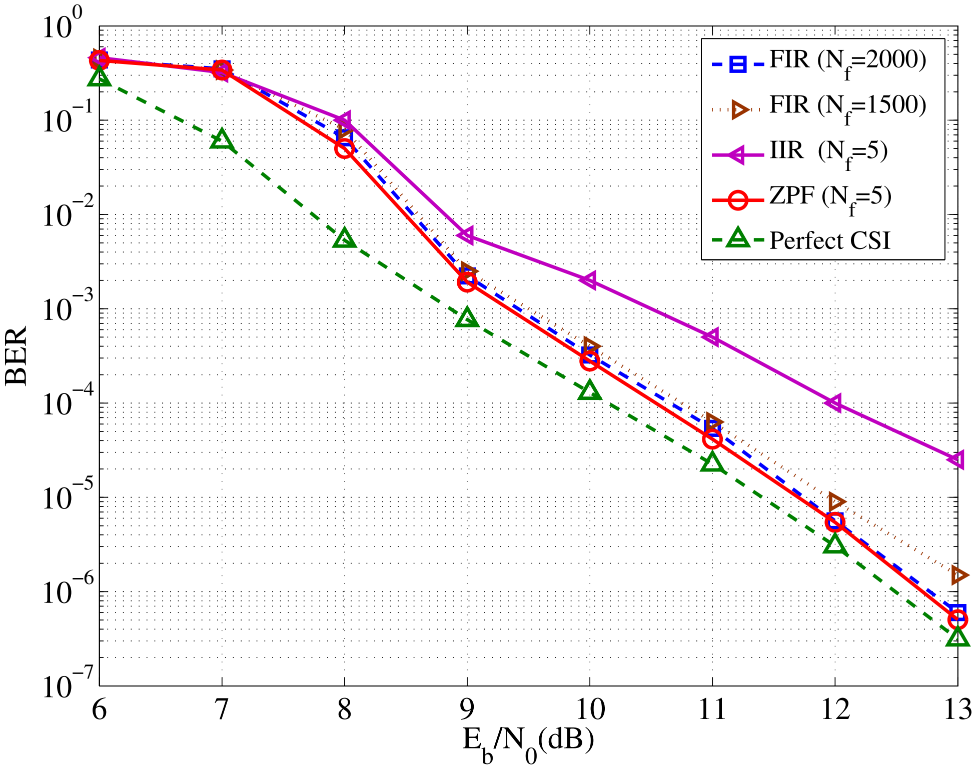

Figure 3.8: BER versus E_b/N_0 for a 64-QAM scheme, comparing ZPF, ordinary IIR and FIR filters.

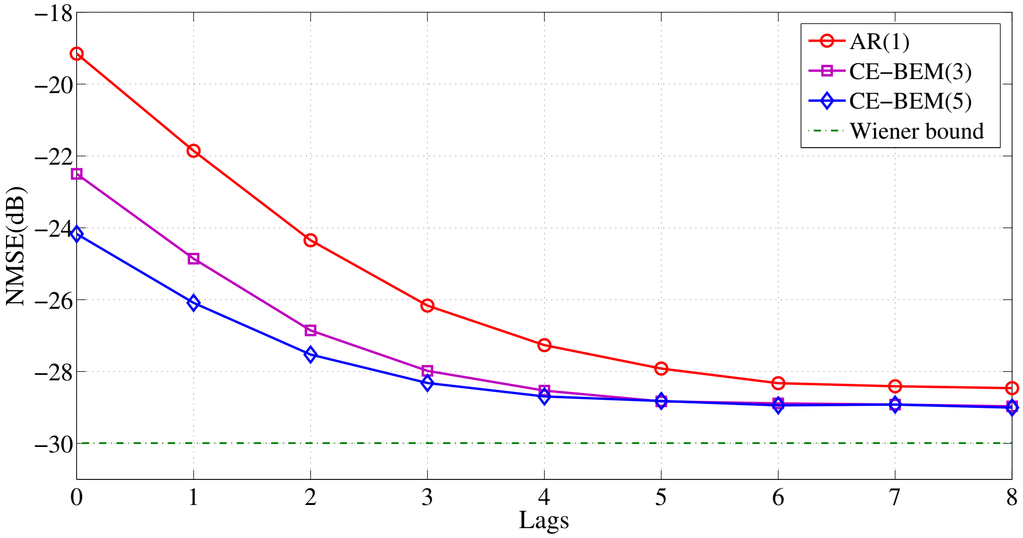

Figure 3.9: NMSE against the number of lags for a 64-QAM scheme assuming known symbols at the receiver.

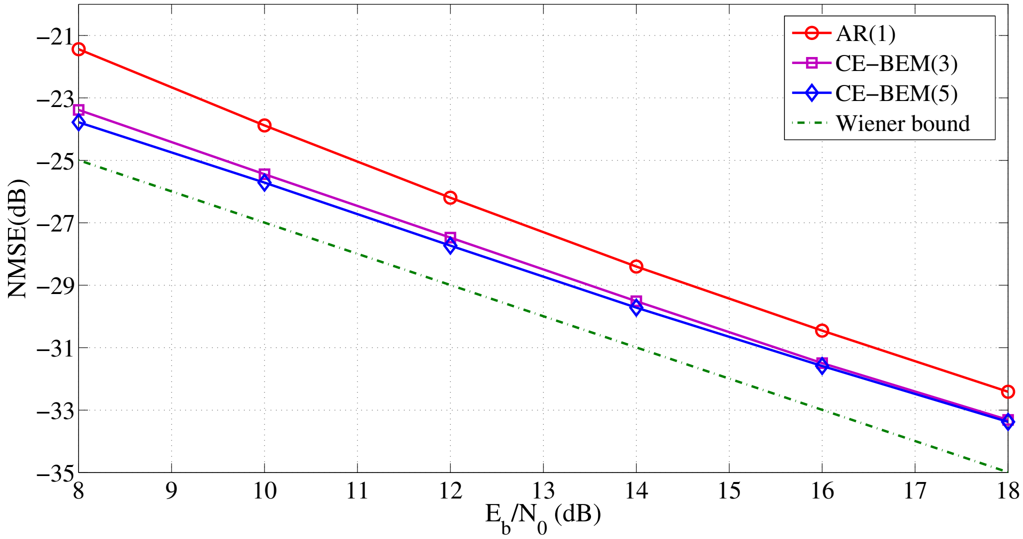

Figure 3.10: NMSE versus E_b/N_0 for 64-QAM, under different channel models using a KF with four lags.

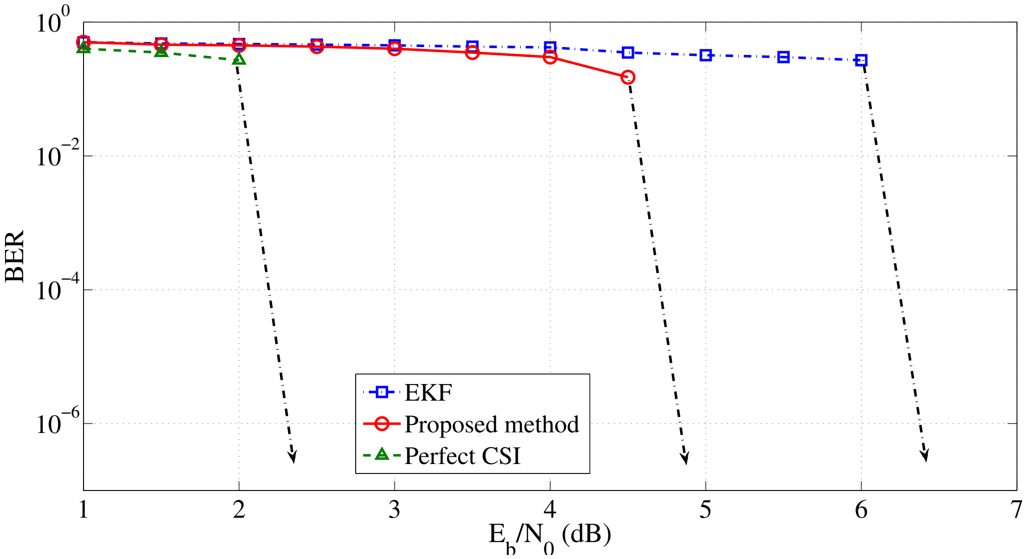

Figure 3.11: Performance of a 4-QAM turbo receiver with LDPC code over an eight-tap channel compared with the perfect-channel and the EKF method.

detector and channel such that the receiver would quickly converge to a low bit error state. EXIT charts have been employed to verify the adequacy of a proposed estimation/equalization and decoding system without having to perform expensive simulations [92, 120]. EXIT charts have been shown to provide insight into turbo processing and are widely used for analysis [6, 7, 73, 80, 88, 92, 120, 125, 142, 166] and

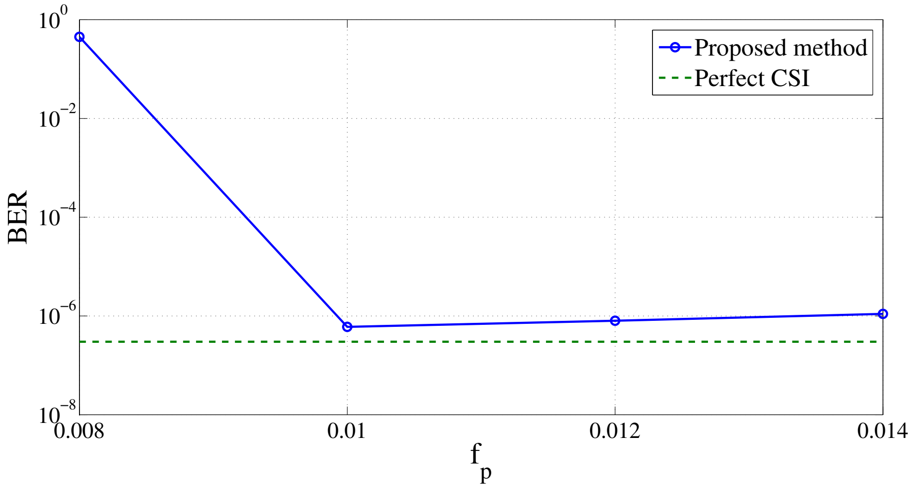

Figure 3.12: BER vs ZPF's passband edge f_p when the normalized Doppler frequency is fixed at f_D =0.01.

design [26, 99, 100, 151] of iterative receivers. A near capacity modulation scheme is designed by Ng et al. [124] based on EXIT charts. The analytic properties of EXIT charts have been exploited to design iterative processors by Ashikhmin, Kramer, and Brink [12].

Mutual information proves an accurate measure to predict the performance of iterative algorithms [29, 39, 160, 162]. Detector and decoder modules in the receiver can be viewed as soft-input-soft-output functional blocks whose performance is characterized by mapping input mutual information to the output mutual information. This concept is used in building EXIT charts, where the function of each block is represented by a curve depicting output mutual information versus input mutual information. The extrinsic mutual information transfer functions of the decoder and equalizer are plotted in a single diagram as shown in Fig. 3.13. For the equalizer, the input mutual information denoted by I_D is shown along the abscissa, whereas the output mutual information, I_E, is plotted along the ordinate. Since the input of the equalizer is the output of the decoder and vice versa, the axes are swapped when the decoder curve is drawn. The exchange of the extrinsic information between the decoder and the equalizer can be represented by a "zigzag" path, i.e., a trajectory, bouncing back and forth between the two curves. The trajectories start at the origin, with each segment representing either the decoder task (horizontal segments) or the equalizer task (vertical segments). The convergence of the iterative process to low BERs is possible if the curves intersect only at a high I_D value. The speed of the

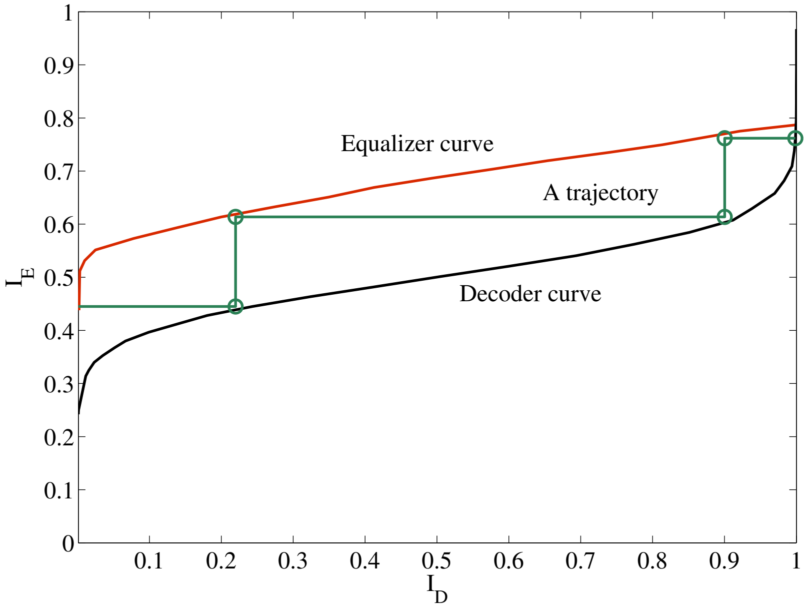

Figure 3.13: An EXIT chart.

convergence can also be inferred from the width of the tunnel; a wide tunnel indicates fast convergence.

In general, the equalizer curve gets closer to the decoder curve for higher order modulations or when the SNR or the length of the error-correcting code is reduced. As a result, the tunnel between the decoder and equalizer curves becomes narrower, and more iterations are needed for convergence. A very low BER will be possible at high-enough SNRs if the system can be designed so that the curves always cross very close to line I_D = 1. A system can achieve good BER with very few iterations if the equalizer curve is well above of the decoder curve for all values of I_D.

The simulation setup used to generate the EXIT function is shown in Fig. 3.15. The input to the soft-in-soft-out decoder $L_e^M\{c^i(k')\}$ was modeled by a Gaussian random process with mean $c^i(k')\sigma_L^2/2$ and variance σ_L^2. The mutual information values I_E and I_D were measured at the input and output of the decoder and are functions of σ_L, where σ_L was varied in the range [0.001, 200]. The interleaver size in this setup was 10^5 bits. We used the MATLAB code developed by Maunder available online [110]. The EXIT curves and average trajectories for the 4-QAM and 16-QAM under different E_b/N_0s are given in Figs. 3.16 and 3.17. Fig. 3.18 shows the case of a 64-QAM scheme with E_b/N_0 = 13 dB. The number of channel taps was L+1 = 3 with a uniform power profile, and the aformentioned convolutional code was employed.

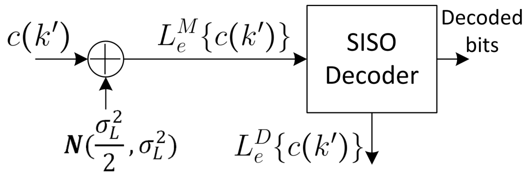

Figure 3.14: Block diagram used to generate the EXIT charts: Decoder setup

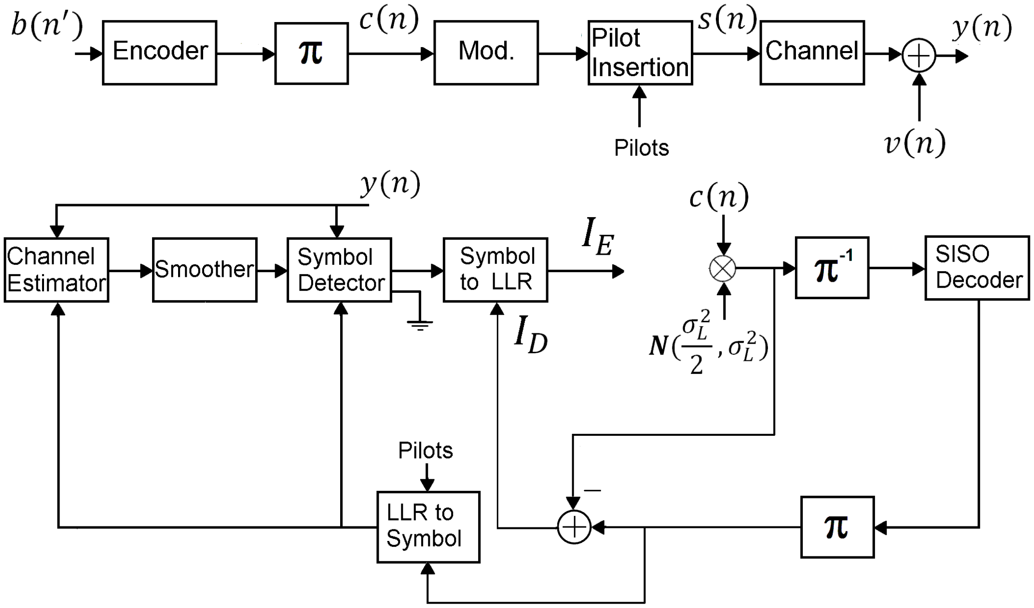

Figure 3.15: Block diagram used to generate the EXIT charts: Equalizer setup

In this figure the equalizer curves for the most significant bit (MSB) and the least significant bit of the symbols are also plotted. The major contribution of the MSB in the convergence is obvious. In Figs. 3.16–3.18, the EXIT curves of the channel estimation/detection subsystem are well above the EXIT curve of the decoder and convergence to a low BER state is achieved after only few iterations.

The convergence performance of a 4-QAM modulation scheme over an $L{+}1 = 8$ tap channel using the previously described LDPC error-correcting code is compared with the EKF method in Fig 3.19. The LDPC code's curve has a higher value I_E for lower values of I_D, creating a higher "knee" point, where the code curve transitions from a steep slope to shallow slope, as compared with the convolutional code. The channel estimation/detector EXIT curve must be above this "knee" for convergence. It can be seen in Fig 3.19, that the proposed channel estimator/detector has a higher EXIT curve, providing convergence at lower SNR values than the EKF. These charts

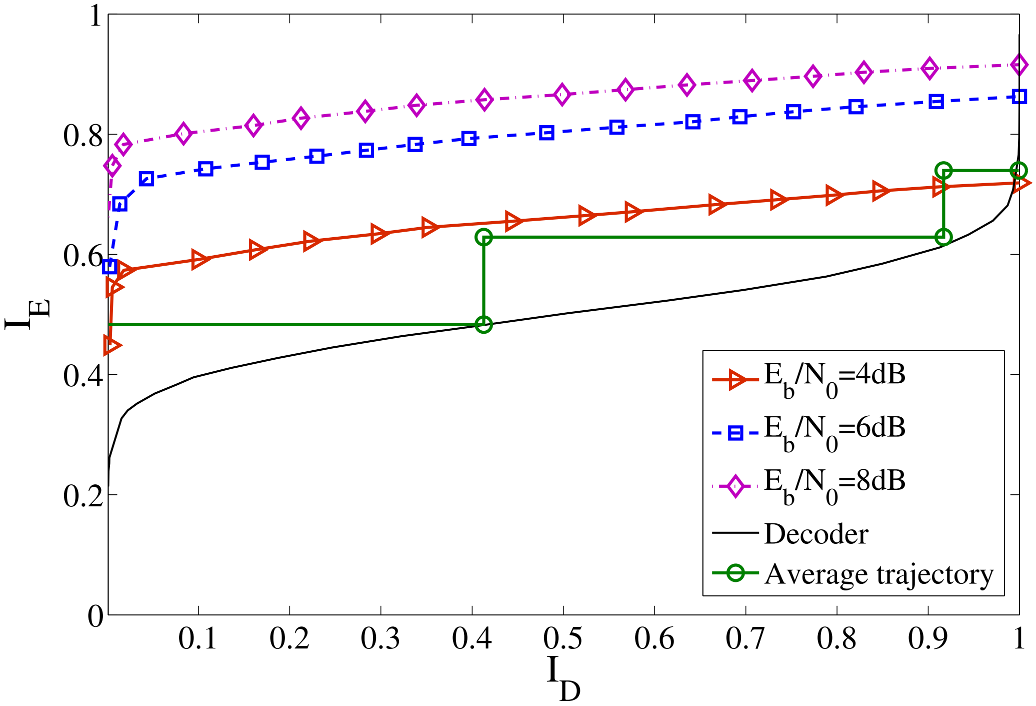

Figure 3.16: EXIT charts for a 4-QAM receiver under different E_b/N_0 including the average trajectory for E_b/N_0 =4dB.

are predicting that in comparison with the EKF method, the intersection of the equalizer and decoder curves of the proposed method would occur at a higher value of I_D and I_E, for 1.4 dB less SNR. Thus, the proposed method is expected to have a lower BER, at lower SNR.

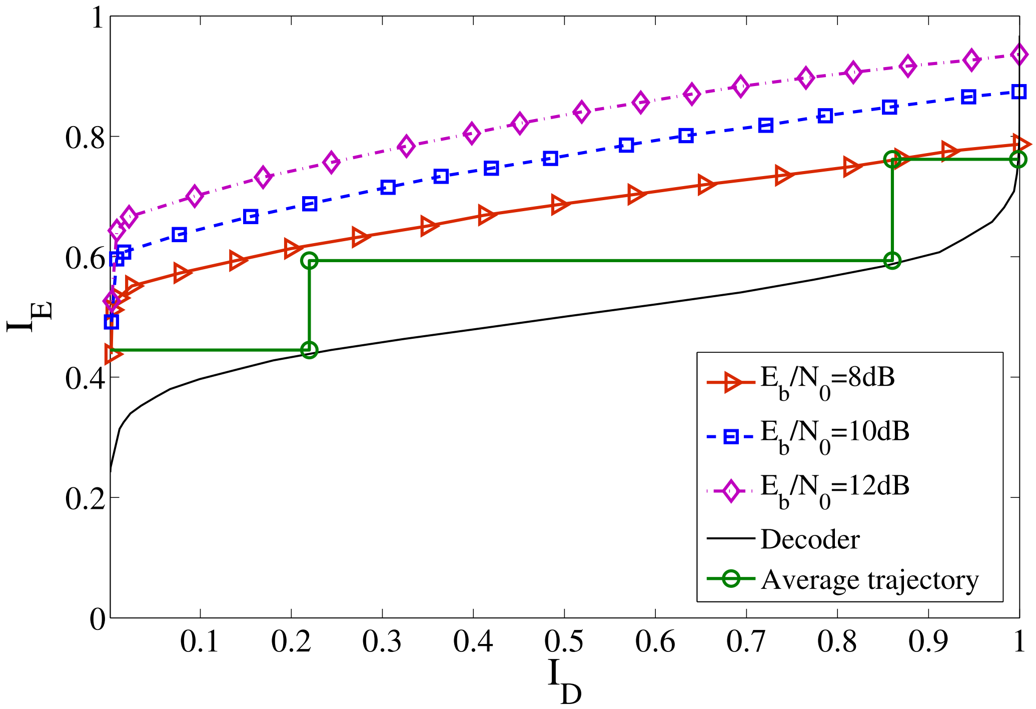

Figure 3.17: EXIT charts for a 16-QAM receiver under different E_b/N_0 including the average trajectory for E_b/N_0 =8 dB.

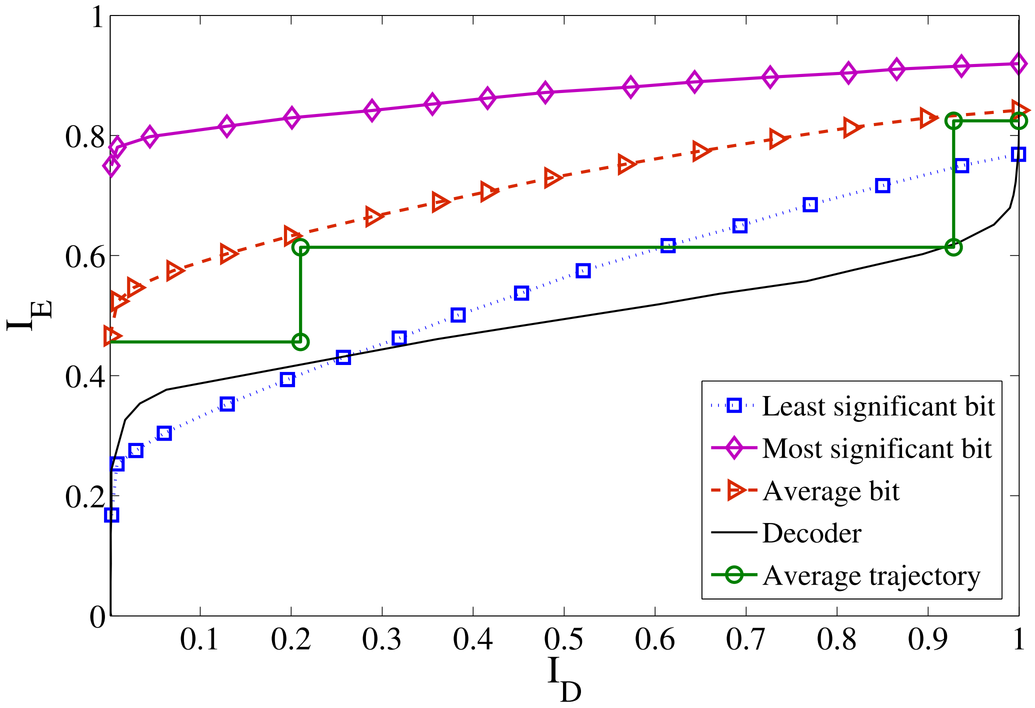

Figure 3.18: EXIT charts for a 64-QAM receiver using CE-BEM with Q =3 and E_b/N_0 =13 dB.

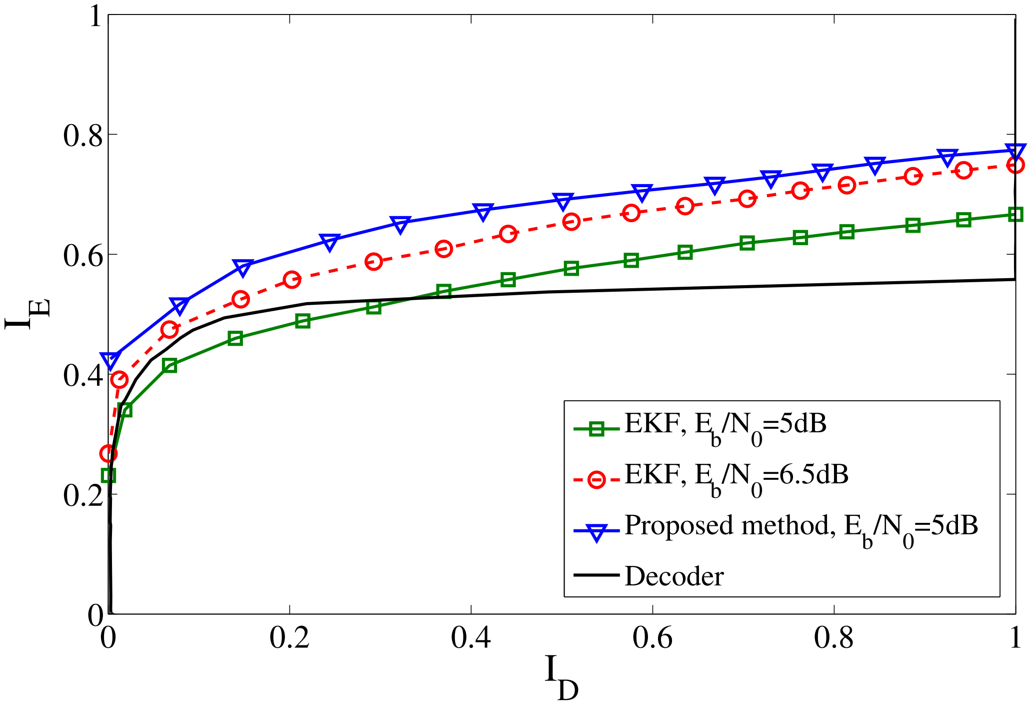

Figure 3.19: EXIT chart for a 4-QAM receiver using LDPC code.

3.9 Estimation of Fast Fading Radio Channels Based on AR(4) for 256-QAM

Using sufficiently large state vectors, the prior methods may be applied to higher order modulation schemes, but the computational cost becomes prohibitive, particularly when the radio channel entails a large number of taps. In this section, instead of a high order CE-BEM we use a lower order AR model to characterize the channel. It is demonstrated that, compared to the previous method, this technique provides a superior bit error rate performance with a low computational cost for fast-fading channels. In contrast to other techniques for fast-fading channels, the error floor for this technique is at a much higher SNR level. The improvement in the BER is significant for a 256-QAM scheme, where the BER does not show any error floor for an E_b/N_0 of as high as 17dB.

3.9.1 Channel Estimation

The receiver structure shown in Fig. 3.1 is considered. The channel is described by (2.2). Since a CE-BEM is not matched to the long-term statistics of the channel, a richer channel model than (2.12) is needed. The evolution of each channel path over time is characterized by the AR(p) process with

$$g(n;l) = -\sum_{j=1}^{p} a_j g(n-j;l) + w(n;l), \tag{3.29}$$

for $l = 0, \dots, L$, where $w(n;l)$ represents the noise process of path l. Define the vector of the AR model for path l as $\mathbf{a} := [a_1 \cdots a_p]^T$. By solving the Yule-Walker equations [30], one obtains

$$\mathbf{a} = -\mathbf{R}_l^{-1}\mathbf{r}_l \tag{3.30}$$

where $\mathbf{R}_l$ represents the $p \times p$ correlation matrix for tap l, defined as

$$(\mathbf{R}_l)_{i,j} = P_l \mathcal{J}_0(2\pi f_D |i-j|), \tag{3.31}$$

with $\mathcal{J}_0(\cdot)$ denoting the first-kind, zero$^{\text{th}}$ order Bessel function. Also, $(\mathbf{r}_l)_i := P_l \mathcal{J}_0(2\pi f_D i)$ for $i = 2, ..., p+1$. The variance of the channel noise process for path l, $\sigma_w^2(l)$, is calculated as

$$\sigma_w^2(l) = P_l - \mathbf{a}^H \mathbf{R}_l \mathbf{a} \tag{3.32}$$

for $l = 0, 1, ..., L$. The system state is characterized by an $(L+1)p \times 1$ vector, defined as

$$\mathbf{h}_p(n) := \left[g(n;0) \cdots g(n-p+1;0) \cdots g(n;L) \cdots g(n-p+1;L)\right]^T . \tag{3.33}$$

The system equations are given by

$$\begin{aligned} \mathbf{h}_p(n) &= \mathbf{F}\mathbf{h}_p(n-1) + \mathbf{w}(n), & (3.34)\\ y(n) &= \mathbf{E}(n)\mathbf{h}_p(n) + v(n), & (3.35) \end{aligned}$$

where $\mathbf{F} = \mathbf{I}_{L+1} \otimes \mathcal{K}$ with

$$\mathcal{K} := \begin{bmatrix} -a_1 & -a_2 & -a_2 & \ldots & -a_p \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}, \tag{3.36}$$

$\mathbf{E}(n) := [s(n)\ s(n-1) \cdots s(n-L)] \otimes \mathbf{e}_0^T$, $\mathbf{e}_0 = [1\ \mathbf{0}_{1\times(p-1)}]^T$, $\mathbf{w}(n) = [w(n;0)\ \mathbf{0}_{1\times p-1} \ldots$, $w(n;L)\ \mathbf{0}_{1\times p-1}]^T$. The autocorrelation matrix for $\mathbf{w}(n)$ is $\mathbf{Q}_w = \text{diag}([\sigma_w^2(0)\ \mathbf{0}_{1\times p-1} \ldots$ $\sigma_w^2(L)\ \mathbf{0}_{1\times p-1}])$.

A KF is used to obtain gain estimates, $\hat{\mathbf{g}}(n) := [\hat{g}(n;0) \cdots \hat{g}(n;L)]^T$ as presented in Algorithm 3, where

$$\hat{g}(n;l) = \mathrm{E}[g(n;l)|y(1), \cdots, y(n+p)] = \left(\mathbf{h}_p(n+p)\right)_{lp+p} \tag{3.37}$$

for $l = 0, 1, \cdots, L$.

The KF makes use of the detected data symbols, $\bar{s}(n)$, fed back by the LLR-to-symbol block. These symbols are contaminated with detection error, $u(n)$, of variance $\gamma(n)$; that is $\bar{s} = s(n) + u(n)$. To take this error into account in the KF calculations, the noise variance is augmented with an additional term to give the "effective noise" $v'(n)$ given by

$$v'(n) = [\hat{g}(n;0) \ldots \hat{g}(n;L)]\,[u(n) \ldots u(n-L)]^T + v(n) \tag{3.38}$$

Assuming that the detection errors are uncorrelated with each other and independent from the channel gains,

$$\sigma_{v'}^2 = \sigma_v^2 + \sum_{l=0}^{L} E\left[|g(n;l)|^2\right]\gamma(n-l) \tag{3.39}$$

In Algorithm 3, $E\,|g(n;l)|^2]$ is approximated with $\left(\hat{\mathbf{h}}(n|n{-}1)\right)^2_{1:p:Lp+1} + \left(\mathrm{diag}(\mathbf{P})\right)^2_{1:p:Lp+1}$.

3.9.2 Complexity Analysis

The computational complexity of the EKF method of [92] is approximated as $\mathcal{O}(\delta[\delta + Q(L+1)]^2)$ floating point multiplications per symbol, per iteration, where δ and Q denote the equalizer delay and the number of basis functions, respectively. However, this method has a significantly high estimation error floor which is too high to allow low bit error rate operation for any modulation scheme with an order higher than 16-QAM. Since higher-order modulations are being proposed to provide the higher spectral efficiency needed for the latest generation of multimedia wireless applications. The computational cost of the CE-BEM based method of [112,120] (called "KF/ZPF-CE-BEM" herein) is on the order of $\mathcal{O}\left(2\delta^3 + 2\left[Q(L+1)\right]^2\right)$. The proposed method has a cost of $\mathcal{O}\left(2\delta^3 + 2\left[p(L+1)\right]^2\right)$, where p represents the order of the AR model. The cost of the ZPF is comparatively negligible for small N_f, and is not considered. In the simulations, we used $L = 2$ (3-tap channel), $N_f = 5, p = 4$ and $\delta = 5$. Comparing with the EKF, where typically $\delta = 5$ and $Q = 9$, the complexity of the proposed method is about one order of magnitude less. For "KF/ZPF-CE-BEM" with 256-QAM, one needs to select $Q = 9$, at least. Therefore, the cost of "KF/ZPF-CE-BEM" is at least three times that of the proposed method. Even with the extra cost, "KF/ZPF-CE-BEM" exhibits a severe error floor for the case of 256-QAM, as illustrated in the next section.

3.9.3 Simulations

An equipower fast-fading Rayleigh channel with $L+1 = 3$ taps and $P_l = 1/(L+1) = 1/3$ was considered. The channel was simulated based on [175]. A sampling interval of $T_s = 25 \times 10^{-6}$, and a Doppler frequency of $f_d = 400$Hz were used, giving a normalized Doppler frequency of $f_D = 0.01$. This Doppler frequency corresponds to a vehicle moving at the speed of 216 km/h, when the carrier frequency is 2GHz. A convolutional code of rate 1/2 and octal generator $[133, 171]$ were employed. Every symbol block

Algorithm 3 Channel estimation Kalman filter for the 256-QAM scheme

```
Inputs:
  • Matrix $\mathbf{E}(n)$ for $n = 1, \cdots, N_E$
  • Signal measurements: $y(n)$ for $n = 1, \cdots, N_E$
  • Symbol detection error variances: $\gamma(n)$ for $n = 1, \cdots, N_E$
Output:
  • Channel gain estimates: $\hat{\mathbf{g}}(n)$ for $n = 1, \cdots, N_E$
Working variables:
  • CE-BEM vector estimates: $\hat{\mathbf{h}}_p(n|n-1)$, $\hat{\mathbf{h}}_p(n-1|n-1)$
  • Covariance matrices for the estimated CE-BEM vector: $\mathbf{P}(n|n-1)$, $\mathbf{P}(n-1|n-1)$
  • KF gain: $\mathbf{K}(n)$
$\hat{\mathbf{h}}_p(0|0) \leftarrow E\left[\hat{\mathbf{h}}_p(0)\right] = \mathbf{0}_{(L+1)p\times 1}$
$\mathbf{P}(0|0) \leftarrow \frac{1}{(L+1)}\mathbf{I}_{(L+1)p}$
for $n = 1, 2, \ldots, N_E$ do
    $\hat{\mathbf{h}}_p(n|n-1) \leftarrow \mathbf{F}\hat{\mathbf{h}}_p(n-1|n-1)$
    $\mathbf{P}(n|n-1) \leftarrow \mathbf{F}\mathbf{P}(n-1|n-1)\mathbf{F}^H + \mathbf{Q}_w$
    $\mathbf{\Gamma}(n) \leftarrow [\gamma(n)\ \gamma(n-1)\ldots\gamma(n-L)]^T$
    $\hat{\mathbf{r}}_{gg}(n|n-1) \leftarrow \left(\hat{\mathbf{h}}_p(n|n-1)\right)^2_{1:p:Lp+1} + \left(\mathrm{diag}(\mathbf{P})\right)^2_{1:p:Lp+1}$
    $\sigma^2_{v'}(n) \leftarrow \sigma^2_v(n) + \mathbf{\Gamma}(n)\hat{\mathbf{r}}_{gg}(n|n-1)$
    $\mathbf{K}(n) \leftarrow \mathbf{P}(n|n-1)\mathbf{E}^H(n)$
        $\times\left[\mathbf{E}(n)\mathbf{P}(n|n-1)\mathbf{E}^H(n) + \sigma^2_{v'}\right]^{-1}$
    $\hat{\mathbf{h}}_p(n|n) \leftarrow \hat{\mathbf{h}}_p(n|n-1)$
        $+\mathbf{K}(n)\left[y(n) - \mathbf{E}(n)\hat{\mathbf{h}}_p(n|n-1)\right]$
    $\mathbf{P}(n|n) \leftarrow \left[\mathbf{I}_{p(L+1)} - \mathbf{K}(n)\mathbf{E}(n)\right]\mathbf{P}(n|n-1)$

    if $n > p-1$ then
        $\hat{\mathbf{g}}(n-p+1) = \left(\hat{\mathbf{h}}_p(n|n)\right)_{p:p:Lp+p}$
    end if
end for
for $k = 1, 2, \ldots, p-1$ do
    $\hat{\mathbf{g}}(N_E-k+1) = \left(\hat{\mathbf{h}}_p(n|n)\right)_{p-j:p:Lp+p-j}$
end for
```

comprised $N = 10^4$ symbols, in which $l_p = 5$ pilots were inserted per $l_s = 20$ data symbols. Pilot segments consisted of an impulse of magnitude $\sqrt{l_p}$, guarded by two zeros on each side. The pilot cost is $10\log_{10}(l_p + l_s)/l_s \approx 0.97$dB, and is not taken into account on the x-axis of the BER plots. The ZPF's normalized passband edge frequency was $f_p = f_D = 0.01$. To pick other parameters, we note that the Wiener bound on the estimation error at the highest SNR of interest, that is at $E_b/N_0 = 17$dB for 256-QAM with a rate 1/2 code, is calculated as $W \approx 2f_D\sigma_v^2 \approx 10^{-4}$. Based on the discussion made in Section 3.4, we selected the passband ripple $R_p = 0.1$dB, stopband attenuation $R_s = 15$dB, and the filter order $N_f = 5$. In block processing, $M = 4000$ and $M_E = 1000$ symbols from the previous block were used to deal with the unwanted ZPF's transient response and to train the KF, respectively (see Section 3.9.1).

An AR process of order $p = 4$ represented the gains variations. A lower p gives inferior BER, while higher order AR models become computationally unstable. The variance of the noise process as given by (3.32) is computed as $\sigma_w^2(l) \approx 6 \times 10^{-13}$ for $l = 0, 1, ..., L$. A better tracking behaviour is obtained with higher values (also reported in [92]), so we selected $\sigma_w^2(l) = 10^{-7}$. The equalizer KF was a fixed-lag KF as described in [101], with $\delta = 5$ lags.

The bit-error rate performances of Gray-mapped 16-QAM, 64-QAM and 256-QAM receivers are demonstrated in Fig. 3.20. In this figure, legend "KF/ZPF; CE-BEM(9)" refers to a CE-BEM based channel modeling using KF and ZPF with $Q = 9$ bases; legend "EKF; CE-BEM(Q=9)" refers to the EKF method of [92]. The proposed method is labeled with "KF/ZPF; AR(4)". The detector/decoder performance with perfect channel state information (CSI) performance is indicated by the line marked "Perfect channel." For the competitor methods, $Q = 9$ was selected despite the fact that the computational cost is at least twice as high as that of the proposed method. The length of the KF's state vector in the "ZP/KF, CE-BEM" method was $Q(L+1) = 27$, and that of the channel and symbol estimator EKF was $\delta + Q(L+1) = 32$, while the length of the state vector for the proposed method is only $p(L+1) = 12$. It can be seen that, in all cases, the performance of the proposed method over the SNRs of interest is within 0.3dB from the perfect channel performance and outperforms the other methods. Specifically, no BER floor is perceptible. The CE-BEM based KF/ZPF method performs well for the 16-QAM case, but suffers serious error floor under higher order modulations. The EKF method did not converge to low BERs over the indicated range for the 256-QAM case.

3.9.4 EXIT Chart Analysis

The convergence properties of the proposed method is analyzed with an EXIT chart. Fig. 3.21 shows the EXIT curves for a 256-QAM receiver at E_b/N_0 = 17dB. A sample trajectory illustrates how the extrinsic information is exchanged between the SISO components of the turbo receiver. A comparison with the EKF-based method of [92] is also made. It can be seen that the proposed method will reliably converge to high mutual information state, thus giving low bit error rates. On the contrary, the EKF cannot make the above-mentioned tunnel for the receiver to converge to a low BER. In addition, the equalizer and decoder curves intersect at higher values of I_D, I_E with the proposed method, thereby predicting a superior BER performance. These conclusions are in compliance with the BER results presented previously.

3.10 Summary

We have applied a low-cost ZPF to the output of a channel estimator KF to accurately estimate a fast-fading channel in a turbo equalizer-decoder scheme. The BER plots for the proposed estimator are within 0.3 dB of the perfect CSI case. The performance of the proposed estimator when used with an LDPC codes is examined and compared with the prior art. By virtue of the long memory of the smoother, the estimation error can be reduced to less than 2 dB of the Wiener bound, without using high-order KFs. An easy-to-deploy method was presented for ZPF design. The NMSE was shown to be consistently decreasing with the SNR. An EXIT chart analysis was performed to examine the convergence properties of the method. We showed that convergence to a low BER state is achieved after only few iterations.

Finally, instead of a high order CE-BEM, we employed a lower order AR model to characterize the channel and used the method to accurately estimate the channel in a 256-QAM scheme. It is demonstrated that this technique provides a superior bit error rate performance with a low computational cost for fast-fading channels.

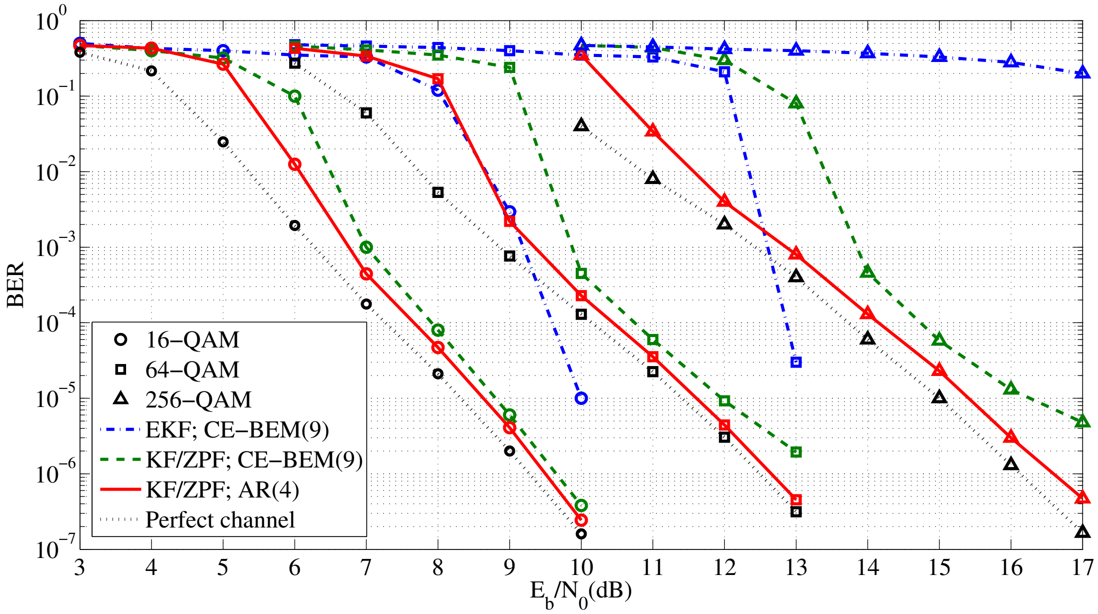

Figure 3.20: BER vs. E_b/N_0 under different modulation schemes

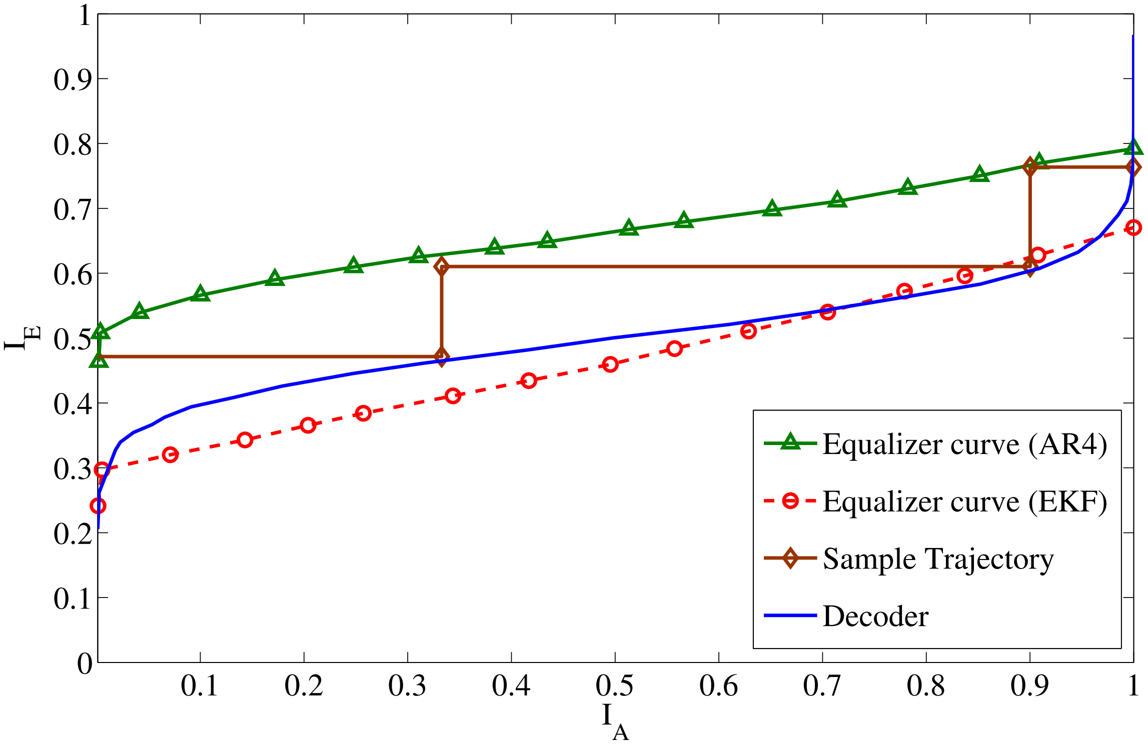

Figure 3.21: EXIT chart for a 256-QAM receiver at E_b/N_0 = 17dB.

Chapter 4

MIMO-OFDM Channel Estimation

In this chapter, the channel estimation techniques used for SISO channels are extended to single-user MIMO-OFDM systems. The focus will be on semiblind channel estimation to better accommodate channels with a large number of antennas, for which previous pilot-assisted schemes do not scale well. From a security standpoint, accurately estimated MIMO channels can provide higher secret key rates compared to the SISO channels due to the larger number of propagation paths.

We propose an efficient and accurate semiblind channel estimation technique for MIMO-OFDM turbo receivers. Once the channel is estimated using a few pilots, a low-order Kalman filter is employed to progressively predict the channel gains for the upcoming blocks. A BEM-based channel estimation scheme is used to allow the channel to vary within a block to make the method compatible with fast-fading radio channels. As the detected data symbols are iteratively used by the Kalman filter to enhance the estimation accuracy, the proposed method compares with iterative pilot-aided systems in terms of computational cost and competes in spectral efficiency with semiblind and blind estimation techniques in fast-fading environments. The BER performance of the proposed estimation approach is 0.3 dB off the perfect CSI case, whereas the computational complexity is on the order of that of near-optimal pilot-assisted methods.

The chapter is organized as follows. The basic structure of a MIMO-OFDM system is described in Section 4.1. The key idea of the approach taken in this chapter concerns a block processing technique, presented in Section 4.2. The channel estimation method is described in Section 4.3. The algorithm used for symbol detection is discussed in Section 4.4. Section 4.5 presents the computational complexity analysis. The simulation results in Section 4.6 compares the BER performance of the method with that of the perfect-CSI. The chapter is summarized in Section 4.7.

4.1 Basic MIMO-OFDM Structure

OFDM is an efficient modulation technique for high data rate communications over frequency selective channels. It simplifies the equalization problem by exploiting the properties of the FFT transform and circulant matrices to convert a delay-dispersive channel into a group of memory-less channels [115]. MIMO-OFDM systems are used for broadband wireless networks and applications such as real time video conferencing and mobile video streaming. As such, MIMO-OFDM techniques have been at the center of interest in many recent works.

MIMO systems offer either *spatial diversity* to provide reliability or *spatial multiplexing* gain to maximize the communication throughput [67]. Reliability is attained by transmitting the signal through multiple independent paths, hence combating the adverse effects of fading and increasing the delay-limited capacity [68]. The spatial diversity may be achieved by space-time codes such as the Alamouti code [9, 149] or space-frequency codes in frequency selective channels [25]. In space-time codes, coding is performed over the space (transmit antennas) and the time, whereas in space-frequency codes, coding may be applied across the space and the subcarriers of OFDM symbols as in [25], or over the space and the OFDM blocks [98].

The focus in this dissertation is on the multiplexing gain of the MIMO channel. The spectral efficiency of the channel in this case increases linearly with the number of antennas. The model for the MIMO-OFDM system is illustrated in Fig. 4.1.

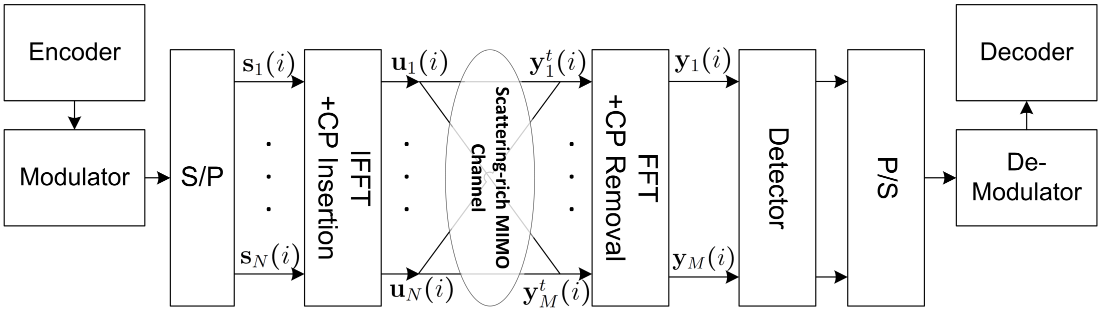

Figure 4.1: MIMO-OFDM structure.

Semi-blind Channel Estimation

As mentioned in Chapter 2, semi-blind channel estimation can save much of the pilot overhead in MIMO systems. The semi-blind technique in this section is based on a BEM representation of the channels as was described in Chapter 3. The channel variations over a block of OFDM symbols are modeled by a BEM. A block processing method then uses the channel estimates of the current BEM block to predict the channel gains over the following block. The variations of the BEM coefficients are modeled by a multi-variate AR model, and tracked by a KF over consecutive blocks. To initialize the KF, the channel is estimated at the beginning of transmission using a few pilots. The KF predicts the channel gains for the upcoming block. For each block, the detected data symbols are iteratively decoded and fed back to the Kalman filter to enhance the estimation accuracy. The proposed method's performance compares favorably with iterative pilot-aided systems and competes with semiblind and blind estimation techniques. The diversity of the MIMO channel is exploited to reduce the interleaver size and, equivalently, the latency, in practical scenarios.

4.2 System Model

In this section, the structure and input-output relationship of the MIMO-OFDM system in the time and frequency domain are presented. It is described how the symbol stream is divided into blocks and processed at the receiver.

A bit-interleaved coded modulation system transmitting as described in Section 2.3.2 over a MIMO time-varying fading channel is considered. The MIMO antenna array consists of N transmitters and M receivers. Each MIMO channel component is characterized by the model described in Eq. (2.2) of Section 2.1. For the fading rates considered in this chapter, the channel gains are assumed to be constant over an OFDM symbol of length $K+L$ where K is the number of OFDM subcarriers.

At transmitter n, a vector $\mathbf{u}_n(i) := [u_n(i,1),\cdots,u_n(i,K)]^T$ is first obtained by computing the inverse FFT of K consecutive modulated symbols, $\mathbf{s}_n(i) := [s_n(i,1)\cdots,s_n(i,K)]^T$. The modulation constellation is selected to have a mean power of unity. The vector $\mathbf{u}_n(i)$ is then prefixed with L cyclic prefix (CP) symbols defined as $u_n(i,-L+1) = u_n(i,K-L+1),\cdots u_n(i,0) = u_n(i,K)$ to form the i^{th} OFDM symbol of length $K+L$. One has $\mathbf{u}_n(i) = \mathbf{B}^H\mathbf{s}_n(i)$ where $\mathbf{B}$ is the FFT matrix. At the receiver, the CP is first removed from each block of the received signal and the remaining K output samples are used to iteratively estimate the channel, detect the data symbols, and decode the codewords. The channel output corresponding to subcarrier k of OFDM symbol i at receiver m, $y^t_m(i,k)$ can be written as

$$y^t_m(i,k) = \sum_{n=1}^{N}\sum_{l=0}^{L} g_{nm}(i,l)u_n(i,k-l) + v^t_m(i,k) \tag{4.1}$$

for $m = 1,\cdots,M$ and $k = 1,\cdots,K$, where $g_{nm}(i,l)$ denotes the channel gain from transmit antenna n to receive antenna m for a relative propagation delay of l samples for $l = 0,\cdots,L$, $u_n(i,k-l)$ represents the symbol sent by transmitter n at subcarrier $k-l$ of the OFDM symbol and $v^t_m(i,k)$ is the white Gaussian noise of variance σ^2_v. The mean power of the received signal to any receive antenna is unity.

The stream of OFDM symbols is logically divided into BEM blocks of length N_b. Each BEM block overlaps with the previous and next block in N_d symbols and differs in $N_a = N_b - N_d$ symbols as shown in Fig. 4.2. Consider tap l of the channel between transmitter n and receiver m. Collect the channel gains over BEM block j in vector $\mathbf{g}^l_{nm}(j) := [g_{nm}((j-1)N_a+1;l)\cdots g_{nm}((j-1)N_a+N_b;l)]^T$. To consider the correlation between the blocks, define $\mathbf{R}_{gg}(l) := P_l\mathcal{J}_0(2\pi f_D(0:N_b+N_a-1))$ where $\mathcal{J}_0(\cdot)$ denotes the Bessel function of the first kind and P_l is the mean power of the path. Then one may write $\mathbf{R}_g(1) := \mathrm{E}[\mathbf{g}^l_{nm}(j+1)\mathbf{g}^l_{nm}(j)^H] = \mathbf{R}_{gg}(1:N_b, N_a+1:N_a+N_b)$ and $\mathbf{R}_g(0) := \mathrm{E}[\mathbf{g}^l_{nm}(j)\mathbf{g}^l_{nm}(j)^H] = \mathbf{R}_{gg}(1:N_b,1:N_b)$, where the correlation matrices are assumed to be identical for all antenna pairs.

A vector form of (5.5) is obtained as

$$\mathbf{y}^t_m(i) = \sum_{n=1}^{N}\mathbf{D}_{nm}(i)\mathbf{u}_n(i) + \mathbf{v}^t_m(i) \tag{4.2}$$

where $\mathbf{y}^t_m(i) := [y^t_m(i,1)\cdots y^t_m(i,K)]^T$, $\mathbf{D}_{nm}(i)$ is a circulant matrix [69] with $[g_{nm}(i,0)\ \mathbf{0}_{1\times(K-L-1)}\ g_{nm}(i,L)\ \cdots\ g_{nm}(i,1)]$ as the first row and $\mathbf{v}^t_m(i) := [v^t_m(i,1)\cdots v^t_m(i,K)]^T$. The counterpart of (4.2) in the frequency domain is obtained by multiplying the equation by $\mathbf{B}$ to yield

$$\mathbf{y}_m(i) = \sum_{n=1}^{N}\mathbf{H}_{nm}(i)\mathbf{s}_n(i) + \mathbf{v}_m(i) \tag{4.3}$$

where $\mathbf{y}_m(i) := \mathbf{B}\mathbf{y}^t_m(i), \mathbf{H}_{nm}(i) := \mathbf{B}\mathbf{D}_{nm}(i)\mathbf{B}^H, \mathbf{s}_n(i) = \mathbf{B}\mathbf{u}_n(i)$ and $\mathbf{v}_m(i) = \mathbf{B}\mathbf{v}^t_m(i)$.

The transmission begins with N_t pilot OFDM symbols. Since the number of unknown channel gains per receive antenna for each BEM block equals NL, for parameter identifiability one needs to send at least $\lceil NL/K\rceil$ pilot blocks. OFDM data

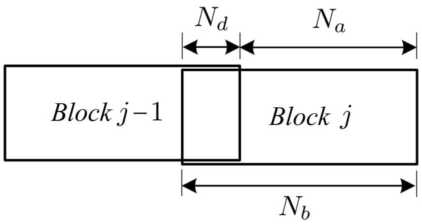

Figure 4.2: Iterative processing of the last N_a symbols of block j is performed using the last N_d symbols of block $j-1$.

follow the pilots. At time step j, the iterative processing for block j starts with estimating the channel gains for this block using the channel estimates as well as the detected data symbols for block $j-1$. Since the last N_d data symbols of block $j-1$ have already been detected, these symbols are used as virtual pilots when block j is processed. Having the channel estimates, N_a OFDM data symbols from block j are detected. The LLR of the detected symbols are input to the decoder to compute the so-called intrinsic LLR information on coded bits [38]. The *extrinsic* LLR is calculated by subtracting the decoder's input LLR from the intrinsic LLR before being fed back to the detector for the next iteration, to prevent positive feedback. The intrinsic LLR is sent to the channel estimator to enhance channel estimation accuracy in the next iteration. The iterative process of channel estimation, symbol detection, and data decoding for block j continues until convergence is reached, at which time the receiver advances to process block $j+1$.

If too short a block size is used, there may be low diversity within a block and the KF may lose the track if the instantaneous SNR at the receiver is *too low* due to channel taps being in deep fade. While this event happens quite frequently in SISO systems with short interleaver size, it becomes less likely as the space diversity of the MIMO channel increases. So, the mean-time before-failure (MTBF) time, the time between blocks where channel estimation and data detection fails, is mainly determined by the number of independent MIMO paths as well as the SNR. When too many receiver errors occur in one block, the processing system can fail to properly estimate the channel in the current block as well as for the following BEM blocks. This loss of channel state estimation (tracking) accuracy may be treated by having the transmitter to reset by sending a few pilots. Simulations show that the MTBF at the SNR range of interest is long enough to make the method viable for most channels of interest.

4.3 Channel Estimation Algorithm

Coherent symbol detection at the receiver uses the channel estimates. A bank of M Kalman filters, one per receive antenna, is employed to estimate the channel jointly with the symbol detector and data decoder. The KFs operate independently because the receive channels are assumed to be independent and thus the output equations can be separated into M independent sets. For correlated channels, this implementation would be sub-optimal. As in [166], the bandlimited channel gains are optimally represented by a Karhunen-Love transform (KLT)-calculated BEM. The BEM coefficients are assumed to be constant over a BEM block of length N_b OFDM symbols, but are allowed to vary between blocks. In this chapter, a multivariate AR(1) model [168] is used to capture the variations of the BEM coefficients between blocks, as in [166].

Consider tap l of the channel between transmit antenna n and receive antenna m. The channel gains over BEM block j for this tap are expressed as $\mathbf{g}^l_{nm}(j) := \mathbf{E}\mathbf{h}^l_{nm}(j)$ where $\mathbf{E}$ is the $N_b \times Q$ BEM matrix with $Q \geq \lceil 2f_D N_b \rceil + 1$, and $\mathbf{h}^l_{nm}(j) := [h_{nm}(j,1) \cdots h_{nm}(j,Q)]^T$. The BEM coefficients for NL taps from all transmit antennas to receive antenna m are put in a state vector defined as $\mathbf{h}_m(j) := [\mathbf{h}^0_{1m}(j) \cdots \mathbf{h}^L_{Nm}(j)]$. Using (4.3), the output equation in the frequency domain for the mth KF is written as

$$\mathbf{y}_m(j) = \mathbf{F}(j)\mathbf{h}_m(j) + \mathbf{v}_m(j) \tag{4.4}$$

where $\mathbf{F}(j) := \sqrt{K}[\text{diag}\{\mathbf{s}_1(j)\}\mathbf{E} \cdots \text{diag}\{\mathbf{s}_N(j)\}\mathbf{E}]\ (\mathbf{I}_{NL} \otimes \mathbf{B}(j,:))$.

The evolution of the BEM coefficients over BEM blocks is captured by a vector process AR(1) model [168] as

$$\mathbf{h}(j+1) = \mathbf{F}_a\mathbf{h}(j) + \mathbf{w}(j) \tag{4.5}$$

where $\mathbf{Q}_a := \text{E}[\mathbf{w}(i)\mathbf{w}^H(j)] = \mathbf{R}_h(0) - \mathbf{F}_a\mathbf{R}^H(1)$, and $\mathbf{F}_a := \mathbf{R}_h(1)\mathbf{R}_h^{-1}(0)$ with $\mathbf{R}_h(1) := \text{E}[\mathbf{h}(j+1)\mathbf{h}^H(j)]$ and $\mathbf{R}_h(0) := \text{E}[\mathbf{h}(j)\mathbf{h}^H(j)]$. These parameters can be computed using the second-order statistics of channel gain vectors over the consecutive BEM blocks as $\mathbf{R}_h(1) = \mathbf{B}^H\mathbf{R}_g(1)\mathbf{B}$, and $\mathbf{R}_h(0) = \mathbf{B}^H\mathbf{R}_g(0)\mathbf{B}$.

The channel estimator for processing the BEM block j follows Algorithm 4. To save computational cost in the case of a tall matrix $\mathbf{G}(j)$, the measurement equation is multiplied by $\mathbf{G}^H(j)$.

Algorithm 4 Channel estimation KF at step j

```
Inputs:
    • Matrix $\mathbf{F}(n)$ for $n = 1, \cdots, N_b$
    • Signal measurements: $\mathbf{y}_m(n)$ for $n = 1, \cdots, N_b$
Output:
    • Channel gain estimates: $\hat{\mathbf{g}}(:, i)$ for $n = 1, \cdots, N_a$
Working variables:
    • CE-BEM vector estimates: $\hat{\mathbf{h}}_m(j|j-1)$, $\hat{\mathbf{h}}_m(j-1|j-1)$
    • Covariance matrices for the estimated CE-BEM vector: $\mathbf{P}(j|j-1)$, $\mathbf{P}(j-1|j-1)$
    • KF gain: $\mathbf{K}$
for $m = 1, \cdots, M$ do
    $\mathbf{G}(j) \leftarrow [\mathbf{F}(1)^T \cdots \mathbf{F}(N_b)^T]^T$
    $\mathbf{y}(j) \leftarrow [\mathbf{y}_m^T(1) \cdots \mathbf{y}_m^T(N_b)]^T$
    if $\mathbf{G}(j)$ is a tall matrix then
        $\mathbf{G}(j) \leftarrow \mathbf{G}^H(j)\mathbf{G}(j)$
        $\mathbf{y}(j) \leftarrow \mathbf{G}^H(j)\mathbf{y}(j)$
        $\mathbf{R}_{vv} \leftarrow \sigma_v^2 \mathbf{G}^H(j)\mathbf{G}(j)$
    else
        $\mathbf{R}_{vv} \leftarrow \sigma_v^2 \mathbf{I}_{NLQ}$
    end if
    $\hat{\mathbf{h}}_m(j|j-1) \leftarrow \mathbf{F}_a \hat{\mathbf{h}}_m(j-1|j-1)$
    $\mathbf{P}_m(j|j-1) \leftarrow \mathbf{F}_a \mathbf{P}(j-1|j-1)\mathbf{F}_a^T + \mathbf{Q}_a$
    $\mathbf{K} \leftarrow \mathbf{P}(j|j-1)\mathbf{G}^H[\mathbf{G}\mathbf{P}(j|j-1)\mathbf{G}^H + \mathbf{R}_{vv}]^{-1}$
    $\hat{\mathbf{h}}_m(j|j) \leftarrow \hat{\mathbf{h}}_m(j|j-1) + \mathbf{K}[\mathbf{y}(j) - \mathbf{G}(j)\hat{\mathbf{h}}_m(j|j-1)]$
    $\mathbf{P}(j|j) \leftarrow [\mathbf{I}_{NLQ} - \mathbf{K}\mathbf{G}(j)]\mathbf{P}(j|j-1)$
    for $i = 1, \ldots, N_a$ do
        $\hat{\mathbf{g}}((m-1)NL + 1 : mNL, i) = \mathbf{I}_{NL} \otimes \mathbf{B}(N_b - N_a + i, :)$
    end for
end for
```

4.4 Symbol Detection

The symbols are detected using an LMMSE filter with co-channel interference cancellation as described in [2]. To cancel the interference, the extrinsic LLR from the decoder is first used to compute the mean $\bar{s}_n(i, k)$ and variance $\gamma_n(i, k)$ of the data symbol estimates (cf. Chapter 3). For each subcarrier k, to estimate $s_n(i,k), n = 1, \cdots, N$, the estimated interference of the other estimated symbols $s_{n'}(i,k), n' \neq n$ in the channel output is subtracted from it. Then the symbol is estimated using an LMMSE estimator with the modified output.

Alternatively, one could employ a soft-output sphere decoder [34] for slightly better performance but at a higher computational cost. Since the matrices $\mathbf{H}_{nm}(i)$'s are diagonal [75, p. 288], $(\mathbf{y}_m(i))_k$ in (4.3) would depend only on the k^{th} entries of the input vectors. So the system of MK equations given by (4.3) can be regrouped into K decoupled equation systems, each consisting of M equations. Define $\mathbf{y}'(i,k) := [(\mathbf{y}_1(i))_k \cdots (\mathbf{y}_M(i))_k]^T$, as comprising the kth elements of the output vectors, $\mathbf{s}'(i,k) := [(\mathbf{s}_1(i))_k \cdots (\mathbf{s}_N(i))_k]^T$, and

$$\mathbf{H}'(i,k) := \begin{bmatrix} (\mathbf{H}_{11}(i))_{k,k} & \cdots & (\mathbf{H}_{1N}(i))_{k,k} \\ \vdots & & \vdots \\ (\mathbf{H}_{M1}(i))_{k,k} & \cdots & (\mathbf{H}_{MN}(i))_{k,k} \end{bmatrix}. \tag{4.6}$$

that is, $\mathbf{H}'(i,k)$ is comprised of the kth diagonal elements of $\mathbf{H}_{mn}(i),\ m = 1, \cdots, M,\ n = 1, \cdots, N$. One has

$$\mathbf{y}'(i,k) = \mathbf{H}'(i,k)\mathbf{s}'(i,k) + \mathbf{v}'(i,k) \text{ for } k = 1, \cdots, K \tag{4.7}$$

For each of the K above-defined system, the soft-input-soft-output LMMSE detector with interference cancellation is used to estimate the data symbols as described below.

To estimate $s'_n(i,k) := (\mathbf{s}'(i,k))_n$, define

$$\Gamma(i,k) := [\gamma'_1(i,k), \cdots, \gamma'_N(i,k)]^T$$

$$\bar{\mathbf{s}'}(i,k) := [\bar{s'}_1(i,k), \cdots, \bar{s'}_N(i,k)]^T$$

where $\gamma'_n(i,k)$ and $\bar{s'}_1(i,k)$ denote the variance and mean of $s'_n(i,k), n = 1, \cdots, N$, respectively, as computed by the LLR-to-Symbol module from the decoder's output extrinsic LLR. Also,

$$\Gamma_n(i,k) := \Gamma(i,k) \,\big|_{(\Gamma(i,k))_n = 1}$$

that is, $\Gamma_n(i,k)$ is obtained by setting the nth element of $\Gamma(i,k)$ to 1. So, the fed back prior information on $s'_n(i,k)$ is not considered by the detector. Likewise,

$$\bar{\mathbf{s}'}_n(i,k) := \bar{\mathbf{s}'}(i,k)\,\Big|_{(\bar{\mathbf{s}'}(i,k))_n=0}$$

because in estimating $s'_n(i,k)$, only the interference from other symbols is canceled. To this end, the contributions made by those symbols must be subtracted from the output. The modified output is defined as

$$\tilde{\mathbf{y}'}(i,k) := \mathbf{y}'(i,k) - \mathbf{H}'(i,k)\bar{\mathbf{s}'}_n(i,k)$$

The LMMSE filter is characterized by

$$\mathbf{w}_n = \left(\mathbf{H}'(i,k)\mathrm{diag}(\Gamma_n(i,k))\mathbf{H}'^H(i,k) + \sigma_v^2\mathbf{I}_M\right)^{-1}\mathbf{H}'_n(i,k)$$

where $\mathbf{H}'_n(i,k)$ is the nth column of matrix $\mathbf{H}'(i,k)$. The symbol estimate and variance are obtained as

$$\hat{s'}_n(i,k) = \mathbf{w}_n^H\tilde{\mathbf{y}'}(i,k)$$

$$\hat{\gamma'}_n(i,k) = \left(\mathbf{H}'(i,k)\mathrm{diag}(\zeta_n(i,k))\mathbf{H}'^H(i,k) + \sigma_v^2\mathbf{I}_M\right)^{-1}$$

where $\zeta_n(i,k))$ is obtained by setting the nth element of $\Gamma(i,k)$ to zero, so that to prevent forwarding the information fed back by the decoder to itself. $\hat{s'}_n(i,k)$ and $\hat{\gamma'}_n(i,k)$ are input to the Symbol-to-LLR module to generate extrinsic LLR for the decoder, as explained in Chapter 3.

4.5 Computational Complexity

The computational complexity of our method is comparable with BEM-based pilot assisted techniques. The proposed method employs a similar BEM model and Kalman filtering as in [166], but each symbol is processed N_b/N_a times. The computational complexity of Algorithm 4 is on the same order as that of the pilot-assisted method in [166]. In Algorithm 4, the two most expensive operations are the matrix multiplication at line 5 with a cost of $\mathcal{O}(M^2N_b(NLQ)^2)$ and the inversion at line 13 with a cost of $\mathcal{O}(M(NLQ)^3)$ complex operations per symbol, for m receivers. The total cost is therefore $\mathcal{O}((M^2N_b(NLQ)^2 + M(NLQ)^3)/N_a)$ complex multiplications per symbol.

The computational complexity of the channel estimator in [35] includes a matrix inversion required to resolve the phase ambiguity [177] with a cost of $\mathcal{O}(KN^2+N^3)$ as well as $\mathcal{O}(M^3K^3L+M^4K^2L^2)$ for matrix multiplication in [35, Eqn. 3.6]. The overall cost is $\mathcal{O}((M^4K^2L^2+M^3K^3L+KN^2+N^3)/N_s)$ operations per channel use where N_s denotes the block size. This cost is greater than that of the proposed method if $K^2N_a > N_bLQ^3N_s$ and $K^3N_a > ML^2Q^3N_s$, assuming $M = N$ and $N_a \approx N_b$. This condition typically holds in the settings of interest in this research.

The approach of [94] incurs a cost of $\mathcal{O}(MK^3 + MNLK^2)$ for matrix inversion and multiplications used in recursive least-squares algorithm. This cost is higher than that of our approach when $K > NLQ$ by a factor of N_a.

For larger numbers of antennas $(M, N > 2)$ however, the cost of symbol detector used in the proposed method (and the method of [94]) is much less than the same detector used with the method of [35]. Since precoding in [35] creates correlations between the symbols, the output equations in frequency domain cannot be split up to K independent sets as explained previously for the proposed method. The LMMSE detector needs to inverse matrices as large as $KN \times KN$, which is not feasible when KN is large. Approximate inversion algorithms may be used with a reduction of accuracy. The proposed method requires inverting matrices of size M. Therefore the computational cost of the detector for the proposed scheme is orders of magnitude lower. The method can also benefit from sphere decoding to enhance the performance.

4.6 Simulations

The simulations consider a MIMO Rayleigh channel with eight-tap between each antenna pair and a power-delay profile of [0 -2.4 -6.5 -9.4 -12.7 -13.3 -15.4 -25.4] dB, normalized to a total power of unity. This profile belongs to a typical vehicular 3G wideband system at a speed of 120 km/h [155]. A nonsystematic convolutional code of rate 1/2 with a generator of (133,171) was used. The normalized Doppler spread was $f_D = f_dT_s = 10^{-4}$ with T_s and f_d denoting the sample duration and Doppler spread, respectively.

Fig. 4.3 shows the BER performance of a 4-QAM 4×4 MIMO-OFDM system compared to that of [33, 35] labeled as "Statistical method" and the approach of [94] tagged with "RLS based" in the figure. Each trial consisted of 2 training OFDM symbols followed by 2×10^5 data symbols. The OFDM symbol length is $K = 32$. For the proposed method, the BEM block size was $N_b = 10$. We picked $N_a = 8$, hence an overlap of $N_b - N_a = 2$ between consecutive BEM blocks. The number of BEM coefficients was $Q = 2$. So, the size of the BEM vector was $LNQ = 64$. For the method of [35], the parameters were tuned for the lowest possible BER. As such, a block length of 50 symbols with precoding parameter $\tau = 0.8$ [35] was used. The phase ambiguity matrix[1] for the method of [35] was assumed *known*, although a minimum number of N impulse pilots would be required to resolve it and the estimation would create some performance degradation.

The performance is as close as 0.3 dB to the perfect CSI case for the proposed method and 1 dB for the other methods. It can also be seen that the pilot rate is at least 0.25% for the statistical method but is practically zero for the proposed method as well as the RLS based approach. Regarding computational cost of the LMMSE detector, the method of [35] would need to invert matrices of size $KM = 128$, whereas the proposed method would invert matrices of size $M = 4$. The pilot-assisted method of [166] could be extended to estimate this channel, but it would require an overhead of $39/288 \approx 14\% \approx 0.6$ dB for pilots.

Fig. 4.4 shows the BER performance for a 16-QAM modulation scheme when $Q = 3$, $N_b = 60$, $N_a = 20$, hence an overlap of $N_b - N_a = 40$ symbols between adjacent BEM blocks. The method of [35] fails to converge for this setup and is not shown. In this case, failures in the convergence of the iterative receiver are observed, calling for taking the MTBF into account. The MTBF would also depend on the SNR. For the proposed method, the MTBF measured for 100 trials is as large as about 40,000 OFDM symbols at SNR=10dB, but reduces to about 4,000 at SNR=9dB and about 200 at SNR=8dB. It should be noted that in practice, these systems are designed for target BER values of about 10^{-6} making the SNR value of 10 dB the most relevant. In Fig. 4.4, the bursty errors account for the sharp rise of the BER curve at 8dB. It is seen that our method is only 0.3dB off from the perfect CSI with a very small pilot overhead. The MTBF for the RLS-based technique from [35] was about 50 symbols on the average at SNR=10 dB in 500 trials.

4.7 Summary

We introduced an accurate semiblind estimation technique for fast fading MIMO channels. The channel variations over a block of OFDM symbols are captured by a KLT based BEM. A block processing technique then used the channel estimates

[1] The channel gain vector is the product of the phase ambiguity matrix and the vector obtained from processing the output covariance matrix [35]

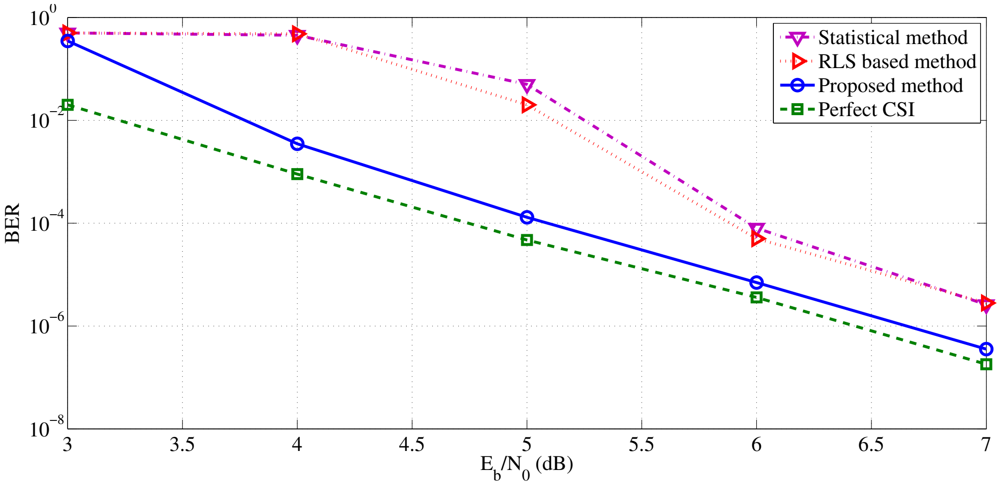

Figure 4.3: 4-QAM 4 × 4 MIMO-OFDM receiver with LMMSE detector compared with the statistical approach of [35] and the RLS based method of [94] at $f_D = 10^{-4}$.

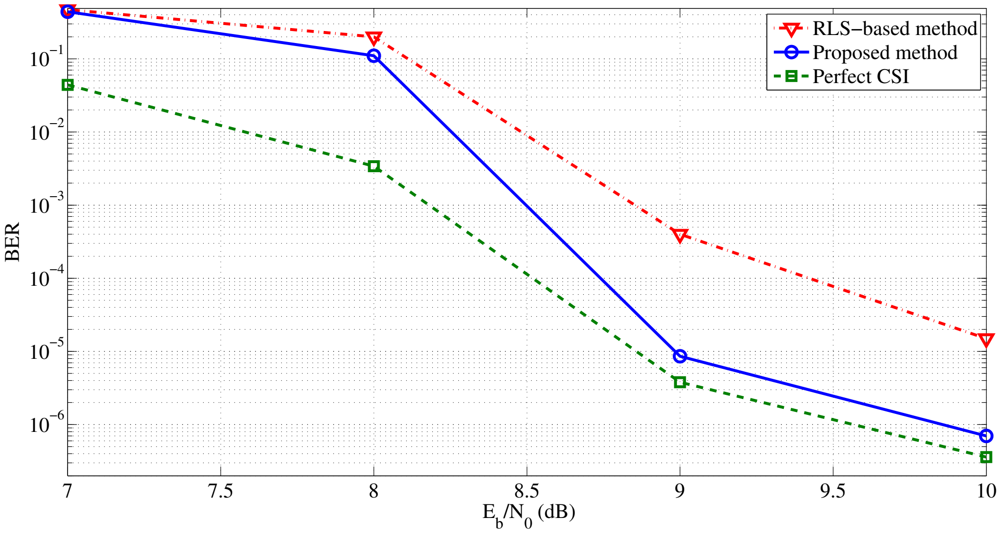

Figure 4.4: 16-QAM 4 × 4 MIMO-OFDM receiver with LMMSE detector compared with the RLS based method of [94] at $f_D = 10^{-4}$.

of the current BEM block to project the channel gains over the next block. The variations of the BEM coefficients were tracked by a KF over consecutive blocks. The proposed method's performance compares favorably with iterative pilot-aided systems and competes with semiblind and blind estimation techniques. The performance is as close as 0.3 dB to the perfect CSI case for the proposed method and 1 dB for the other methods. The diversity of the MIMO channel was exploited to reduce the interleaver size and, hence, the latency, in practical scenarios. Compared to the previous art, the proposed method was shown to excel in the MTBF, especially in higher order modulations. The complexity of the proposed method is on the same order as the pilot-based iterative methods. This technique specially lends itself to large scale MIMO radio systems due to its tiny pilot overhead and can be considered a viable approach in low-latency broadband systems as an important application of MIMO-OFDM.

The iterative channel estimation techniques proposed in this dissertation, enhance the accuracy of the channel estimates and provide better BER performances at lower computational costs compared to the prior art. However, the effect of iterative channel estimation on the capacity has not been discussed. In the next chapter, the capacity gain provided by iterative processing is studied.

Chapter 5

Capacity of Iteratively Estimated Doubly-Selective Channels

Previously in this dissertation, efficient, low-complexity methods for iterative channel estimation were introduced. It was not addressed however, how the achievable capacity of the estimated channel would be affected by iterative processing. This chapter explores the capacity of radio channels when iterative channel estimation, data detection, and decoding are employed. Knowing the capacity gain from iterative detection versus purely pilot-based channel estimation helps a designer to compare the performance of an iterative receiver against a non-iterative one and select the best balance between performance and cost.

As the secrecy capacity of a communication channel between Alice and Bob is determined by the capacity of their channel as well as the capacity of the Eve's channel, the calculations of the secrecy capacity must also consider the effect of iterative channel estimation on the capacity of these channels. Specifically, if performing iterative channel estimation by Eve significantly affects the capacity of her channel, this capacity gain must be taken into account by Alice and Bob when they rely on the inferior capacity of Eve's channel to secure the communication.

The interaction between the symbol detector and the decoder is analytically characterized and depicted in an EXIT chart, where a bound on the detector curve is found. With optimal LMMSE pilot-based channel estimation, the results of this chapter demonstrate that iterative channel estimation provides insignificant capacity advantage at fading rates below 1% of the symbol rate, though a computational-cost gain is still available. Iterative channel estimation provides a capacity benefit if sub-optimal pilot signaling is used to provide initial channel estimates.

5.1 Introduction

The iterative approach to estimating wireless channels incorporates detected data symbols into the channel estimation algorithm (see, e.g., [92,101,112,118,120]). This chapter calculates the capacity of radio channels when an iterative channel estimation system is used at the receiver.

This chapter concerns the capacity of iteratively estimated channels, where unlike [4,105,178], both pilots and soft decisions on data symbols contribute in channel estimation. Iterative processing provides a *capacity gain* by enhancing the accuracy of channel estimation [84], thereby reducing the effective noise seen by the receiver. The capacity is examined by evaluating the mutual information between the transmitted signal and the signal at the receiver given the estimated channel. This mutual information tends to increase as the result of interaction between the channel estimator/equalizer (detector) and decoder modules as the receiver iterates. Knowing the capacity gain helps a designer to compare the performance of an iterative receiver against a non-iterative one and select the best balance between performance and computational cost.

The chapter presents bounds on the EXIT curve for the joint channel estimation and detector system for iterative receivers. By calculating bounds on the LMMSE channel estimation error, a lower bound on channel capacity is calculated. From this channel error bound, a bound on the iterative detector/channel estimator EXIT curve for multipath Rayleigh fading channels is derived. For a given receiver, it is known that the available capacity is proportional to the area under the detector's EXIT curve [69]. Therefore, by bounding the EXIT curve in the presence of channel estimation error, considering iterative channel estimation, bounds on the available capacity for a given receiver are being measured. The EXIT curve also determines the type of error correction codes that should be used with a given detector/channel estimator to provide a given capacity [69]. To provide bounds on the EXIT curve, this chapter derives bounds on the channel estimation error variance for different levels of knowledge about the transmitted data signal provided by the extrinsic decoder feedback in the previous iteration. The error variance is then used to put a lower bound on the achievable channel capacity when iterative channel estimation is in use.

This chapter considers a receiver with an LMMSE channel estimator, soft-input-soft-output symbol detector, and soft-input-soft-output error correction code decoder. LMMSE channel estimators and soft-input-soft-output decoders are ubiquitous in digital radio receivers [120, 166]. To improve the accuracy, the channel estimation block uses the estimated data symbol values based on the extrinsic information fed back from the decoder. To prevent unwanted positive feedback, the input to a given block at each iteration must be almost independent of its output from the previous iteration. As in turbo decoders, this independent-information requirement is fulfilled by interleaving the coded bits before modulation. If a long memory interleaver is used and the extrinsic information from the decoder is fed back to the channel estimator/data detector of the next iteration, the estimated channel error will be independent of the feedback information. This independence assumption is parallel to the independence assumption used in EXIT chart analysis for error correction codes [29]. Ideally, the interleaver should be of infinite length. In practice, the interleaver need only be long enough to capture multiple (10 or more) independent samples of the channel gains for the system performance to approach that predicted by the EXIT chart.

To the best of our knowledge, no prior work has analytically investigated the capacity gain from iterative estimation of a doubly selective channel. The effect of non-iterative channel estimation error on capacity has been extensively studied [44, 113, 123, 174]. The capacity of pilot-based estimation schemes in non-iterative receivers has been explored in [85–87, 105]. A lower bound on the capacity of flat fading radio channels for iterative receivers in slow fading channels was derived in [53], but the detected symbols' uncertainty was not addressed. An upper bound on the capacity of non-iteratively-estimated frequency-selective channels is derived in [36]. Knowing the capacity gain of iterative processing is important when the secrecy capacity of a communication channel is evaluated based on the capacities of the main and Eve's channel. In particular, Eve's channel capacity must be calculated based on the assumption that she is capable of iterative channel estimation. Otherwise, the true secrecy capacity may be significantly lower than the one calculated based on the assumption that Eve does not perform iterative channel estimation.

The chapter demonstrates that the worst case for pilot-based channel estimation is when the received signal power is equally spread over all propagation paths with the gain of each propagation path being independent. This also holds for data aided estimation under low fading rates. This permits the easy calculation of a general lower bound on channel capacity.

The major results of this research assume that the data symbols sequence are sampled independently from a Gaussian distribution to reduce the cost of the calculations. The Gaussian assumption is made in [88] for EXIT analysis of MMSE turbo equalizers. To validate the Gaussian assumption to other realistic scenarios, the case of finite-order modulations used in practical systems is also explored. It is shown that the Gaussian approximation provides a good approximation of the EXIT curves for higher order modulations which can be calculated at a much lower cost than the full discrete modulation constellation capacity calculations.

The chapter is organized as follows. In Section 5.2, a model of the channel for the transmitted and received signals is presented. In Section 5.3, a bound on the mean-square error (MSE) for channel estimation using prior estimated data symbol values is presented. Section 5.4 presents the capacity bound of the radio channel using the channel estimates. The case of finite order modulation is explored in Section 5.5. Section 5.6 contains numerical results on the use of the capacity calculations for fast-fading radio channels. More bandwidth efficient pilot schemes are discussed in Section 5.7. A summary of this chapter is presented in Section 5.8.

5.2 System model

The linear model of a communication channel is described by $\mathbf{y} = \mathbf{Hs} + \mathbf{v}$, where $\mathbf{H}$ is the $M \times N$ impulse response matrix, and $\mathbf{s}$, $\mathbf{y}$ and $\mathbf{v}$ are the input, output and noise vectors, respectively. The capacity is achieved by Gaussian inputs and given by the mean value of the conditional information as

$$\mathcal{C} = (1/N)\mathrm{E}[\max_{\mathbf{R_{ss}}} \mathcal{I}(\mathbf{s}, \mathbf{y}|H)] \quad \text{bits/s/Hz}, \tag{5.1}$$

where $\mathbf{R_{ss}}$ is the auto-covariance of the input vector, and the average input power is constrained to some P_s. With Gaussian noise and power-constrained input, the average capacity with known $\mathbf{H}$ is [150]

$$\mathcal{C} = (1/N)\mathrm{E}\left[\max_{\mathbf{R_{ss}}} \log\det\left(\mathbf{I} + \mathbf{R_{vv}^{-1}}\mathbf{H}\mathbf{R_{ss}}\mathbf{H}^{H}\right)\right] \quad \text{bits/s/Hz} \tag{5.2}$$

with $\mathbf{R_{vv}}$ being the auto-covariance matrix of the noise and $(1/N)\mathrm{trace}(\mathbf{R_{ss}}) = P_s$.

The goal of this study is to bound (5.2) for iteratively *estimated* channels using the analytical properties of EXIT charts. The receiver iteratively performs channel estimation, symbol equalization and decoding as shown in Fig. 5.1. The equalizer and

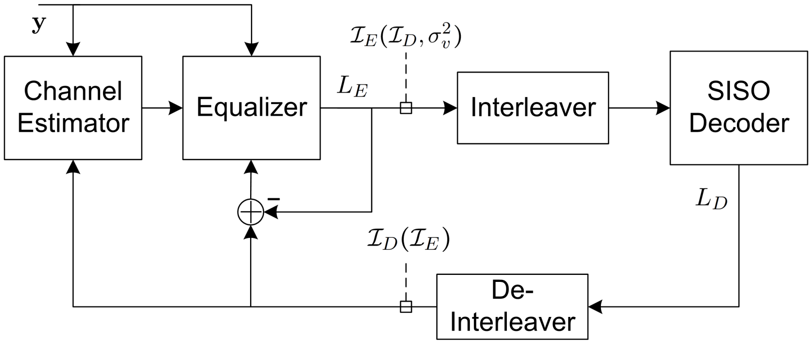

Figure 5.1: Receiver structure

soft-input-soft-output decoder generate soft information on the coded bits denoted as L_E and L_D, respectively. This soft information consists of a vector of the LLR information for the bits. The LLR for each bit is the log of the ratio of the probability of the bit's value being one over its probability of having the value zero. The equalizer and decoder blocks in Fig. 5.1 can alternatively be viewed as functions of the input mutual information. The mutual information measured at the output of the decoder, $\mathcal{I}_D$, is a function of $\mathcal{I}_E$ coming from the equalizer. The mutual information at the equalizer's output is a function of $\mathcal{I}_D$ as well as the power of channel noise σ_v^2. The problem is how iterative processing affects the channel capacity when perfect channel state information (CSI) is not available at the receiver. To make the problem tractable, the following assumptions are made:

(A1) The interleaver is sufficiently long to permit the EXIT analysis and validate the independence of channel estimates from detection errors.

(A2) An orthogonal pilot scheme such as that of [105] or [87] is used. The former was shown to be optimal in terms of bounds on capacity and LMMSE and has been used in numerous works for channel estimation, see, e.g., [120, 121, 130, 148, 157, 166, 171]. The latter is effective at high SNR regimes.

(A3) The gains for each channel propagation path are assumed to be independent. We will show that this assumption corresponds to the worst-case scenario for pilot-based estimation (cf. Lemma 2) and thus, is needed when one calculates a lower bound on capacity.

These assumptions are common in the literature of iterative systems and do not invalidate the applicability of our approach to commonly proposed iterative receiver practical systems.

For the iterative model considered in this chapter, the channel capacity is proportional to the area under the equalizer's EXIT curve $\mathcal{I}_E(\mathcal{I}_D, \sigma_v^2)$ as described below. In [12], some analytical properties of EXIT charts for a general abstract model of interacting component decoders are presented. It was shown that the area under the EXIT chart is related to the mutual information between the input ant output of the communication channel. We use this property to prove a lemma on the channel capacity for the turbo-equalizer model of Fig 5.1. It will be shown that the capacity is the average value of the equalizer EXIT function $\mathcal{I}_E(\mathcal{I}_D, \sigma_v^2)$. Before that, we recite the relevant results of [12] with slight notational changes as follows. A message is encoded to codeword $\{x_i; i = 1, \cdots, m\}$, sent through the communication channel, and received as $\{y_i; i = 1, \cdots, m\}$ by a component decoder. The codeword is also scrambled with a long interleaver, transmitted through an independent *extrinsic channel*, and received by the decoder as w_i. This independence is a common assumption in EXIT analysis that is approximated in practical receivers through the use of long interleavers between the encoding and modulation stages of the transmitter. In turbo decoders, the extrinsic channel models *a priori* information coming from a second component decoder. The first decoder then uses y_i and w_i to compute the extrinsic information e_i for the other decoder. This model includes the turbo equalizer model of this chapter as a special case where the component decoders correspond to the equalizer and decoder blocks in Fig 5.1. Moreover, the extrinsic channel corresponds to the feedback path from the decoder to the equalizer. The random variables corresponding to x_i, w_i, e_i, and y_i are denoted as X_i, W_i, E_i and Y_i, respectively. The mutual information at the input and output of the decoder is defined as $\mathcal{I}_E := (1/m)\sum_{i=1}^{m} \mathcal{I}(X_i, W_i)$ and $\mathcal{I}_D := (1/m)\sum_{i=1}^{m} \mathcal{I}(X_i, E_i)$, respectively. The area under the $\mathcal{I}_E(\mathcal{I}_D)$, $\mathcal{A}$, is given by [12]

$$\mathcal{A} = \mathcal{I}_{D,max}^2 \left[1 - \frac{H(\mathbf{X}, \mathbf{Y})}{\sum_{i=1}^{m} H(X_i)} \right] \tag{5.3}$$

where $\mathcal{I}_{D,max} := (1/m)\sum_{i=1}^{m} H(X_i)$, $\mathbf{X} := [X_1, \cdots, X_m]^T$ and $\mathbf{Y} := [Y_1, \cdots, Y_m]^T$. The following lemma can be verified.

Lemma 1. *For an M-ary modulation scheme, the average value of $\mathcal{I}_E(\mathcal{I}_D, \sigma_v^2)$ is the channel capacity given by*

$$\mathcal{C} = \frac{1}{\mathcal{I}_{D,max}} \int_0^{\mathcal{I}_{D,max}} \mathcal{I}_E(\mathcal{I}_D, \sigma_v^2) \mathrm{d}\mathcal{I}_D \tag{5.4}$$

where $\mathcal{I}_{D,max} := \log_2 M$.

Proof. In the case of M-ary modulation with all symbols having equal transmission probability, $H(X_i) = \log_2 M$, therefore $\mathcal{I}_{D,\max} = \log_2 M$. Inserting this into (5.3) and using $\mathcal{C} = \mathcal{I}(\mathbf{Y}, \mathbf{X}) = \sum_{i=1}^{m} H(X_i) - H(\mathbf{X}, \mathbf{Y})$ and the definition of $\mathcal{A}$ yields (5.4). ■

This chapter considers a bit interleaved coded modulation scheme. A block of N transmit symbols, $\mathbf{z} = [z_1 \ \ldots \ z_N]^T$ is expressed as the sum of data symbols s_i and pilots p_i, $i = 1, ..., N$ as $\mathbf{z} = \mathbf{s} + \mathbf{p}$, where $\mathbf{s} = [s_1 \cdots s_N]^T$ and $\mathbf{p} = [p_1 \cdots p_N]^T$. If a data symbol is sent at instant i then $p_i = 0$ and if a pilot symbol is transmitted at instant i then $s_i = 0$. It is assumed that $E[|s_i|^2] = 1$ for data symbol with index i. Moreover, data symbols are Gaussian distributed unless otherwise stated. We mainly focus on the pilot arrangement of [105], where every l_s data symbols in $\mathbf{s}$ are followed by l_p *null* data symbols, corresponding to the positions of the l_p pilot symbols in $\mathbf{p}$. The pilot pattern is comprised of an impulse of magnitude $\sqrt{l_p}$ guarded by at least L zeros on each side. This scheme has widely been employed for estimating fast-fading channels, and it has been shown that periodic use of this pilot pattern maximizes channel capacity for purely pilot-based channel estimation [105]. We used this arrangement in [120, 121].

The transmit signal is distorted by a doubly-selective noisy channel with a normalized Doppler frequency of $f_D = f_d T_s$ with T_s and f_d denoting the sampling interval and Doppler spread, respectively. The channel is modeled as a linear time-varying filter with $L + 1$ taps. The mean power of path l is denoted P_l, for $l = 0, ..., L$. The channel gain at time n for a relative propagation delay of l samples is denoted $g(n; l)$. One can write the channel output $y(n)$ at time n as

$$y(n) = \sum_{l=0}^{L} g(n; l) z(n - l) + v(n) \tag{5.5}$$

for $n = 1, 2, \ldots, N$, where $v(n)$ denotes the Gaussian zero-mean complex white noise with variance σ_v^2.

The channel gains for N samples are stored in vector $\mathbf{g} := [g(1; 0) \ \ldots g(N; 0) \ldots g(1; L) \ \ldots g(N; L)]^T$. The received signal, $\mathbf{y} := [y_1 \ldots y_N]^T$ is described as

$$\mathbf{y} \ = \ (\mathcal{S} + \mathcal{P})\,\mathbf{g} + \mathbf{v} \tag{5.6}$$

with the measurement noise vector $\mathbf{v} = [v_1 \ \ \ldots \ v_N]^T$ with auto-correlation matrix $\mathbf{R_{vv}} = \sigma_v^2 \mathbf{I}_N$. Matrix $\mathcal{S}$ is formed by data symbols $\mathbf{s}$, as $\mathcal{S} = [\mathcal{D}_0[\mathbf{s}] \ \ldots \mathcal{D}_L[\mathbf{s}]]$, where $\mathcal{D}_l[\mathbf{s}]$ denotes a diagonal matrix with main diagonal as the l-sample delayed version of vector $\mathbf{s}$, i.e., vector $\mathbf{s}$ prefixed with l zeros. Matrix $\mathcal{P}$ is defined likewise in terms of $\mathbf{p}$. To calculate the channel capacity, (5.6) is rewritten as

$$\mathbf{y} = \mathbf{H_s s} + \mathbf{v} \tag{5.7}$$

where $\mathbf{H_s}$ is the $N_s \times N_s$ matrix of channel gains corresponding to data symbols in the channel output equation [105].

The channel estimator calculates the optimal LMMSE estimate of $\mathbf{g}$, denoted as $\hat{\mathbf{g}}$, based on the measurement in (5.6). The estimation uses estimates of the transmitted symbols, calculated in the previous iteration of the receiver algorithm. At the first iteration, data symbols are unknown and channel estimation is based solely on the pilot symbols. At each iteration, the data symbol estimates $\bar{\mathbf{s}}$, fed back to the channel estimator, are assumed to be contaminated with *detection* error $\mathbf{u} := [u_1 \; ... \; u_N]^T$ so that $\bar{\mathbf{s}} = \mathbf{s} - \mathbf{u}$ or

$$\mathbf{u} := \mathbf{s} - \bar{\mathbf{s}}. \tag{5.8}$$

The detection error is zero mean and independent for each symbol with the variance for each symbol's error stored in vector $\Gamma := [\gamma_1 \; ... \; \gamma_N]^T$. The detection variance for pilot symbols is zero. The data symbol's estimates are based on the extrinsic information from the error correction code decoder which makes the detection error independent for each sample if a long memory interleaver is used [92]. Extrinisic information is obtained by subtracting the LLR input to the decoder from the intrinsic LLR of the decoder output [29]. Extrinsic information feedback creates statistical independence of channel estimation/data detection and decoder blocks' extrinsic outputs, simplifying the analysis of these systems [66]. The error of the channel estimation using the feedback information is derived below. This error is then used to find bounds on the EXIT curve of the detector. Equation (5.6) may be rewritten to consider symbol uncertainty as

$$\mathbf{y} = \left(\bar{\mathcal{S}} + \mathcal{P}\right)\mathbf{g} + \mathcal{U}\mathbf{g} + \mathbf{v}, \tag{5.9}$$

where $\bar{\mathcal{S}} = [\mathcal{D}_0[\bar{\mathbf{s}}\,] \; \ldots \mathcal{D}_L[\bar{\mathbf{s}}\,]]$, and $\mathcal{U} = [\mathcal{D}_0[\mathbf{u}] \; \ldots \mathcal{D}_L[\mathbf{u}]]$.

The channel gains for N samples and $L+1$ propagation paths can be described using a length $Q\,(L+1)$ vector $\mathbf{h}$ where $Q << N$ using a CE-BEM [60], so that $\mathbf{h} := [h_0(1) \; \ldots \; h_0(Q) \; h_1(1) \ldots h_1(Q) \ldots h_L(1)$
$\ldots \; h_L(Q)]^T$, and $\mathbf{g} = \mathcal{B}\mathbf{h}$, where $\mathcal{B} := \mathbf{I}_{L+1} \otimes \mathbf{E}$ and $(\mathbf{E})_{ik} = (1/\sqrt{N})e^{\jmath\omega_k i}$ for $i = 1, ..., N;\; k = 1, ..., Q$ when $Q \geq \lceil 2f_D N \rceil, \omega_q := \frac{2\pi}{N}(q - (Q+1)/2)$, $q = 1, ..., Q$ correspond-

ing to normalized frequency range of $[-f_D, f_D]$. We let $Q = \lfloor 2f_D N \rfloor + 1$. Defining $\mathbf{F} := \left(\bar{\mathcal{S}} + \mathcal{P}\right)\mathcal{B}$, the measurement equation (5.9) is then rewritten into a more useful form for channel estimation as

$$\mathbf{y} = \mathbf{F}\mathbf{h} + \mathcal{U}\mathbf{g} + \mathbf{v}, \tag{5.10}$$

As mentioned before, the detection errors are ideally independent and identically distributed for the data symbols if a long interleaver and extrinsic feedback is used. In practice, the analysis seems to give good predictions of system performance for an interleaver length $N > 10/f_D$. The vectors $\mathbf{Fh}$ and $\mathcal{U}\mathbf{g}$ are therefore uncorrelated; so, $\mathcal{U}\mathbf{g}$ can be treated as noise for LMMSE estimation.

5.3 Bound on the Mean Square Error for Channel Estimation

The estimate of the channel coefficients is denoted as $\hat{\mathbf{h}}$ so the channel coefficient error is defined as $\tilde{\mathbf{h}} := \hat{\mathbf{h}} - \mathbf{h}$. The autocorrelation matrix of $\tilde{\mathbf{h}}$ when an LMMSE estimator is used can be written as [90]

$$\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} \;=\; \left(\mathbf{F}^H \mathcal{G}^{-1}\mathbf{F} + \mathbf{R}_{\mathbf{hh}}^{-1}\right)^{-1} \tag{5.11}$$

where $\mathbf{R}_{\mathbf{hh}}$ represents the autocorrelation matrix of $\mathbf{h}$ and

$$\mathcal{G} := \mathrm{cov}(\mathcal{U}\mathbf{g} + \mathbf{v}) = \mathrm{E}[\mathcal{U}\mathbf{g}\mathbf{g}^H\mathcal{U}^H] + \sigma_v^2 \mathbf{I}_N. \tag{5.12}$$

The channel estimation error is defined as $\tilde{\mathbf{g}} := \mathbf{g} - \hat{\mathbf{g}} \;=\; \mathcal{B}\tilde{\mathbf{h}}$. Since $\mathcal{B}^H\mathcal{B} = \mathbf{I}$, we have $\mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}\} = \mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\}$. Using (5.11), the mean squared error of the channel estimator denoted by $\sigma^2_{\tilde{g}\tilde{g}}$ is given by

$$\begin{aligned}\sigma^2_{\tilde{g}\tilde{g}} &= (1/N)\mathrm{tr}\{(\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F} + \mathbf{R}_{\mathbf{hh}}^{-1})^{-1}\} \\ &= \frac{1}{N}\sum_{i=1}^{Q(L+1)} \lambda_i^{-1}\left(\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F} + \mathbf{R}_{\mathbf{hh}}^{-1}\right) \end{aligned} \tag{5.13}$$

An upper bound on $\sigma^2_{\tilde{g}\tilde{g}}$ is found by using the Weyl's inequality[1] [20] to obtain $\lambda_i(\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F} + \mathbf{R}_{\mathbf{hh}}^{-1}) \geq \lambda_i\left(\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F}\right) + \lambda_{\min}\left(\mathbf{R}_{\mathbf{hh}}^{-1}\right)$ in (5.13). When $\mathbf{R}_{\mathbf{hh}} = r\mathbf{I}$, $\lambda_i\left(\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F} + \mathbf{R}_{\mathbf{hh}}^{-1}\right) =$

[1] For Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$, $\lambda_{\min}(\mathbf{A} + \mathbf{B}) \geq \lambda_{\min}(\mathbf{A}) + \lambda_{\min}(\mathbf{B})$

$\lambda_i\left(\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F}\right) + r^{-1}$.

The rest of this section relies on the following lemmas to present a derivation for $\sigma^2_{\tilde{g}\tilde{g}}$ containing the effect of detection error on channel estimation error.

5.3.1 Worst case scenario for channel estimation error

The following lemma describes the worst case scenario for channel estimation error with purely pilot based estimation and serves as a basis to (A3).

Lemma 2. *If data symbols are unknown, then the LMMSE channel estimator's mean squared error is maximized if all channel taps are uncorrelated with equal mean power and the power spectral density of each tap is flat.*

Proof. If no transmitted data symbols are known at the receiver, $\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F}$ becomes a scaled identity matrix as $N \to \infty$, then channel estimation variance approaches

$$\begin{aligned}\sigma^2_{\tilde{g}\tilde{g}} &= \frac{1}{N}\mathrm{tr}\left(\frac{l_p}{\sigma_v^2(l_s+l_p)}\mathbf{I} + \mathbf{R}_{\mathbf{hh}}^{-1}\right)^{-1} \\ &= \frac{1}{N}\sum_{i=1}^{Q(L+1)}\left(\frac{l_p}{\sigma_v^2(l_s+l_p)} + \lambda_i^{-1}\left(\mathbf{R}_{\mathbf{hh}}\right)\right)^{-1} \qquad (5.14)\end{aligned}$$

using $\lambda_i(\mathbf{A}+\alpha\mathbf{I}) = \lambda_i(\mathbf{A})+\alpha, \forall\mathbf{A}$. The sum of eigenvalues of $\mathbf{R}_{\mathbf{hh}}$ is equal to the known mean received power from the channel. Using the fixed received channel power, the optimization of Lagrange multipliers is used to maximize (5.14) if

$$\lambda_i = \frac{N}{Q(L+1)} = \frac{1}{2f_D(L+1)} \;\forall i \qquad (5.15)$$

So, $\mathbf{R}_{\mathbf{hh}}$ will be $(1/2f_D(L+1))\mathbf{I}_{Q(L+1)}$. As $\mathbf{R}_{\mathbf{gg}} = \mathbf{B}\mathbf{R}_{\mathbf{hh}}\mathcal{B}^H$ and $\mathcal{B}$ is block diagonal, then $\mathbf{R}_{\mathbf{gg}}$ is block-diagonal also, implying that the channel taps are uncorrelated. ■

Lemma 3. *The matrix $\mathcal{G}$ defined by* (5.12) *is diagonal with diagonal elements given by*

$$(\mathcal{G})_{ii} = \sigma_v^2 + \sum_{l=0}^{L} P_l\gamma_{i-l} \text{ for } i = 1,2,...,N. \qquad (5.16)$$

Proof. Consider the block form of the $(L+1)N \times (L+1)N$ matrix $\mathbf{R}_{\mathbf{gg}}$, given by

$$\mathbf{R}_{\mathbf{gg}} := \begin{bmatrix}\mathbf{R}_{gg}^{0,0} & \cdots & \mathbf{R}_{gg}^{0,L} \\ \vdots & & \vdots \\ \mathbf{R}_{gg}^{L,0} & \cdots & \mathbf{R}_{gg}^{L,L}\end{bmatrix}. \qquad (5.17)$$

Using the definition of $\mathcal{U}$ in (5.12) and applying the property $\mathrm{E}[\mathbf{A}\mathbf{B}\mathbf{A}^H] = \mathrm{E}[\mathbf{A}(\mathrm{E}[\mathbf{B}])\mathbf{A}^H]$ for uncorrelated matrices $\mathbf{A}$ and $\mathbf{B}$ we have

$$\mathcal{G} \quad = \quad \mathrm{E}\left[\sum_{l'=0}^{L}\sum_{l=0}^{L}\mathcal{D}_l(\mathbf{u})\mathbf{R}_{gg}^{l',l}\mathcal{D}_{l'}(\mathbf{u})^H\right] + \sigma_v^2\mathbf{I}_N.$$

The entries of $\mathcal{G}$ may be expressed as

$$(\mathcal{G})_{i,j} \quad = \quad \sum_{l=0}^{L}\sum_{l'=0}^{L}\left(\mathbf{R}_{gg}^{l',l}\right)_{i,j}\mathrm{E}[u_{i-l}u_{j-l'}^*] + \sigma_v^2\delta_{ij}$$

Since the channel taps are assumed to be uncorrelated as per (A3), for $l \neq l'$, $\mathbf{R}_{gg}^{l',l} = 0$ and thus, $(\mathcal{G})_{i,j} = \sigma_v^2\delta_{ij}$. For $l = l'$, as the detection errors are assumed to be uncorrelated according to (A1), $\mathrm{E}[u_{i-l}u_{j-l}^*] = 0$ when $i \neq j$, leading to $(\mathcal{G})_{i,j} = 0$. For $i = j$,

$$(\mathcal{G})_{i,i} \quad = \quad \sigma_v^2 + \sum_l P_l\gamma(i-l)$$

This result completes the proof. ■

Lemma 4. *The matrix $\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F}$ approaches a scaled identity matrix, namely,*

$$\lim_{N\to\infty}\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F} = \beta\mathbf{I}_{Q(L+1)}, \tag{5.18}$$

almost surely, where

$$\beta := \frac{1}{l_s + l_p}\left(\frac{l_p}{\sigma_v^2} + \frac{l_s\,(1-\sigma_u^2)}{\sigma_v^2 + \sigma_u^2}\right). \tag{5.19}$$

Proof. Matrix $\mathbf{F}$ in (6.3) can be treated as the sum of $\mathbf{F_s} := \mathcal{S}\mathcal{B}$ and $\mathbf{F_p} := \mathcal{P}\mathcal{B}$. Using Assumption (A2), data and non-zero pilot symbols are fully decoupled at the channel output; that is, ${\mathbf{F_s}}^H\mathbf{F_p} = \mathbf{0}$. Therefore, one can write

$$\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F} = {\mathbf{F_s}}^H\mathcal{G}^{-1}\mathbf{F_s} + {\mathbf{F_p}}^H\mathcal{G}^{-1}\mathbf{F_p}. \tag{5.20}$$

The diagonal entries of matrix $\mathcal{G}$ in (5.16) are approximated as being σ_v^2 for pilot symbols and $\sigma_v^2 + \sigma_u^2$ for data symbols. The exact entries of the diagonal would have lower values for the L data symbols following each pilot block. Increasing these values makes the following error variance calculations slightly pessimistic with respect to the true noise variances. As a result,

$$\mathbf{F_p}^H\mathcal{G}^{-1}\mathbf{F_p} \approx l_p/(l_s+l_p)\sigma_v^2\mathbf{I} \tag{5.21}$$

by noting that $\mathbf{F_p}^H\mathbf{F_p} = l_p/(l_s+l_p)\mathbf{I}$ It should be noted that this approximation is not an essential part of the method. One may hold on to the exact values and perform the analysis accordingly. Also note that the off-diagonal elements of the matrix are strictly zero. For LMMSE estimators and at fading rates of up to 1%, $l_s >> l_p$, the approximation is excellent. . Now define $\mathbf{M_s} := \mathbf{F_s}^H\mathcal{G}^{-1}\mathbf{F_s} = (\mathbf{I}_{L+1}\otimes\mathbf{E}^H)\bar{\mathcal{S}}^H\mathcal{G}^{-1}$ $\bar{\mathcal{S}}(\mathbf{I}_{L+1}\otimes\mathbf{E})$ organized as

$$\mathbf{M_s} := \begin{bmatrix} \mathbf{M_s}^{0,0} & \cdots & \mathbf{M_s}^{0,L} \\ \vdots & & \vdots \\ \mathbf{M_s}^{L,0} & \cdots & \mathbf{M_s}^{L,L} \end{bmatrix} \tag{5.22}$$

where $\mathbf{M_s}^{l,l'} = \mathbf{E}^H\mathcal{D}_l^H(\bar{\mathbf{s}})\mathcal{G}^{-1}\mathcal{D}_{l'}(\bar{\mathbf{s}})\mathbf{E}$. Since the diagonal elements of the diagonal matrix $\mathcal{G}^{-1}$ corresponding to data symbols are approximated as $(\sigma_v^2+\sigma_u^2)^{-1}$, the elements of block $\mathbf{M_s}^{l,l'}$ can be written as

$$(\mathbf{M_s}^{l,l'})_{i,j} = \frac{1}{N(\sigma_v^2+\sigma_u^2)}\sum_{n=1}^{N}\bar{s}_{n-l}^*\bar{s}_{n-l'}e^{-\jmath(\omega_i-\omega_j)n} \tag{5.23}$$

Kolmogorov's strong law of large numbers (SLLN) [2] [82, p. 178] can be invoked to find the limit of $(\mathbf{M}^{l,l'})_{i,j}$ as $N\to\infty$. Two case are considered.

Case 1: $l = l'$

When $l = l'$ and $i \neq j$, the SLLN applies to the RHS of (5.23) to obtain

$$\lim_{N\to\infty}(\mathbf{M_s}^{l,l'})_{i,j} = \lim_{N\to\infty}\frac{1}{N(\sigma_v^2+\sigma_u^2)}\sum_{n=1}^{N}\mathrm{E}\left[|\bar{s}_{n-l}|^2\right]e^{-\jmath(\omega_i-\omega_j)n} \text{ for } l = l' \tag{5.24}$$

almost surely. To show that the RHS of (5.24) equals zero, note that $\mathrm{E}\left[|\bar{s}_{n-l}|^2\right]$ is equal to $1-\sigma_u^2$ for n corresponding to data symbols and equal to zero elsewhere. Therefore, it is periodic with a fundamental period of $T_0 := l_s+l_p$, so its frequency representation only has non-zero components at frequencies which are integer multiples of $2\pi/T_0$. On the other hand, the sampling theorem requires $1/T_0 > 2f_D = Q/N$ in order for the

[2] Let $X_1, X_2, ...$ be independent RVs and $a_n > 0$ with a_n being unbounded as $N \to \infty$. Then $a_N^{-1}\sum_{i=1}^{N}\{X_i - \mathrm{E}(X_i)\} \overset{a.s.}{\to} 0$ provided that $\sum_{i=1}^{\infty}\mathrm{Var}(X_i)/a_i^2 < \infty$

channel to be identifiable at the first iteration when the channel is estimated using only pilots. This requirement implies that $\max|\omega_i - \omega_j| = 2\pi Q/N < 2\pi/T_0$; that is the frequency component of the sequence $\mathrm{E}\left[|\bar{s}_{n-l}|^2\right]$ at $\omega_i - \omega_j$ for all $i \neq j$ is zero assuming that $N = mT_0$ for some large integer m. So $\lim_{N\to\infty}({\mathbf{M_s}}^{l,l})_{i,j} = 0$ for $i \neq j$.

When $l = l'$ and $i = j$, the SLLN-induced Equation (5.24) becomes

$$\lim_{N\to\infty} ({\mathbf{M_s}}^{l,l'})_{i,j} = \lim_{N\to\infty}\frac{1}{N(\sigma_v^2+\sigma_u^2)}\sum_{n=1}^{N} E\left[|\bar{s}_{n-l}|^2\right]\Big\} \tag{5.25}$$

To simplify (5.25), recall $\mathrm{E}\left[|\bar{s}_{n-l}|^2\right]$ is equal to $1 - \sigma_u^2$ for n corresponding to data symbols and equal to zero elsewhere. The average in (5.25) simplifies to

$$\lim_{N\to\infty}\left({\mathbf{M_s}}^{l,l}\right)_{i.i} = \lim_{N\to\infty}\frac{1}{N(\sigma_v^2+\sigma_u^2)}\sum_{n=1}^{N}|\bar{s}_{n-l}|^2 \approx \frac{l_s\,(1-\sigma_u^2)}{(l_s+l_p)(\sigma_v^2+\sigma_u^2)} \tag{5.26}$$

Equations (5.21) and (5.26) give the diagonal entries of $\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F}$ as specified by (5.19)
Case 2: $l \neq l'$

To apply the SLLN theorem to this case, the summation in (5.23) is split up into separate summations with each of the r summations containing only independent terms as as follows

$$\lim_{N\to\infty} ({\mathbf{M_s}}^{l,l'})_{i,j} = \lim_{N\to\infty}\frac{1}{N(\sigma_v^2+\sigma_u^2)}\Big(\sum_{\substack{n=rm+1\\ m=0,1,\cdots}} d_{i-j}^{l,l'}(n) + \cdots + \sum_{\substack{n=rm+r\\ m=0,1,\cdots}} d_{i-j}^{l,l'}(n)\Big) \tag{5.27}$$

where $d_{i-j}^{l,l'}(n) := \bar{s}_{n-l}^{*}\bar{s}_{n-l'}e^{-\jmath(\omega_i-\omega_j)n}$, and r is selected such that r does not divide $l - l'$. For example, if $r = l - l' + 1$, each sub-summation involves independent RVs only. One can verify that the premise of the theorem that $\sum_{n}\mathrm{Var}(d_{i-j}^{l,l'}(n))/n^2 < \infty$ is also satisfied for each sub-summation. Using the SLLN, the right hand side of (5.27) goes to zero when $l \neq l'$ because $\mathrm{E}[d_{i-j}^{l,l'}(n)] = 0$ for all n. ■

5.3.2 Approximation to LMMSE estimation error

It will be shown that if the assumptions of Lemma 4 hold true, the LMMSE of a multipath channel given by (5.13) is approximated as

$$\sigma_{\tilde{g}\tilde{g}}^2 \approx \frac{1}{N}\sum_{i=1}^{Q(L+1)}\left(\beta + \lambda_i^{-1}\left(\mathbf{R_{hh}}\right)\right)^{-1} \tag{5.28}$$

with β as in (5.19).

To validate the approximation of (5.28), the eigenvalues of matrix $\mathbf{M_s}$ are calculated to be substituted in (5.13) (Recalling from (5.20) and (5.21) that $\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F} = \mathbf{F_s}^H\mathcal{G}^{-1}\mathbf{F} + l_p/(l_s+l_p)\sigma_v^2\mathbf{I})$. The matrix $\mathbf{M}_1 := \mathbf{F}^H\mathbf{F}\mathcal{G}^{-1}$ has the same non-zero eigenvalues as $\mathbf{M_s}$ [20]. $\mathbf{M}_1$ is a random Wishart matrix for which analysis techniques have been developed in recent theoretical work [40]. The matrix $\mathbf{F}$ has N rows and $Q\,(L+1)$ columns but is completely determined by the N estimated symbol values. So, $\mathbf{F}$ is not a Wishart matrix. This creates a high degree of dependency between most entries of $\mathbf{F}$ which cannot be handled by current random matrix theory. Current state-of-the-art theory with random Wishart matrices of the form $\mathbf{Z}^H\mathbf{Z}$ only deal with cases where most entries of $\mathbf{Z}$ are independent of each other [40]. In particular, the current theory requires that each of row of $\mathbf{Z}$ is independent which is not the case with $\mathbf{M}_1$. This strong dependency of the rows prevents us from applying currently known methods of random matrix theory to eigenvalue analysis for $\mathbf{M}_1$.

Although the exact distribution of the eigenvalues seems to be beyond the current random matrix theory, Monte Carlo simulations show that the Marčenko-Pastur law [40] for Wishart matrices and the Gamma distribution are excellent approximations as far as the computation of $\sigma^2_{\tilde{g}\tilde{g}}$ is concerned. The gamma distribution describes the distribution of the sums of exponentially distributed random variables. These distributions are compared in Fig. 5.2 when $\sigma_u^2 = 0$ (for known-data case), $f_D = 0.01$ and SNR=10 dB. The parameters of the Marčenko-Pastur law and the Gamma distribution are selected such that the mean is equal to $\beta = 10$ and the variance is equal to $Q(L+1)/N$. The Gamma distribution is a good match to the actual eigenvalue distributions. Fig. 5.3 illustrates the variance of the channel estimation error $\sigma^2_{\tilde{g}\tilde{g}}$ against various f_D for $N = 2\times10^5$. The case of identical eigenvalues as the approximation we made in this chapter is also shown. As expected, this latter corresponds to a lower variance of estimation error. The cases of Marčenko-Pastur law and the Gamma distribution are indistinguishable in the figure for the fading rates of interest, and either can be used to obtain an upper bound $\sigma^2_{\tilde{g}\tilde{g}}$.

In the worst-case scenario for channel estimation error as outlined by Lemma 2, the channel estimation LMMSE takes an even simpler form as follows.

Corollary 1. *For channels with flat power spectral density, the LMMSE is given by*

$$\sigma^2_{\tilde{g}\tilde{g}} \approx \left(\frac{\beta}{2f_D(L+1)} + 1\right)^{-1} \tag{5.29}$$

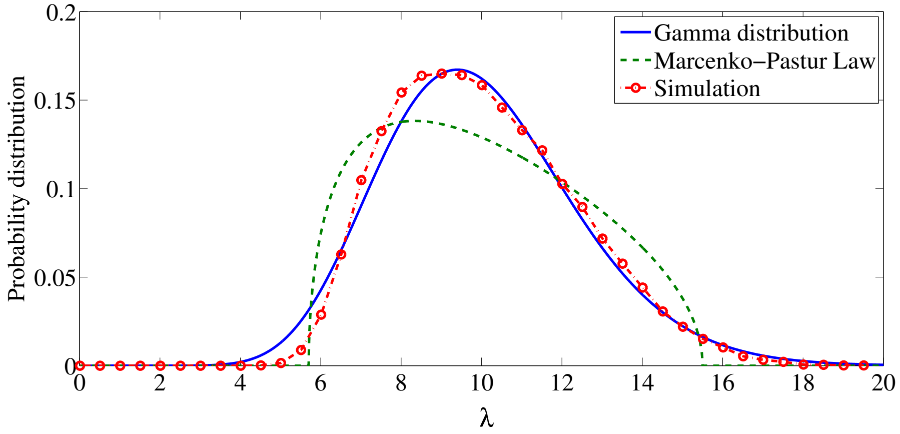

Figure 5.2: Distribution of eigenvalues for known data symbols ($\sigma_u^2 = 1$); $L = 2$, $f_D = 0.01$, SNR=10dB

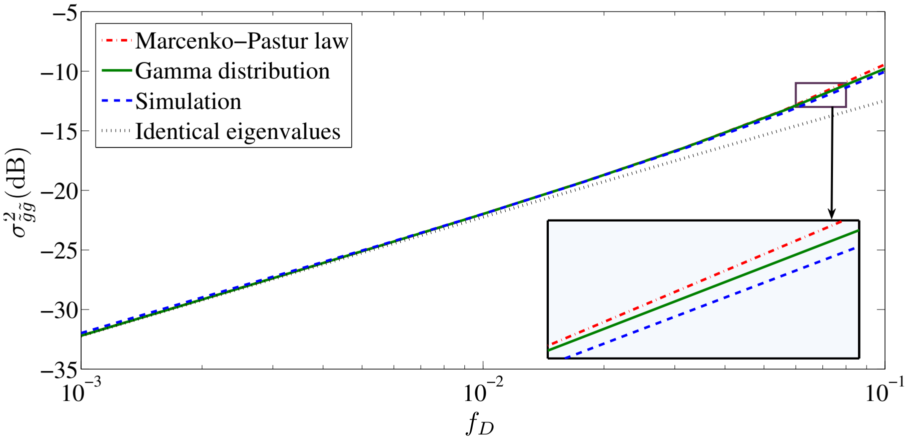

Figure 5.3: $\sigma^2_{\tilde{g}\tilde{g}}$(dB) versus f_D for known data symbols ($\sigma_u^2 = 0$); $L = 2$, SNR=10dB

Proof. Inserting λ_i from (5.15) into (5.28) and noting $Q \approx 2f_D N$, gives (5.29). ■

The relevance of the above result comes from the experimental observation that for Rayleigh channels, (5.29) and (5.28) can be used interchangeably, even if the eigenvalues of channel process are not equal.

5.4 A Lower Bound on Capacity

In this section, a lower bound on the channel capacity is calculated using (5.4). For an iterative receiver with perfect CSI at the receiver, where independent extrinsic information is fed back to the equalizer, Eq. (5.2) gives the mutual information at the equalizer's output, $\mathcal{I}_E$. With imperfect channel knowledge, if the channel estimates are assumed to be independent of the decision on data symbols as per (A1), the results of [105] on the capacity with channel estimation error hold. The output equation is rewritten as $\mathbf{y} = (\mathcal{S}+\mathcal{P})\mathcal{B}\hat{\mathbf{h}} - (\mathcal{S}+\mathcal{P})\mathcal{B}\tilde{\mathbf{h}} + \mathbf{v}$. The *effective noise* [16] is defined as $\mathbf{v}' := -(\mathcal{S}+\mathcal{P})\mathcal{B}\tilde{\mathbf{h}} + \mathbf{v}$. Now (5.2) turns into the following lower bound on $\mathcal{I}_E$ [105].

Lemma 5. *For Gaussian transmit vector* $\mathbf{s}$*, if the channel estimates are assumed to be independent of the decision on data symbols, then the mutual information at the equalizer's output is lower bounded by*

$$\mathcal{I}_E \geq \frac{1}{N}\mathrm{E}\left[\log\det\left(\mathbf{I} + \mathbf{R}_{\mathbf{v}'\mathbf{v}'}^{-1}\hat{\mathbf{H}}_{\mathbf{s}}\hat{\mathbf{H}}_{\mathbf{s}}^H\right)\right] \quad \textit{bits/s/Hz} \tag{5.30}$$

where $\mathbf{R}_{\mathbf{v}'\mathbf{v}'}$ *is the covariance matrix of the effective noise and* $\hat{\mathbf{H}}_{\mathbf{s}}$ *denotes the LMMSE-estimated channel matrix.*

Note that it is assumed $\mathbf{R}_{\mathbf{ss}} = \mathbf{I}_{N_s}$ in (5.2) because no CSI is available at the transmitter [150]. The covariance matrix of the effective noise is written as $\mathbf{R}_{\mathbf{v}'\mathbf{v}'} = E[(\mathcal{S}+\mathcal{P})\mathcal{B}\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\mathcal{B}^H(\mathcal{S}+\mathcal{P})^H] + \sigma_v^2\mathbf{I}$. This matrix has a similar structure to matrix $\mathbf{F}^H\mathcal{G}^{-1}\mathbf{F}$ discussed in the previous section with $\sigma_u^2 = 0$ and thus, can be approximated as a scaled diagonal matrix $\mathbf{R}_{\mathbf{v}'\mathbf{v}'} := (\sigma_{\tilde{g}\tilde{g}}^2 + \sigma_v^2)\mathbf{I}$ as $N \to \infty$ using the similar argument and (5.11). Because for the LMMSE estimator of $\mathbf{g}$ one has $\sigma_{\hat{g}\hat{g}}^2 = 1 - \sigma_{\tilde{g}\tilde{g}}^2$, the effective SNR for capacity calculation is obtained as $\rho_{\text{eff}} = (1-\sigma_{\tilde{g}\tilde{g}}^2)/(\sigma_{\tilde{g}\tilde{g}}^2 + \sigma_v^2)$. So, the SNR is reduced by two different factors: an increase in effective channel noise and a reduction in the variance of the channel gains [16]. Using the effective SNR and following the approach of [16, 36], the lower bound on $\mathcal{I}_E$ described by (5.30) can be written as

$$\mathcal{I}_E \geq \frac{N_s}{N^2}\sum_{i=1}^{N_s}\mathrm{E}\left[\log\left(1 + \frac{(1-\sigma_{\tilde{g}\tilde{g}}^2)}{\sigma_{\tilde{g}\tilde{g}}^2 + \sigma_v^2}\lambda_i\left(\mathbf{H}_{\mathbf{s}}\mathbf{H}_{\mathbf{s}}^H\right)\right)\right]\text{bits/s/Hz} \tag{5.31}$$

where $N_s = l_sN/(l_s+l_p)$ is the number of data symbols for the pilot scheme of [105], and the scale factor N_s/N accounts for the throughput loss due to pilots. Equation (5.31) takes into account the effect of detection error in the previous iteration when $\sigma_{\tilde{g}\tilde{g}}^2$ is given as in (5.28), and is used to analyze the turbo receiver performance using (5.4).

At low SNR's, the effective noise $\mathbf{v}'$ is almost Gaussian, and the bound in (5.31) is tight [3, 71, 164]. At high SNR's, as in [3], one may reasonably assume that (5.31) presents a tight lower bound, since the likelihood that an error in channel estimation will improve the error performance is very slim.

In a turbo receiver, data symbols are detected based on the estimated channel. Then an error correction decoder is used to reduce error in the data symbols. Refined estimates of data symbols are fed back to the channel estimator to further generate more accurate channel estimates, which in turn contribute to improve symbol detection. As the receiver iterates, more accurate data symbols are available to the channel estimator. At the final stage when the receiver converges, the channel estimation is performed with almost all data symbols being perfectly known. Therefore, the "known data" performance of channel estimator determines the actual channel capacity for the final iterations of the receiver.

The EXIT chart method of [29, 66] combines the EXIT curve of the decoder with the bound on the EXIT curve of the channel estimator and detector to find the full capacity, $\mathcal{I}_E$, of the iterative receiver. The mutual information from the decoder's feedback is given by $\mathcal{I}_D = \mathcal{I}\left(s, \bar{s}\right)$ where $\mathcal{I}\left(s, \bar{s}\right)$ is the mutual information between the true data symbols and the estimated data symbols. $\mathcal{I}_D$ is expressed in terms of the error variance of the estimated data symbols σ_u^2. As $E[|s|^2] = 1$, we have

$$\mathcal{I}_D := \mathcal{I}(\bar{s}; s) = \mathcal{H}(s) - \mathcal{H}(s|\bar{s}) = \log(1/\sigma_u^2) \tag{5.32}$$

where $\sigma_u^2 \in [0, 1]$ using that s and $\bar{s}$ are jointly Gaussian [41]. Supposing that all other parameters are fixed, the detector's mutual information as given by (5.31) is a function of $\sigma_{\tilde{g}\tilde{g}}^2$, which is in turn a function of σ_u^2 through (5.28) and (5.19). Equation (5.32) gives σ_u^2 as a function of $\mathcal{I}_D$. Therefore, $\mathcal{I}_E$, the mutual information at the output of the symbol detector, is a function of $\mathcal{I}_D$. At each iteration, the error correction code uses the output of the symbol detector to perform soft decoding and recompute the extrinsic information to be fed back to the channel estimator and symbol detection systems for the next round. This function of the decoder is represented by the *code curve* in the EXIT plot, whereas the detector function is depicted with the *detector curve*. The latter takes $\mathcal{I}_D$ as input and outputs detector's extrinsic information, denoted with $\mathcal{I}_E$ in (5.31). As one block's output is the other's input, the curves can be plotted in the same diagram with an axis swap. The exchange of mutual information between the detector and the decoder is shown by a trajectory bouncing back and forth between the two curves. The information exchange continues until the system converges to the state where the detector and code EXIT curves intersect. The convergence state of the decoder's mutual information is the actual capacity of the system, and evidently, depends on the performance of both blocks. The area under the detector's EXIT curve is proportional to the capacity of the system; a higher EXIT curve indicates the system has higher capacity [69]. Prior literature on detector EXIT curves are normalized for mutual information in the range $[0,1]$. To convert the EXIT curves derived in this chapter to these curves, scale the $\mathcal{I}_D$ and $\mathcal{I}_E$ values by $B = log_2 M$ where M is the order of modulation used after clipping the curves for $\mathcal{I}_D > B$ and $\mathcal{I}_E > B$.

5.5 Finite Order Modulation

This section considers the case of channel capacity when symbols are selected from a finite constellation $\chi = \{\zeta_1, \zeta_2, ..., \zeta_M\}$ of size $|\chi| = M$. Reference [23] derives a formula for the capacity of a non-fading AWGN channel when inputs selected from a finite order constellation. This formula is extended to doubly selective channels below.

Consider a transmit block of N symbols, $\mathbf{s} \in \chi_{\mathbf{s}}$, where $|\chi_{\mathbf{s}}| = N^M$. As in [23], it is assumed that all data symbols have equal probability $P(\mathbf{s} = \mathbf{s}_j) = N^{-M}$ for $\mathbf{s}_j \in \chi_{\mathbf{s}}$. Under this assumption, the channel capacity with finite order modulation as a function of SNR, denoted with $\mathcal{I}_{E_f}(\text{SNR})$, is obtained by averaging over channel realizations $\mathbf{H_s}$ as

$$\mathcal{I}_{E_f}(\text{SNR}) = \frac{1}{N} E\{\mathcal{I}(\mathbf{y}; \mathbf{s}|\mathbf{H_s})\} \text{ bits/s/Hz} \tag{5.33}$$

where

$$\mathcal{I}(\mathbf{y}; \mathbf{s}|\mathbf{H_s}) = \mathcal{H}(\mathbf{y}|\mathbf{H_s}) - \mathcal{H}(\mathbf{y}|\mathbf{s}, \mathbf{H_s}) \tag{5.34}$$

The conditional random vector $\mathbf{y}$ given $\mathbf{s}$ and $\mathbf{H_s}$ is Gaussian with mean $\mathbf{H_s s}$ and covariance $\sigma_v^2 \mathbf{I}_N$. Using this fact and the properties of entropy, the channel capacity given the channel condition is given as

$$\begin{aligned} \mathcal{I}(\mathbf{y}; \mathbf{s}|\mathbf{H_s}) &= -N \log(\pi e \sigma_v^2) - \sum_{\mathbf{s}_j \in \chi_{\mathbf{s}}} N^{-M} \int_{\mathbf{y} \in \mathcal{Y}} \Big[p(\mathbf{y}|\mathbf{s}_j, \mathbf{H_s}) \\ &\times \ \log \sum_{\mathbf{s}_k \in \chi_{\mathbf{s}}} N^{-M} p(\mathbf{y}|\mathbf{s}_k, \mathbf{H_s}) \Big] \mathrm{d}\mathbf{y} \end{aligned} \tag{5.35}$$

Note that $p(\mathbf{y}|\mathbf{s}_j, \mathbf{H_s}) = (\pi\sigma_v^2)^{-N} \exp(-(\mathbf{y} - \mathbf{H_s s}_j)^H (\mathbf{y} - \mathbf{H_s s}_j)/\sigma_v^2)$. The value of (5.35) can be computed using Monte Carlo techniques.

For estimated channels, Eq. (5.33) replaces Eq. (5.31) with $\hat{\mathbf{H}}_{\mathbf{s}} = \sqrt{1-\sigma^2_{\tilde{g}\tilde{g}}}\mathbf{H}_{\mathbf{s}}$ replacing $\mathbf{H}_{\mathbf{s}}$ in (5.33). The effective noise has a variance of $E[|s\tilde{g}+v|^2] = \sigma^2_{\tilde{g}\tilde{g}} + \sigma^2_v$. Assuming that detection error $\mathbf{u}$ is Gaussian and noting that $E|\hat{s}|^2 = 1-\sigma^2_u$, Eq (5.32) is rewritten in the finite modulation case as

$$\mathcal{I}_D = \mathcal{I}_{E_f}((1-\sigma^2_u)/\sigma^2_u) \tag{5.36}$$

where function $\mathcal{I}_{E_f}(\cdot)$ is defined in (5.33).

5.6 Numerical Results

In this section, we first consider the case of Gaussian distributed symbols. A doubly selective Rayleigh model with three equal power taps (delay $L = 2$ samples) characterizes the radio channel. The time-variant channel impulse response $g(n;l)$ for tap l; $l = 0,\cdots,L$ is an independent complex Gaussian process with zero-mean and variance $P_l = 1/(L+1)$, independent from other taps impulse responses, and is generated using the method of [175]. The lower bound on detector's mutual information as given by (5.31) is computed through Monte Carlo simulations with 500 trials per experiment. The mean power of data symbols was taken to be $\bar{\mathcal{P}}_s = 1$. For the worst-case scenarios, (5.29) and (5.19) are used to compute $\sigma^2_{\tilde{g}\tilde{g}}$. For the case of Rayleigh channel, (5.28) and (5.19) are used to evaluate $\sigma^2_{\tilde{g}\tilde{g}}$. The default setting for most of the simulations is $l_s = 20, l_p = 5, f_D = 0.01$. The normalized Doppler frequency of $f_D = 0.01$ corresponds to a sampling interval of $T_s = 25\mu$s and a Doppler spread $f_d = 400$Hz, which matches a fading process for a radio signal with a carrier frequency of 2GHz, to communicate with a vehicle moving at 216 km/h. The mutual information generated by the decoder $\mathcal{I}_D$ is computed using (5.32), where σ^2_u is varied in the interval $[10^{-4}, 1]$. In the capacity calculations for finite order modulations, the pilot overhead is taken into account by multiplying $\mathcal{I}_{E_f}$ in (5.33) by $l_s/(l_s+l_p)$.

Fig. 5.4 shows the effect of pilot rate on capacity when $l_p = 5$ and l_s is varied. As l_s increases, the capacity with pilot based estimation (the value of the curves for $\mathcal{I}_D = 0$) eventually hits a maximum value for $l_s = 30$ after which it decreases. The starting point is important to the convergence of the iterative receiver [29]. One may use this plot to find the minimum pilot rate corresponding to the highest capacity. This plot also shows how fast the receiver would converge to the final state of maximum mutual information at the right, and what the capacity gain of iterative estimation and decoding would be.

As mentioned earlier, the attainable capacity is equal to the average value of the detector curve $\mathcal{I}_E(\mathcal{I}_D)$. For any detector curve, the non-iterative capacity is simply the level of $\mathcal{I}_E$ for $\mathcal{I}_D = 0$, the curve's starting point. Fig. 5.5 depicts the effect of the normalized Doppler frequency, f_D, on the capacity gain at SNR=$7dB$. This figure shows that with higher Doppler frequency, the more capacity gain that iterative channel estimation provides. Iterative detection provides only a minor improvement for $f_D = 0.005$ but is significant for $f_D = 0.02$. This shows that iterative channel estimation for slower fading channels offers little extra capacity over purely pilot-based channel estimation if optimal LMMSE channel estimation is used.

The SNR penalty at a given BER is defined by $\eta := (E_b/N_0)_1 - (E_b/N_0)_2$, where $(E_b/N_0)_1$ and $(E_b/N_0)_2$ represent the SNR required to achieve the BER by using iterative and non-iterative estimation, respectively. The SNR penalty η of using a purely pilot-based channel estimation is illustrated in Fig. 5.6 for various data to pilot ratio l_s/l_p. Here, a 64-QAM receiver with a target BER of 10^{-7} is considered. In Fig. 5.6 we set $(E_b/N_0)_1 = 13dB$. The data depicted in this figure clearly validate the analytical results presented by this chapter, and are in accordance with the intuition that the iterative estimation is more beneficial with lower pilot ratios.

In Fig. 5.7, a comparison between the theoretical capacity lower bound as obtained in this chapter for $\mathcal{I}_D \to \infty$, and that of a simulated turbo-receiver under different modulation schemes has been made. The channel was estimated with a CE-BEM based Kalman filter using $Q = 15$. The term "Gaussian" in the legends denote Gaussian distributed data symbols, as opposed to the finite constellations of 4-QAM, 16-QAM and 64-QAM with a rate 1/2 convolution error correction code [120]. Legends "Known-data C.E" and "Pilot-based C.E." refer to the cases when the channel is estimated based on either known data symbols or only pilots symbols at the receiver. Legend "Perfect CSI" refers to the capacity of a receiver with perfect channel state information (CSI). These results show that optimal LMMSE pilot-based estimation can achieve almost the same capacity as presented iterative receivers. Simulations for finite-order modulation were performed by using (5.29) to calculate $\sigma^2_{\tilde{g}\tilde{g}}$. Monte Carlo method was used to evaluate the detector's mutual information using (5.33). For estimated channel, $\hat{\mathbf{H}}_\mathbf{s} = (1 - \sigma^2_{\tilde{g}\tilde{g}})\mathbf{H}_\mathbf{s}$ replaced $\mathbf{H}_\mathbf{s}$ in (5.33) and the effective noise power was modified to $\sigma^2_{\tilde{g}\tilde{g}} + \sigma^2_v$ (as in [105]). Eq. (5.36) was used to calculate $\mathcal{I}_D$ in the case of finite order modulation. The widening gap between the capacity of Gaussian transmission and that of the finite-order modulations as SNR increases is due to the suboptimality of the modulation and code used in the simulations and has been reported in the previous work [143]. Conventional coded modulation schemes allocate the same amount of parity bits to different modulation levels, ignoring the fact that less significant bits of modulated symbols need more error protection whereas more significant bits need less [165]. This non-optimal use of the parity bits leads to an SNR loss. By contrast, more efficient coding schemes such as multi-level codes make better use of channel capacity at higher SNRs.

Fig. 5.8 illustrates the capacity of an iteratively-estimated single-tap fading channel at SNR=7dB, when finite constellations are used, $l_p = 1$ and $l_s = 20$. The capacity of Gaussian transmission is also shown. In this figure, $\mathcal{I}_E$ denotes the mutual information at the output of the detector as given by (5.33). The capacity of an independent identically distributed (i.i.d.) flat fading channel was obtained by setting $N = 1$ in (5.33). The functional curve of a nonsystematic convolutional code of rate 1/2 with an octal generator of (133,171) for the 4-QAM modulation is also shown. The interchange of mutual information between the decoder and the detector is represented by a typical trajectory. Note that in this figure, the code curve for modulation of order M spans the range $[0 \ \log_2 M]$ bits on both axes; an inverse scaling factor of $R_{max} = log_2 M$ must be applied to convert to the $\mathcal{I}_D \in [0, R_{\max}]$ and $C \in [0, R_{\max}]$ region of the capacity graphs to obtain the standard EXIT charts used in much of the prior literature.

This type of plot can be used to select a code for the given modulation scheme. The code should be selected so that a tunnel between the detector and decoder curves to the low BER region of the exit chart ($\mathcal{I}_D = 1$) exists. After plotting the detector curves, one for example can verify that while the code rate 1/2 would work for a 4-QAM at SNR=7dB, a code rate of 1/4 with generator (117,127,155,171) should be used for 16-QAM at this SNR, but this code would waste a great deal of capacity as indicated by the large area between the detector and code curves. It can also be seen that iterative estimation benefits a Gaussian transmission roughly as much as it does a sufficiently high order modulation. Also, for higher order modulations, the capacity is close to the Gaussian case. This suggests that the results obtained in a previous section for Gaussian symbols can be extended to finite order modulation, provided that, for the given SNR, the order of modulation is large enough so that the finite-order modulation performs close to the Gaussian transmission.

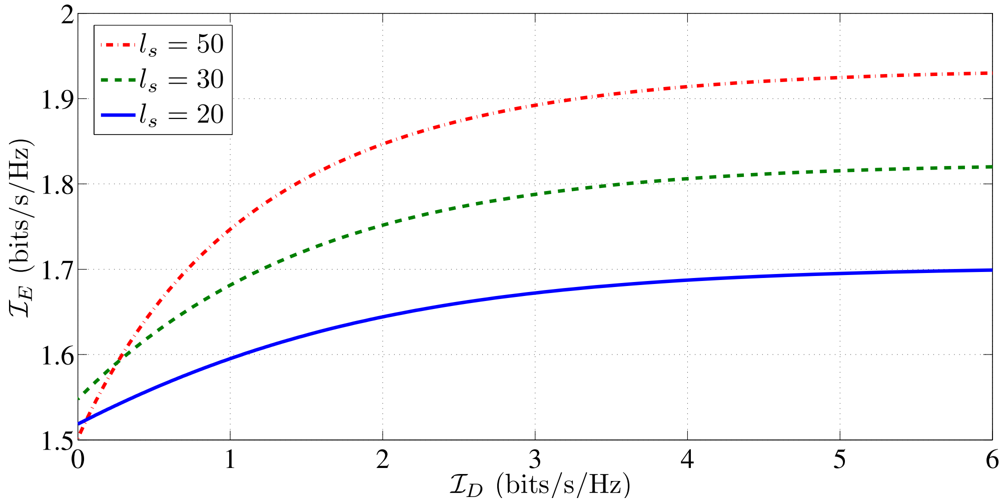

Figure 5.4: Lower bound on capacity under various pilot rates, for SNR=$7dB$, $l_p = 5, f_D = 0.01, L = 2$

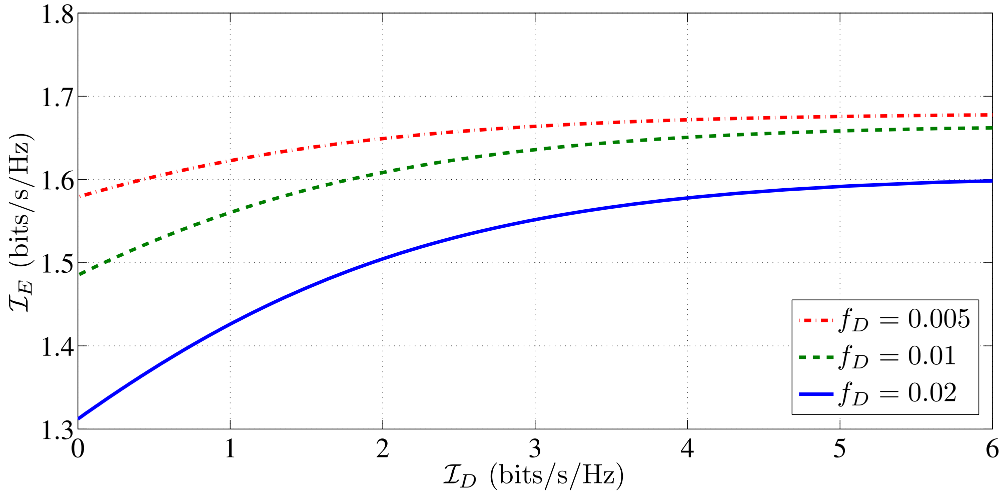

Figure 5.5: Lower bound on capacity under various fading rates, for SNR=$7dB$, $l_p = 5, l_s = 20, L = 2$

5.7 More bandwidth efficient pilot schemes

The operating SNR of the system can be traded for bandwidth efficiency if the pilot segments are concentrated in the beginning of the block [87], where each pilot segment is comprised of an impulse followed by L zeros. At lower SNRs, this technique calls for a short block length, in which case the resulting truncation error of basis expansion modeling can no longer be ignored. In this section, a somewhat more general form of

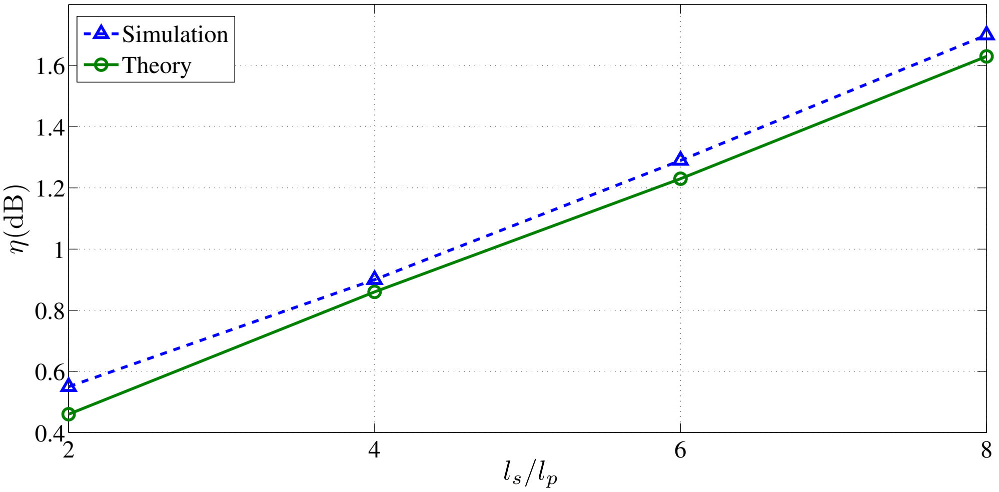

Figure 5.6: SNR penalty versus data-to-pilot ratio l_s/l_p, $f_D = 0.01, L = 2$

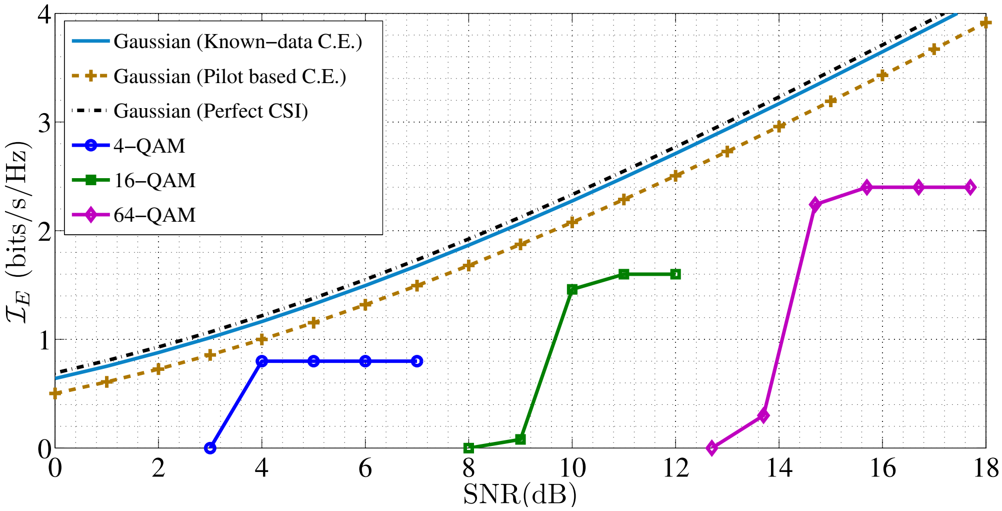

Figure 5.7: Theoretical bounds on the channel capacity compared to the capacity of an iterative receiver with $l_s = 20, l_p = 5, f_D = 0.01, L = 2$

the pilot scheme of [87] is considered. Given positive integers N, l_s and M, we define the pilot segment as $\mathbf{p}_0 := [\mathbf{0}_L \; p \; \mathbf{0}_L \cdots p \; \mathbf{0}_L]$ of length $N_{pb} := L + M(L+1)$. After each pilot segment $\mathbf{p}_0$ are Ml_s data symbols. The arrangement for $M = 1$ corresponds to the pilot symbol regime of [105]. The pilot power is fixed at $P_p := (2L+1)/(2L+1+l_s)$, and data power as $P_d := 1 - P_p$. This power allocation gives a mean power of unity to both pilots and data symbols when $M = 1$ and is a common practice [92,101,120,157].

The performance of the iterative receiver with different pilot arrangements for

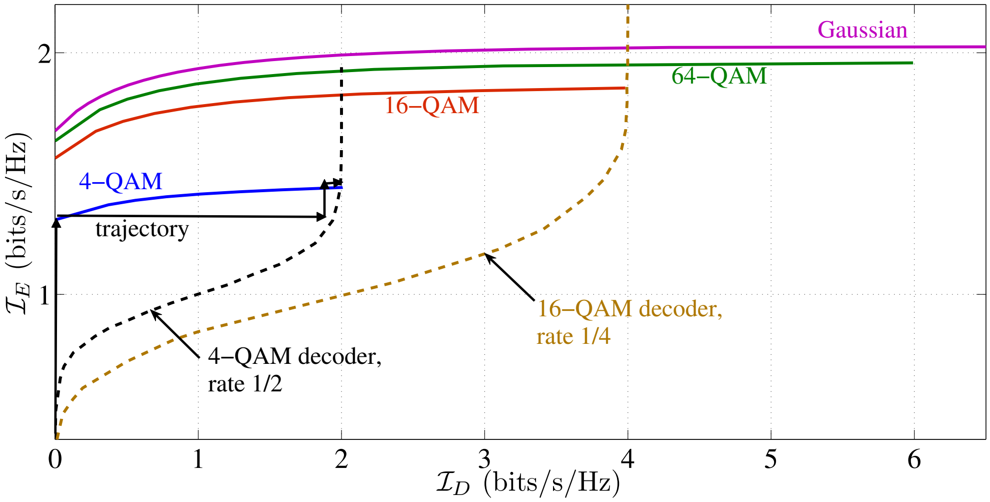

Figure 5.8: Detector's mutual information $\mathcal{I}_{E_f}$ versus $\mathcal{I}_D$ for an iteratively estimated single-tap fading channel using 4-QAM, 16-QAM and 64-QAM modulation compared with Gaussian transmission, at SNR=7dB

$N = 1000$ is shown in Fig. 5.9. The "Lumped pilots" line refers to a system using the scheme from [87] where a pilot block for $M = \lceil (N - L)/(l_s + L + 1) \rceil = 44$ is used to estimate a channel block for a block of length N samples. Larger values of M give poor mutual information results for $\mathcal{I}_D = 0$, so these configurations will give a poor performance with non-iterative receivers. The capacity with iterative detection/decoding depends on the area under the whole detector curve so higher values of M may give better performance overall. As M is reduced, the initial channel estimate error is improved, making the capacity for $\mathcal{I}_D = 0$ higher, but lowering the capacity as $\mathcal{I}_D \to \infty$ as the pilot signals consume more of the transmitted signal samples. For large M, the convergence of the receiver becomes an issue. If the receiver is able to converge, these results show that iterative processing for $M > 1$ offers higher capacity gain compared to the case $M = 1$.

5.8 Summary

The capacity advantage of an iterative receiver over a non-iterative channel estimator was evaluated. By taking the uncertainty in decoded data bits into account, the channel estimation LMMSE of an iterative receiver with a given pilot ratio was obtained. The LMMSE was then used to derive a bound on capacity. The simulations results

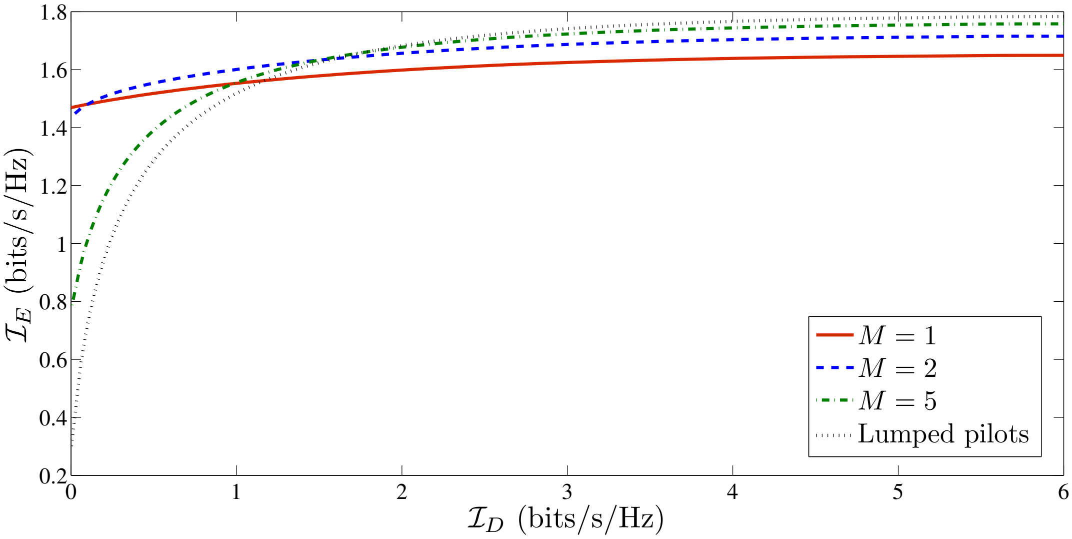

Figure 5.9: Capacity versus $\mathcal{I}_D$ under different values of M with $L = 2, N = 1000$, $l_s = 20$ and SNR=7 dB

are consistent with theory and can aid with system design.

Chapter 6

Secret Key Generation

6.1 Secret Key Generation In Vehicular Wireless Networks

Wireless channels are intrinsically prone to unauthorized access and confidentiality attacks. Securing the communication channel is vital in controlling the risks posed by malicious acts in mobile networking. A solution to ensure data security uses the characteristics of the communication channel to generate secret keys, known only to the legitimate communicating parties. The secret keys can then be used to encrypt the message. Using the channel to produce secret keys requires that the channel gains be estimated. The rate of the secret key generated in this way, i.e, the key capacity is determined by the estimation accuracy as well as the fading rate. The high accuracy methods proposed in the previous chapters can be employed to generate the secret keys at reasonable cost.

Upper bounds on the secret key capacity of estimated wireless channels is the subject of discussion in this chapter. These bounds have not been studied in prior art for realistic communication scenarios where the channel measurements at the two ends may not be simultaneous. In this research, we investigate upper bounds on the secret key capacity of doubly-selective channels assuming a practical IEEE 802.11 wireless network with full-duplex and half-duplex transmission modes. In the next section, the impact of channel estimation error on the performance of a key generation scheme is delineated.

6.1.1 Channel Estimation Error

The iterative approach to radio channel estimation exploits the detected data symbols to enhance the accuracy beyond the reach of purely pilot based systems. We proposed low complexity high accuracy channel estimators for fast-fading channels in iterative receivers in SISO and MIMO systems [112, 120]. The performance of the proposed estimator was close to that of a Wiener estimator, which is the optimal in the LMMSE sense. To start studying the theoretical aspects of key generation in iterative receivers, we consider the case where a Wiener filter is used to estimate the channel gains. The results will serve as a baseline to compare the performance of non-optimal estimators.

Wiener Estimator of Channel Gains

To take the full advantage of channel estimation in key generation, we assume that a Wiener filter is used by the receiver to estimate the channel gains. A Wiener filter is an LMMSE estimator, which for Gaussian RVs, is also the MMSE estimator. In this section, the error of an LMMSE estimator is calculated.

An $L+1$-tap Rayleigh fast-fading channel with a normalized Doppler frequency of f_D is considered. A block of N i.i.d. symbols, $\mathbf{s} = [s_0\ s_1\ \dots\ s_{N-1}]^T$ with zero mean and unit variance are input into the channel. The symbols are assumed to be known at the iterative receiver, after sufficient rounds of channel estimation and data detection. The corresponding channel gains are represented by $\mathbf{g} := [g(0;0)\ g(1;0)\ \dots g(N-1;0)\dots\ g(0;L)\ g(1;L)\ \dots g(N-1;L)]^T$. The output of the channel is contaminated with additive noise $\mathbf{v} = [v_0\ v_1 \dots v_{N-1}]^T$, with correlation $\mathbf{R}_{\mathbf{vv}} = \sigma_v^2 \mathbf{I}_N$.

The channel gains are expressed as

$$\mathbf{g} = \mathcal{B}\mathbf{h} \tag{6.1}$$

where $\mathcal{B} = \mathbf{I}_{L+1} \otimes \mathbf{E}$ for some matrix $\mathbf{E}$, is an $N(L+1)\times Q(L+1)$ non-singular transformation matrix and $\mathbf{h}$ is the $Q\times 1$ vector to be estimated. In case of CE-BEM, $\mathcal{B}$ is simply comprised of complex exponentials, with $Q \geq 2f_D N$. If $Q < N$, the CE-BEM representation is only an approximate one [129].

Define

$$\mathcal{S} = [\operatorname{diag}(\mathcal{D}_0[\mathbf{s}])\ \operatorname{diag}(\mathcal{D}_1[\mathbf{s}])\ \dots \operatorname{diag}(\mathcal{D}_L[\mathbf{s}])], \tag{6.2}$$

where $\mathcal{D}_l[\mathbf{s}]$ denotes an l-sample delayed version of vector $\mathbf{s}$, i.e., vector $\mathbf{s}$ prefixed with l zeros. The channel output is given by

$$\mathbf{y} = \mathbf{F}\mathbf{h} + \mathbf{v}, \tag{6.3}$$

where

$$\begin{aligned} \mathbf{F} &:= \mathcal{S}\mathcal{B} \\ &= \{F_{ll',ik} = s_{i-l'-1}E_{ik} \\ &\quad \text{for } l = 1,\ l' = 1,2,...,L+1,\ i = 1,2,...,N;\ k = 1,2,...,Q;\} \end{aligned}$$

in which E_{ik} refers to the elements of $\mathbf{E}$, and dummy symbols $s_{-1} = s_{-2} = ... = 0$ are used to simplify the formulation. The Wiener estimator for (6.3) is expressed in terms of cross-covariance and auto-covariance matrices as $\hat{\mathbf{h}} = \mathbf{C}_{\mathbf{yh}}^{H}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{y}$ [47,76]. Using (6.3) we obtain that

$$\hat{\mathbf{h}} = \mathbf{R}_{\mathbf{hh}}\mathbf{F}^{H}\left(\mathbf{F}\mathbf{R}_{\mathbf{hh}}\mathbf{F}^{H} + \sigma_v^2\mathbf{I}_N\right)^{-1}\mathbf{y} \tag{6.4}$$

The estimation error for $\hat{\mathbf{h}}$ is defined as $\varepsilon := \hat{\mathbf{h}} - \mathbf{h}$. Its covariance matrix is computed as

$$\mathbf{R}_{\varepsilon\varepsilon} = \mathbf{R}_{\mathbf{hh}} - \mathbf{R}_{\mathbf{hh}}\mathbf{F}^{H}\left(\mathbf{F}\mathbf{R}_{\mathbf{hh}}\mathbf{F}^{H} + \sigma_v^2\mathbf{I}_N\right)^{-1}\mathbf{F}\mathbf{R}_{\mathbf{hh}} \tag{6.5}$$

where we used identities $\mathbf{R}_{\mathbf{hy}} = \mathbf{R}_{\mathbf{yh}}^{H} = \mathbf{R}_{\mathbf{hh}}\mathbf{F}^{H}$ and $\mathbf{R}_{\mathbf{yy}} = \mathbf{F}\mathbf{R}_{\mathbf{hh}}\mathbf{F}^{H} + \sigma_v^2\mathbf{I}_N$. Using the Searle's identity for matrix inversion[1], after some manipulations, (6.5) can be written as

$$\begin{aligned} \mathbf{R}_{\varepsilon\varepsilon} &= \sigma_v^2\mathbf{R}_{\mathbf{hh}}\left(\mathbf{F}^{H}\mathbf{F}\mathbf{R}_{\mathbf{hh}} + \sigma_v^2\mathbf{I}_{(L+1)N}\right)^{-1} \\ &= \sigma_v^2\left(\mathbf{F}^{H}\mathbf{F} + \sigma_v^2\mathbf{R}_{\mathbf{hh}}^{-1}\right)^{-1} \\ &= \left(\frac{1}{\sigma_v^2}\mathbf{F}^{H}\mathbf{F} + \mathbf{R}_{\mathbf{hh}}^{-1}\right)^{-1} \end{aligned} \tag{6.6}$$

Least-squares Estimator

Assuming $s_i \neq 0$ for $i = 1,2,\ldots N$, $\mathbf{F}$ and $\mathbf{F}^{H}\mathbf{F}$ are full-rank and the LS estimator is given by

$$\hat{\mathbf{h}} = \left(\mathbf{F}^{H}\mathbf{F}\right)^{-1}\mathbf{F}^{H}\mathbf{y} \tag{6.7}$$

[1] $(\mathbf{AB} + \mathbf{I})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{BA} + \mathbf{I})^{-1}$

The correlation of the estimation error $\varepsilon \coloneqq \hat{\mathbf{h}} - \mathbf{h}$ is obtained as

$$\mathbf{R}_{\varepsilon\varepsilon} = \sigma_v^2(\mathbf{F}^H\mathbf{F})^{-1}. \tag{6.8}$$

An expression for the correlation of the gain estimation error is obtained as $\mathbf{R}_{ee} = \sigma_v^2\mathcal{B}(\mathbf{F}^H\mathbf{F})^{-1}\mathcal{B}^H$.

From (6.6), (5.18), and (6.8), when the CE-BEM coefficients are uncorrelated, the estimation errors are uncorrelated also. This important conclusion will be used to compute the key capacity of a multiple path channel.

Secret key generation from fast-fading channels has to consider the effect of time-division duplexing (TDD) on the correlation between the gains estimated by Alice and Bob. Also, the existing estimators need to be modified to be compatible with TDD. The design is then optimized based on the criteria given in the previous section. To put the problem into a context, consider an IEEE802.11p mobile system in which Alice and Bob estimate a fading Rayleigh channel. At a speed of 100 km/h, a carrier frequency of $f_c = 5.9$ GHz, a sampling rate of $f_s = 10$ MHz, a symbol duration of $T_s = 10^{-7}$ sec [139], the normalized Doppler frequency is $f_D = f_d T_s = v f_c T_s / c \approx 0.00005$, where c is the speed of light. In order for Alice and Bob to be able to obtain the same channel estimate sequence in a TDD transmission of period T_d, one needs to have $T_d < 1/2f_d = T_s/2f_D = 10^4 T_s$ according to the sampling theorem. Therefore, the frame length has to be < 10000/2 = 5000 symbols. The typical frame of IEEE 802.11p consists of about 5000 bits [58, p. 181] or $5000/\log M$ symbols, where M denotes the modulation order. An M-QAM modulation with $M = 4$ or higher will provide a frame length of less than 5000/2 = 2500 < 5000 symbols. For lower order modulation, the channel gains cannot be perfectly recovered from the samples, but there still exist a correlation between the Alice's and Bob's measurements. Here, a relationship between the key rate and the fading rate as well as the estimation accuracy may be sought. When the fading rate increases, the channel gain samples de-correlate faster over time. Therefore, the entropy of the samples per second is higher, which leads to higher key rates. At the same time however, the correlation between the Alice and Bob's observations is reduced assuming the frame rate is fixed. Moreover, the more accurate channel estimation corresponds to a higher correlation between the gain estimates obtained by Alice and Bob, resulting in more mutual information between Alice and Bob and, thus, more key capacity.

In the next section, we will study key generation in iterative receivers. Iterative estimation can significantly reduce the error, hence enable higher key rates, compared to the present techniques.

6.1.2 Secret Key Capacity of Iterative Receivers: Full-duplex Transmission

In this section, we give a closer look at the capabilities of a doubly-selective fast-fading multipath channel as a source of common randomness for legitimate users. We study the case where Alice and Bob estimate the same channel gains, $g(n)$, contaminated by their estimation errors $\varepsilon_A(n)$ and $\varepsilon_B(n)$, respectively, while an eavesdropper being kept sufficiently away from them, will experience an almost independent channel if the channel's propagation effects are random and unpredictable [106]. This independence condition only holds for the fast fading portion of the channel impulse response. Shadow fading and the deterministic path losses and portions of the channel response which can be predicted from ray-tracing are not suitable sources of key bits and must be excluded when secret key generation is performed. The case of non-simultaneous channel measurements will be studied in a following section. The estimation error for either party is assumed to be a sequence of Gaussian RVs of zero mean and variance σ_ε^2, with a PSD $S_{\varepsilon\varepsilon}(f) = \sigma_\varepsilon^2/(2f_D)$ over $[-f_D, f_D]$, where $f_D \in [0, 0.5]$ denotes the normalized Doppler frequency. The channel gain process follows the Jake's model where the PSD of $g(n)$ is $S_{gg}(f) = P/(\pi\sqrt{f_D^2 - f^2})$, $|f| < f_D$, with P representing the power of a channel tap.

Channel gain estimates are generally correlated over time. Secret key bits derived from correlated RVs would not be independent. Before the estimated channel gains are used to generate a key, a *whitening* procedure is in order, to produce a set of uncorrelated RVs.

The input to the whitening module of Fig. 6.1 is $q_A(n) = g(n) + \varepsilon_A(n)$ for Alice, and $q_B(n) = g(n) + \varepsilon_B(n)$ for Bob. In the frequency domain, the whitening process is equivalent to first, passing the channel estimates through a low-pass filter with a frequency response of $W(f) = 1/\sqrt{S_{gg}(f)/P} = \sqrt{\pi\sqrt{f_D^2 - f^2}}$, $|f| < f_D$, and then, downsampling with a rate of $1/2f_D$ in the time domain, as shown in Fig. 6.2. This procedure results in a flat power spectrum over all the frequency range for whitened $g(n)$, implying independence for jointly Gaussian RVs.

Whitening is parallel to the orthogonal decomposition of [173] and enables the use of the key generation method described therein. It is inspired by the method used

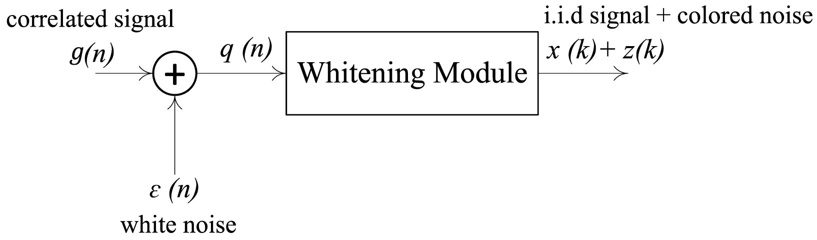

Figure 6.1: Signal model

by [175] to generate simulated Rayleigh channel fading sequence from i.i.d. RVs. It also makes key distillation easier as the key bits obtained from different samples would be independent.

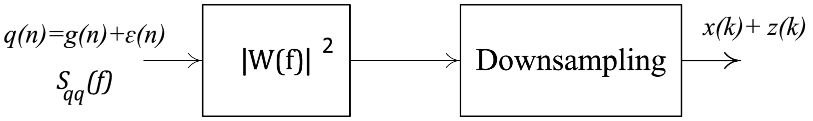

Figure 6.2: Equivalent whitening process

An upper bound on secret key rate for a single-tap channel is the mutual information $\mathcal{I}(q_A(n); q_B(n))$ between Alice's and Bob's measurements [5, 111]. For the case of jointly Gaussian RVs and stationary channel, the mutual information is given by [13]

$$\mathcal{I}(q_A; q_B) = -\int_{-f_D}^{f_D} \log_2 \left[1 - \frac{|S_{AB}(f)|^2}{S_{AA}(f) S_{BB}(f)}\right] \mathrm{d}f \tag{6.9}$$

where $S_{AA}(f)$ and $S_{BB}(f)$ denote the PSD of q_A and q_B, respectively, and $S_{AB}(f)$ represents the cross spectral density of q_A and q_B. Assuming that the estimation errors $\varepsilon_A(n)$ and $\varepsilon_B(n)$ for Alice and Bob are uncorrelated, $S_{AB}(f) = S_{gg}(f) = P/|W(f)|^2$. Inserting $S_{AB}(f)$ and $S_{AA}(f) = S_{BB}(f) = P/|W(f)|^2 + \sigma_\varepsilon^2/(2f_D)$ into (6.9), we have

$$\mathcal{I}(q_A; q_B) = -\int_{-f_D}^{f_D} \log_2 \left(1 - \left(\frac{P}{P + [\sigma_\varepsilon^2/(2f_D)]|W(f)|^2}\right)^2\right) \mathrm{d}f \tag{6.10}$$

For a channel with flat PSD, $|W(f)|^2 = 2f_D$ and (6.10) simplifies to

$$\mathcal{I}(q_A; q_B) \approx -2f_D \log_2 \left(1 - \left(\frac{P}{P + \sigma_\varepsilon^2}\right)^2\right) \tag{6.11}$$

Eq. (6.10) will be generalized to multipath channels in the next section.

Key Capacity of Multiple Path Channels

The key capacity of multiple path channel is simply the addition of the secret key generation capacity of each path, assuming that different channel paths are independent. To simplify the derivation, let the channel power spectrum have a uniform distribution over $[-f_D, f_D]$. Let $P_l, l = 0, \ldots, L$ denote the mean power of path l where $\sum_l P_l = 1$. Using (6.11), the key capacity (bits per second per Hertz) is obtained as

$$C_k = -\sum_{l=0}^{L} 2f_D \log_2 \left(1 - \left(\frac{P_l}{P_l + \sigma_{\varepsilon,l}^2}\right)^2\right) \tag{6.12}$$

with $\sigma_{\varepsilon,l}^2$ denoting the variance of estimation error for path l. When the PSD is uniform, an intuitive closed form expression for the key capacity can be obtained. In this case $\mathbf{R}_{hh} = (2f_D(L+1))^{-1}\mathbf{I}_{Q(L+1)}$. Assuming equi-power channel taps, using (6.6) and noting that $\mathbf{F}^H\mathbf{F} \approx \mathbf{I}$, one may write

$$\mathbf{R}_{\varepsilon\varepsilon} = \left(\frac{1}{\sigma_v^2} + 2f_D(L+1)\right)^{-1} \mathbf{I}_{Q(L+1)} \tag{6.13}$$

In addition, the estimation errors for different paths are identical, i.e.,

$$\sigma_{\varepsilon,0}^2 = \cdots = \sigma_{\varepsilon,L}^2 = \sigma_{\varepsilon}^2 = \frac{Q}{N}\left(\frac{1}{\sigma_v^2} + 2f_D(L+1)\right)^{-1} = 2f_D\left(\frac{1}{\sigma_v^2} + 2f_D(L+1)\right)^{-1} \tag{6.14}$$

Inserting (6.14) into (6.12) and noting that $2f_D(L+1)\sigma_v^2 << 1$ for the cases of interest here (when $f_D \le 0.01$), we have that

$$C_k \approx -2(L+1)f_D \log_2\left(4f_D(L+1)\sigma_v^2\right) \tag{6.15}$$

When f_D is fixed, C_k approximates a linear function of SNR (dB), the slope of which is determined by f_D.

6.1.3 Secret Key Capacity of Iterative Receivers: Time-division Duplexing

Consider a time-division-duplexing (TDD) radio link, through which Alice and Bob take turns sending subblocks of length N symbols to each other. Further, suppose that K subblocks are transmitted by each party. For simplicity of explanation, let

$K = 1$. The results can easily be extended to $K > 1$. We only consider a single tap channel here. The results are easily extended to the multipath case by adding the key capacities of different channel paths when they are assumed to be independent. Define $\mathbf{y}_A := [y_A(1)...y_A(N)]^T$ and $\mathbf{y}_B := [y_B(N+1)...y_B(2N)]^T$ as the received signals by Alice and Bob, respectively. We have

$$\mathbf{y}_A = \mathcal{S}_A \mathbf{g}_A + \mathbf{v}_A \tag{6.16}$$

$$\mathbf{y}_B = \mathcal{S}_B \mathbf{g}_B + \mathbf{v}_B \tag{6.17}$$

where $\mathbf{g}_A := [g(1)...g(N)]^T$ and $\mathbf{g}_B := [g(N+1)...g(2N)]^T$ denote the (common) channel gains as seen by Alice and Bob respectively. Matrix $\mathcal{S}_A$ represents a diagonal $N \times N$ matrix with the symbols sent by Bob on diagonal. For Alice, vector $\mathbf{v}_A$ is the Gaussian i.i.d. noise, independent from $\mathbf{v}_B$. Define the covariance matrices $\mathbf{R}^g_{AA} := E[\mathbf{g}_A \mathbf{g}_A^H] = \mathbf{R}^g_{BB} := E[\mathbf{g}_B \mathbf{g}_B^H]$ and $\mathbf{R}^g_{AB} := E[\mathbf{g}_A \mathbf{g}_B^H]$.

To avoid ill-conditioned matrices, the singular value decomposition (SVD) is used as $\mathbf{R}^g_{AB} = \mathbf{U_A S U_B}$ where $\mathbf{U_A}$ and $\mathbf{U_B}$ are unitary matrices and $\mathbf{S}$ is a diagonal matrix. Define $\mathbf{E}_A := \mathbf{U_A}(:, \mathbf{1}:\mathbf{Q})$ and $\mathbf{E}_B := \mathbf{U_B}(:, \mathbf{1}:\mathbf{Q})$ where $Q := \lceil 2 f_D N \rceil + 1$. Let $\mathbf{h}_A := \mathbf{E}_A^H \mathbf{g}_A$, $\mathbf{h}_B := \mathbf{E}_B^H \mathbf{g}_B$. So,

$$\mathbf{R}^h_{AA} \;:=\; E[\mathbf{h}_A \mathbf{h}_A^H] = \mathbf{E}_A^H \mathbf{R}^g_{AA} \mathbf{E}_A \tag{6.18}$$

$$\mathbf{R}^h_{BB} \;:=\; E[\mathbf{h}_B \mathbf{h}_B^H] = \mathbf{E}_B^H \mathbf{R}^g_{BB} \mathbf{E}_B \tag{6.19}$$

$$\mathbf{R}^h_{AB} \;:=\; E[\mathbf{h}_A \mathbf{h}_B^H] = \mathbf{E}_A^H \mathbf{R}^g_{AB} \mathbf{E}_B \tag{6.20}$$

Assuming identical noise variances of σ_v^2 for both users, and defining $\mathbf{F}_A := \mathcal{S}_A \mathbf{E}$, one may write the LMMSE estimator of $\mathbf{h}_A$ as

$$\hat{\mathbf{h}}_A = \left(\mathbf{I}_Q + \sigma_v^2 {\mathbf{R}^h_{AA}}^{-1}\right)^{-1} \mathbf{F}_A^H \mathbf{y}_A \tag{6.21}$$

$$= \left(\mathbf{I}_Q + \sigma_v^2 {\mathbf{R}^h_{AA}}^{-1}\right)^{-1} \mathbf{F}_A^H \left(\mathbf{F}_A \mathbf{h}_A + \mathbf{v}_A\right) \tag{6.22}$$

$$= \left(\mathbf{I}_Q + \sigma_v^2 {\mathbf{R}^h_{AA}}^{-1}\right)^{-1} \left(\mathbf{h}_A + \mathbf{F}_A^H \mathbf{v}_A\right) \tag{6.23}$$

where (6.22) follows from $\mathbf{F}_A^H \mathbf{F}_A \approx \mathbf{I}_Q$ (in Lemma 4 set $\sigma_u^2 = 0$ so that $\beta = 1$ and $\mathcal{G} = \mathbf{I}$). Similarly,

$$\hat{\mathbf{h}}_B = \left(\mathbf{I}_Q + \sigma_v^2 {\mathbf{R}^h_{BB}}^{-1}\right)^{-1} \left(\mathbf{h}_B + \mathbf{F}_B^H \mathbf{v}_B\right) \tag{6.24}$$

The secret key capacity per channel symbol may be expressed in terms of the mutual information between $\hat{\mathbf{h}}_A$ and $\hat{\mathbf{h}}_B$ as [134]

$$C_k = \frac{1}{2N}\mathcal{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B) = \frac{1}{2N}\log_2\frac{\det\{\mathbf{R}^{\hat{h}}_{AA}\}\det\{\mathbf{R}^{\hat{h}}_{BB}\}}{\det\{\mathbf{R}^{\hat{h}}_{AB,AB}\}} \tag{6.25}$$

where

$$\mathbf{R}^{\hat{h}}_{AA} := E[\hat{\mathbf{h}}_A\hat{\mathbf{h}}_A^H] = \mathbf{M}_A\mathbf{R}^{h}_{AA}\mathbf{M}_A^H + \sigma_v^2\mathbf{M}_A\mathbf{M}_A^H \tag{6.26}$$

$$\mathbf{R}^{\hat{h}}_{BB} := E[\hat{\mathbf{h}}_B\hat{\mathbf{h}}_B^H] = \mathbf{M}_B\mathbf{R}^{h}_{BB}\mathbf{M}_B^H + \sigma_v^2\mathbf{M}_B\mathbf{M}_B^H \tag{6.27}$$

$$\mathbf{R}^{\hat{h}}_{AB} := E[\hat{\mathbf{h}}_A\hat{\mathbf{h}}_B^H] = \mathbf{M}_A\mathbf{R}^{h}_{AB}\mathbf{M}_B^H \tag{6.28}$$

$$\mathbf{M}_A := \left(\mathbf{I}_Q + \sigma_v^2{\mathbf{R}^{h}_{AA}}^{-1}\right)^{-1} \tag{6.29}$$

$$\mathbf{M}_B := \left(\mathbf{I}_Q + \sigma_v^2{\mathbf{R}^{h}_{BB}}^{-1}\right)^{-1} \tag{6.30}$$

$$\mathbf{R}^{\hat{h}}_{AB,AB} := \begin{bmatrix} \mathbf{R}^{\hat{h}}_{AA} & \mathbf{R}^{\hat{h}}_{AB} \\ (\mathbf{R}^{\hat{h}}_{AB})^H & \mathbf{R}^{\hat{h}}_{BB} \end{bmatrix} \tag{6.31}$$

6.1.4 Analysis of Long Measurement Blocks

In practice, transmission over continuous-fading channels encompasses many symbol blocks. Unfortunately, calculation of the secret key capacity using the method described in the previous section involves matrices of large dimensions, handling of which is a burden. In this section, the special structure of the covariance matrices is exploited to extend the results of the previous section to the case where Alice and Bob perform TDD transmission over a long period of time, i.e., $K \to \infty$. To simplify the analysis, let $\mathcal{S}_A = \mathcal{S}_B = \mathbf{I}$ in (6.16). A sequence of K measurements by each party is considered. Define $\mathbf{Y}_{AB} := [(\mathbf{y}_A^1)^T\ (\mathbf{y}_B^1)^T \cdots (\mathbf{y}_A^K)^T\ (\mathbf{y}_B^K)^T]^T$, $\mathbf{Y}_A := [(\mathbf{y}_A^1)^T \cdots (\mathbf{y}_A^K)^T]^T$ and $\mathbf{Y}_B := [(\mathbf{y}_B^1)^T \cdots (\mathbf{y}_B^K)^T]^T$, where $\mathbf{y}_A^k := [y_A((k-1)N_C+1) \cdots y_A((k-1)N_C+N)]^T$, $\mathbf{y}_B^k := [y_B((k-1)N_C+N+1) \cdots y_B(kN_C)]^T$ and $N_C := 2N$.

Note that $y_A(n) = \hat{g}_A(n) = g_A(n) + v_A(n)$ and $y_B(n) = \hat{g}_B(n) = g_B(n) + v_B(n)$. Using a BEM representation, $\mathbf{y}_A^k$ and $\mathbf{y}_B^k$ can be replaced with $\mathbf{x}_A^k = \mathbf{E}^H\mathbf{y}_A^k$ and $\mathbf{x}_B^k = \mathbf{E}^H\mathbf{y}_B^k$. Now define $\mathbf{X}_{AB} := [\mathbf{x}_A^1\ \mathbf{x}_B^1 \cdots \mathbf{x}_A^K\ \mathbf{x}_B^K]^T$, $\mathbf{X}_A := [\mathbf{x}_A^1 \cdots \mathbf{x}_A^K]^T$ and $\mathbf{X}_B := [\mathbf{x}_B^1 \cdots \mathbf{x}_B^K]^T$.

The secret key capacity for infinitely long TDD transmission is defined as

$$C_k^\infty := \lim_{K\to\infty}\frac{1}{KN_C}\mathcal{I}(\mathbf{X}_A;\mathbf{X}_B) = \lim_{K\to\infty}\frac{1}{KN_C}\log_2\frac{\det\{\mathbf{R}^X_{AA}\}\det\{\mathbf{R}^X_{BB}\}}{\det\{\mathbf{R}^X_{AB,AB}\}} \tag{6.32}$$

where the covariance matrices are defined similar to (6.26). Since the channel gain process and the noise process are wide sense stationary, these matrices are block Toeplitz given as below.

$$\mathbf{R}_{AA}^{X} := \begin{bmatrix} A_0 & A_1 & \dots & A_{K-1} \\ A_1^H & A_0 & \dots & A_{K-2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{K-1}^H & A_{K-2}^H & \dots & A_0 \end{bmatrix} \tag{6.33}$$

$\mathbf{R}_{BB}^{X}$ is similarly defined.

$$\mathbf{R}_{AB,AB}^{X} := \begin{bmatrix} A_0 & C_0 & \dots & A_{K-1} & C_{K-1} \\ C_0^H & A_0 & \dots & C_{K-2} & A_{K-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{K-1}^H & A_{K-1}^H & \dots & C_0^H & A_0 \end{bmatrix} \tag{6.34}$$

where

$$\begin{aligned} A_i = B_i \quad &:= \quad \mathrm{E}[\mathbf{x}_A^k(\mathbf{x}_A^{k+i})^H] = \mathbf{E}^H\mathbf{J}_i^A\mathbf{E} + \sigma_v^2\delta_i\mathbf{I} && (6.35) \\ C_i \quad &:= \quad \mathrm{E}[\mathbf{x}_A^k(\mathbf{x}_B^{k+i})^H] = \mathbf{E}^H\mathbf{J}_i^C\mathbf{E} && (6.36) \end{aligned}$$

where δ_i denotes the Kronecker delta function, $(\mathbf{J}_i^A)_{m,n} := \mathcal{J}_0(2\pi f_D|N_C i + n - m|)$, $(\mathbf{J}_i^C)_{m,n} := \mathcal{J}_0(2\pi f_D|N_C i + N + n - m|)$, and $A_{-k} = A_k^H$.

Algorithm 5 is used to compute the logarithm of the determinants in (6.32). This algorithm is based on the Whittle recursion for the inversion of large Block-Toeplitz matrices [108, 146]. To compute $\det\{\mathbf{R}_{AA}^X\}$ and $\det\{\mathbf{R}_{BB}^X\}$, the matrices A_i and B_i are fed to the algorithm for $i = 0, \cdots, K-1$. To compute $\det\{\mathbf{R}_{AB,AB}^X\}$, the matrices $A_0, C_0, A_1, C_1, \cdots, A_{K-1}, C_{K-1}$ are input into the algorithm. To calculate C_k^∞, the limit as $K \to \infty$ is needed. Fortunately, the prediction error covariances $\tilde{P}$ approach a steady-state value as $K \to \infty$ for the above inputs. Therefore, one only needs to continue the computations up to the value of K where $\tilde{P}$ converges to its asymptotic value.

Algorithm 5 Block-Toeplitz Matrix Determinant Calculation

Inputs:

- Block matrices A_k for $k = 0, \cdots, K-1$

Output:

- Logarithm of determinant of matrix: LogDet

Working variables:

- Prediction error covariances: P and $\tilde{P}$
- Prediction matrix coefficients: D_m and $\tilde{D}_m$ for $m = 1, \cdots, K-1$
- Prediction matrix coefficients for previous iteration: C_m and $\tilde{C}_m$ for $m = 1, \cdots, K-1$

```
LogDet $\leftarrow 0$
$P \leftarrow A_0$
$\tilde{P} \leftarrow A_0$
for $k = 1, 2, \ldots, K-1$ do
    $\Delta \leftarrow A_k + \sum_{m=1}^{k-1} C_m A_{k-m}$
    $D_k \leftarrow -\Delta \tilde{P}^{-1}$
    $\tilde{D}_k \leftarrow -\Delta^H P^{-1}$
    for $m = 1, 2, \ldots, k-1$ do
        $D_m \leftarrow C_m + D_k \tilde{C}_{k-m}$
        $\tilde{D}_m \leftarrow \tilde{C}_m + \tilde{D}_k C_{k-m}$
    end for
    $C_m \leftarrow D_m$ for $m = 1, \cdots, k$
    $\tilde{C}_m \leftarrow \tilde{D}_m$ for $m = 1, \cdots, k$
    $P \leftarrow P + C_k \Delta^H$
    $\tilde{P} \leftarrow \tilde{P} + \tilde{C}_k \Delta$
    LogDet $\leftarrow + \log_2 |\tilde{P}|$
end for
```

6.2 Simulations

First, the NMSE of a typical channel estimator versus SNR is shown in Fig. 6.3 and compared with the optimal-MMSE (Wiener) estimator. In this figure, the line labeled "KF Estimator" refers to a Kalman-filter based estimator followed by a low-pass filter as described in [120]. It can be seen that practical system can indeed approach the MMSE-optimal performance. This suggests that accurate estimators can be employed for higher secret key rates.

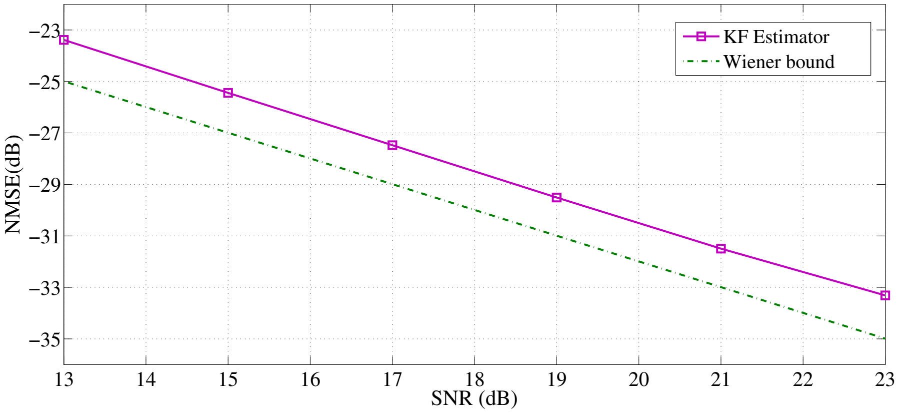

Figure 6.3: NMSE versus SNR for a Kalman filter compared to a Wiener filter .

The secret key rate for three different wireless standard scenarios, as shown in Table 6.1, is computed from (6.12) and compared. Given the signal bandwidth B_w, the symbol duration calculates as $T_s = 1/B_w$. As a rule of thumb, the number of resolvable paths can be approximated as [89],

$$L + 1 \approx \lfloor 2T_{D,rms}B_w \rfloor + 1, \tag{6.37}$$

where $T_{D,rms}$ denotes the root-mean-squared (RMS) delay spread. The channel gains have unit power ($\sum_l P_l = 1$). The power profile is uniform over paths, unless otherwise noted. The PSD is uniformly distributed over normalized range of $[-f_D, f_D]$. The vehicle's velocity is 100 km/h or $v \approx 28$m/s unless otherwise stated. The normalized Doppler frequency is $f_D = vf_c/c$ as noted earlier.

Table 6.1: Wireless Standard Parameters

Standard	Carrier frequency (f_c)	Bandwidth (B_w)	Symbol rate (SR)
UWB (IEEE 802.15.4a) [8, 116]	4GHz	0.5GHz	112Msps
IEEE 802.11b	2.4GHz	20MHz	1.375Msps
IEEE 802.11p	5.9GHz	10MHz	10Msps

The secret key rate in bits per second is plotted in Fig. 6.4 for different wireless standards. To generate this graph, the secret key capacity in (6.15) is multiplied by the bandwidth B_w to give the key rate in bits per second. The number of propagation paths for each standard is obtained from (6.37) for $T_{D,rms}$ = 100ns, typical of an urban microcellular radio channel [11]. According to (6.15) the slope of the curves is determined by f_D. In the “UWB” case, despite the low fading rate, the key rate is significant due to its high bandwidth and the large number of propagation paths.

The effect of channel diversity on the key rate is illustrated in Fig. 6.5 for the IEEE 802.11p standard. It can be seen that channel diversity has a significant contribution to the key capacity.

The effect of channel diversity on the key rate when a single path dominates the rest is illustrated in Fig. 6.6, where several scenarios for an IEEE 802.11p network with $L + 1 = 10$ paths are compared. The case of a single-tap channel is also shown for comparison. For the non-uniform power profiles, there is one dominant path and nine non-dominant path of a total power of either −10 dB, −20 dB or −30 dB. As far as the channel capacity is concerned, there is no point in estimating non-dominant paths. From a security point of view however, the difference made by estimating the non-dominant paths in terms of key capacity is significant if the non-dominant paths are not much weaker.

Fig. 6.7 shows the key capacity (bits per symbol) versus SNR for full-duplex transmission using (6.12) as well as TDD with different block sizes N when $NK = 4000$ symbols and $f_D = 0.01$ based on (6.25). There exists a 3-dB penalty due to TDD at $N = 1$ which is expected because the TDD in this case comes down to discarding half the samples, leading to the estimation error being doubled. This fact can be deduced from (6.14) noting that downsampling is equivalent to doubling f_D and, thus, the estimation error. Moreover, it is seen that as we start to go down the Nyquist rate, the key capacity notably declines.

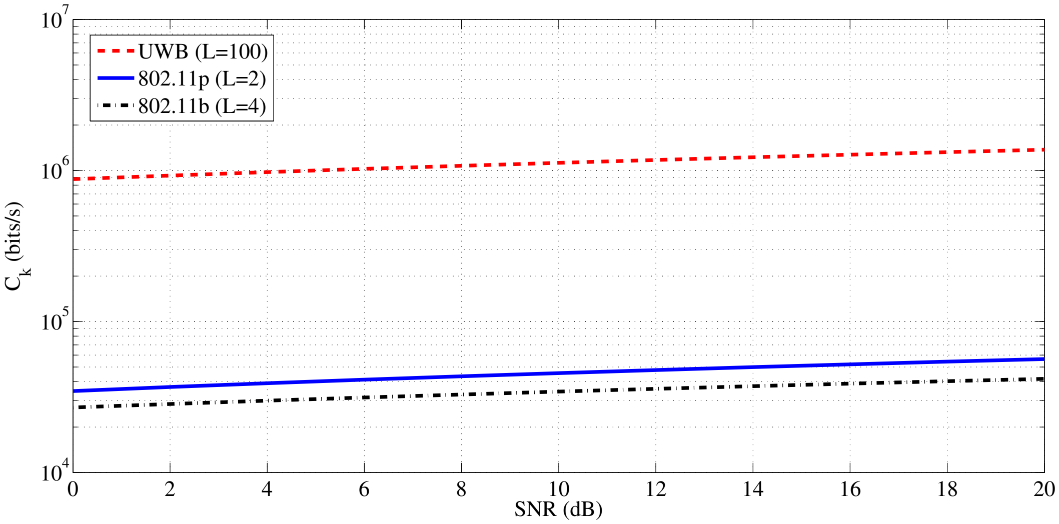

Figure 6.4: Key rate in bits per second versus SNR for different wireless standards.

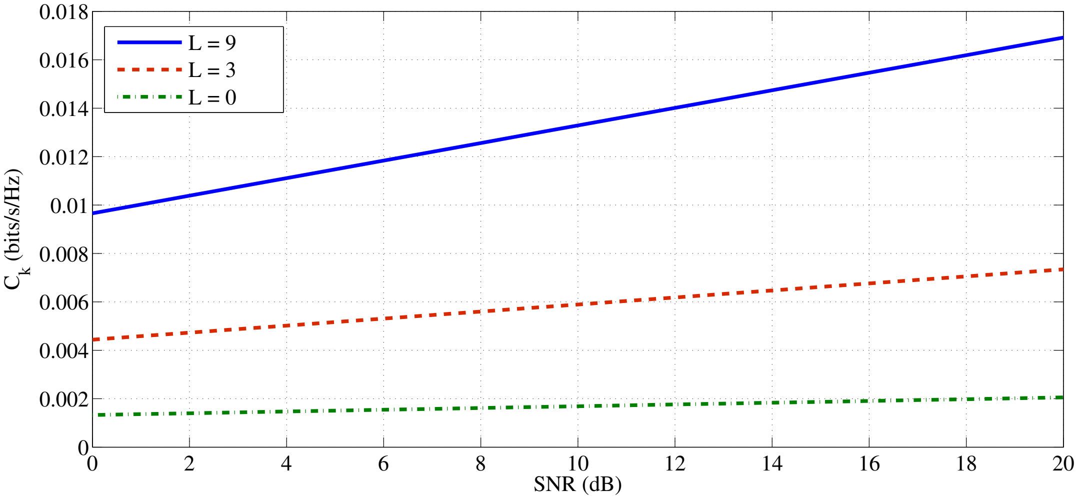

Figure 6.5: Key capacity of IEEE 802.11p versus SNR for different number of paths.

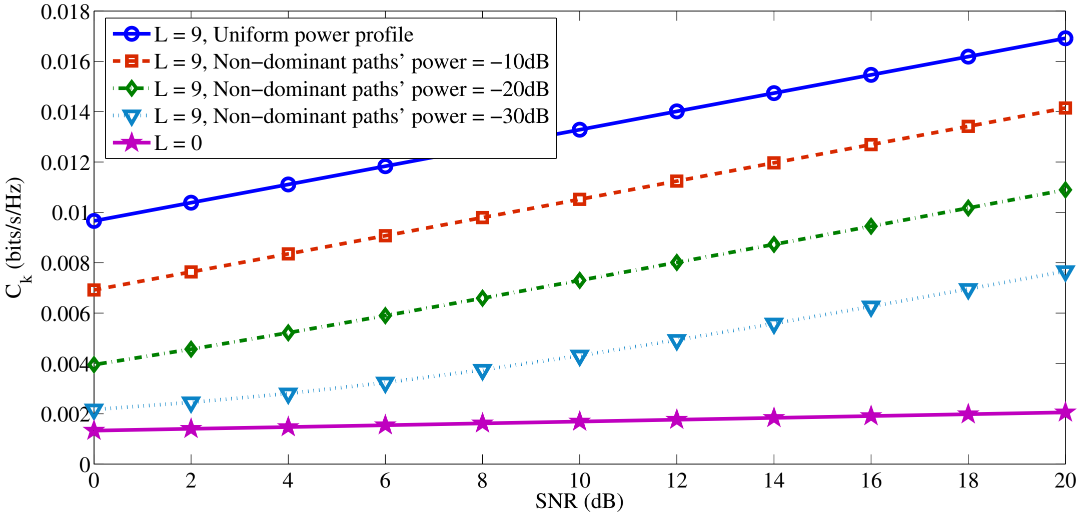

Figure 6.6: Key capacity of IEEE 802.11p versus SNR for a 10-tap channel under different delay-power profile.

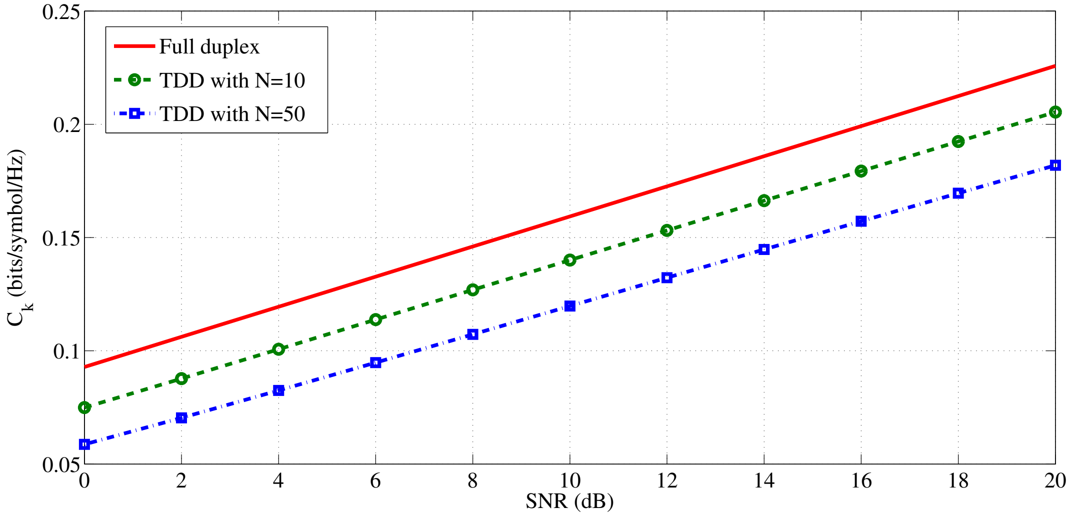

Figure 6.7: Key capacity versus SNR under full-duplex and time-division duplexing at $f_D = 0.01$.

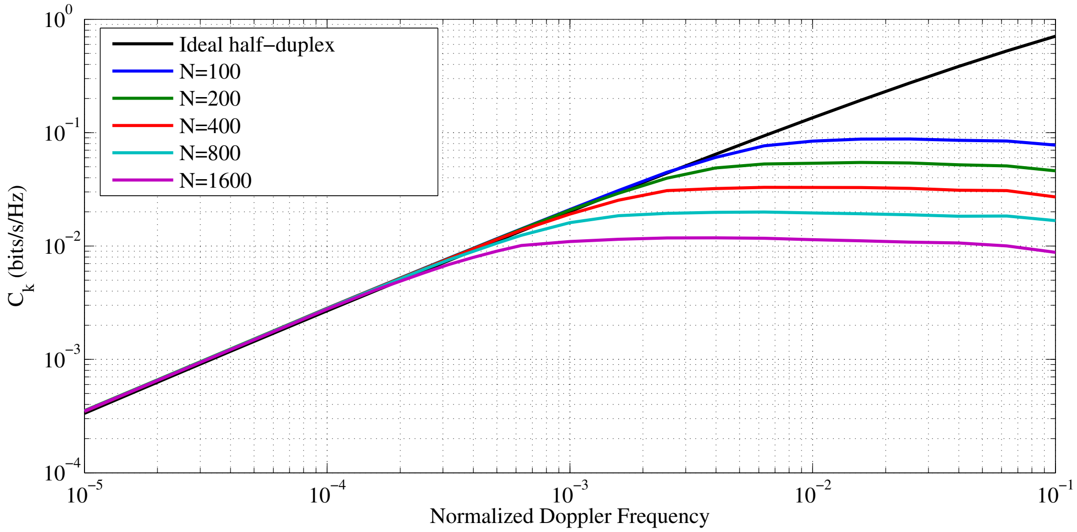

Figure 6.8: Key capacity versus f_D under full-duplex and time-division duplexing at SNR= 10dB.

In Fig. 6.8, the key capacity (bits per symbol) versus normalized Doppler frequency f_D for various block sizes is depicted when SNR= 10 dB. Algorithm 5 is used to compute the key capacity for TDD transmission based on (6.32) when the channel follows the Rayleigh fading model. The ideal half-duplex case corresponds to $N = 1$. To explain the shape of the capacity curves, let us consider the case when $N = 400$. At this N, the maximum normalized Doppler frequency less than which the channel gains can be fully recovered from samples, is $1/2N_C = 1/4N = 1/(4 \times 400) = 6.25 \times 10^{-4}$ by sampling theorem. Therefore, one expects that the two curves grow at the same rate. Then there is a turning point beyond which the performance of TDD transmission rapidly deteriorates compared to the ideal half-duplex case, as the correlation between the samples belonging to different blocks drops off. Yet, the figure shows that even at higher speeds of the vehicle or large transmission blocks when $f_D > 1/4N$, significant secret key rates can be generated.

6.3 Summary

We studied secret key generation from the physical properties of the wireless channel. A realistic doubly selective channel scenario with half-duplex transmission was considered and upper bounds on the secret key rate were calculated. It was shown that even when the channel is sampled under the Nyquist rate, that is, when the transmission intervals are too long to allow for a full reconstruction of the channel impulse response, a significant key capacity is still available.

Chapter 7

Conclusions and Future Work

7.1 Research Results

In this dissertation, methods for iterative estimation of doubly selective radio channels are proposed. The capacity gain of iterative channel estimation is studied. Upper bounds on secret key rate generated from channel estimates are derived.

To enhance the performance of the channel estimator, a smoother is employed to reduce the estimation error at the output of the Kalman filter, without requiring a long memory Kalman filter, hence saving the computational complexity. The computational cost can be reduced further by using higher order AR models instead of CE-BEM. The results are compared to the previous art to demonstrate the advantages of the proposed techniques in terms of BER, convergence speed and cost. Convergence analysis using EXIT charts demonstrates the fast convergence of the proposed methods to low BER states. We show that convergence to a low BER state is achieved after only few iterations.

In order to save the bandwidth, an accurate and efficient approach to semiblind estimation of MIMO-OFDM channels based on KLT-BEM is proposed. A block processing technique is employed to use the channel estimates of the current BEM block to project the channel gains over the next block. Unlike the precoding-based methods, the proposed scheme can be used with more accurate non-linear equalizers such as sphere decoders without inflicting unacceptable computational cost. The proposed method's performance compares favorably with existing iterative pilot-aided systems and competes with existing semiblind and blind estimation techniques. The performance of our method is as close as 0.3 dB to the perfect CSI case for the proposed method and 1 dB for the other methods. Compared to the previous art, the proposed method is shown to excel in the MTBF, especially in higher order modulations.

To study the effectiveness of iterative processing, the capacity of iteratively estimated radio channels is investigated. It is demonstrated that how the knowledge of the capacity gain from iterative detection versus purely pilot-based channel estimation helps a designer to compare the performance of an iterative receiver against a non-iterative one and select the best balance between performance and cost. By taking the uncertainty in decoded data bits into account, the channel estimation LMMSE of an iterative receiver with a given pilot ratio is obtained. The LMMSE is then used to derive a bound on capacity. The simulations results are consistent with theory and can aid with system design. The interaction between the symbol detector and the decoder is characterized in an EXIT chart. With optimal LMMSE pilot-based channel estimation, the results of this research reveal that iterative channel estimation provides insignificant capacity advantage at fading rates below 1% of the symbol rate, though a computational-cost gain is still available. Iterative channel estimation provides a capacity benefit if sub-optimal pilots are used to provide initial channel estimates.

In the last part of the research, we study the problem of secret key generation from the channel gains. We consider a realistic doubly selective channel scenario based on IEEE 802.11p standard, where half-duplex transmission is allowed and then calculated upper bounds on the secret key rate. It is shown that even when the transmission intervals in the half-duplex transmission mode are too long to allow for a full reconstruction of the channel impulse response, a significant key capacity is still available.

The main contributions of the dissertation are listed as follows.

- Introducing Low-complexity and accurate channel estimation algorithms for iterative receivers
- Investigating the capacity of iteratively estimated channels
- Proposing a semiblind channel estimation technique for MIMO-OFDM
- Calculating bounds on secret key capacity in realistic scenarios

7.2 Future Work

The error correction coding considered in this dissertation included conventional LDPC or convolutional codes. Future work may consider also multilevel coding [165] and analog codes [143].

The idea of using a low complexity zero phase filter designed with the method introduced in Chapter 3 seems to have a good potential to be used in other estimation algorithms. For instance, a smoother can be employed to enhance the performance of the MIMO channel estimator of Chapter 4 and reduce the cost. For this to work, the block processing scheme of Chapter 3 needs to be modified.

The capacity advantage of an iterative receiver over a non-iterative channel estimator was evaluated. By taking the uncertainty in decoded data bits into account, the channel estimation LMMSE of an iterative receiver with a given pilot ratio was obtained. The LMMSE was then used to derive a bound on capacity. The simulations results are consistent with theory and can aid with system design. Although this study considers single-input single-output systems, the approach can be extended to MIMO systems in a future work. Future work will also consider general distribution of power and correlations of the channel tap gains. The calculation of capacity bounds when different error correction codes are employed at different levels of modulation will also be the subject of future research.

Calculations of the upper bounds on the secret key rate optimistically assumed that the eavesdropper is completely ignorant about the communication channel. In practice, there may exist cases where Eve is located close to either party, able to measure the same channel as Alice and Bob do, and capable of performing iterative channel estimation and decoding. Future work may use a technique to hinder Eve from estimating the channel by sending “random” pilot symbols. The pilot pattern is known to Alice and Bob, but unknown to Eve. As such, she will not be able to initialize iterative processing and is bound to use blind techniques, which are not as accurate. This gives Alice and Bob some advantage in deriving a secret key. Privacy amplification is then used to obtain completely secret key bits.

Bibliography

[1] Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017. Technical report, Cisco Systems Inc., 2012.

[2] T. Abe and T. MATSUMOTO. Space-time turbo equalization in frequency-selective MIMO channels. *Vehicular Technology, IEEE Transactions on*, 52(3):469–475, May 2003.

[3] S. Adireddy, Lang Tong, and H. Viswanathan. Optimal placement of training for frequency-selective block-fading channels. *Information Theory, IEEE Transactions on*, 48(8):2338–2353, Aug. 2002.

[4] M. Agarwal and M.L. Honig. Wideband fading channel capacity with training and partial feedback. *Information Theory, IEEE Transactions on*, 56(10):4865–4873, Oct. 2010.

[5] R. Ahlswede and I. Csiszar. Common randomness in information theory and cryptography. I. Secret sharing. *Information Theory, IEEE Transactions on*, 39(4):1121 –1132, July 1993.

[6] S. Ahmed, T. Ratnarajah, M. Sellathurai, and C. Cowan. EXIT chart analysis of a reduced complexity iterative MIMO-OFDM receiver. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 2430–2434, Apr. 2007.

[7] S. Ahmed, T. Ratnarajah, M. Sellathurai, and C. Cowan. Iterative receivers for MIMO-OFDM and their convergence behavior. *Vehicular Technology, IEEE Transactions on*, 58(1):461–468, Jan. 2009.

[8] G. R. Aiello and G. D. Rogerson. Ultra-Wideband wireless systems. *Microwave Magazine, IEEE*, 4(2):36–47, 2003.

[9] S. Alamouti. A simple transmit diversity technique for wireless communications. *Selected Areas in Communications, IEEE Journal on*, 16(8):1451–1458, Oct. 1998.

[10] A. Antoniou. *Digital Signal Processing: Signals, Systems, and Filters.* McGraw-Hill, Toronto, Ontario, 2006.

[11] A.A. Arowojolu, A.M.D. Turkmani, and J.D. Parsons. Time dispersion measurements in urban microcellular environments. In *Vehicular Technology Conference, 1994 IEEE 44th*, pages 150 –154 vol.1, June 1994.

[12] A. Ashikhmin, G. Kramer, and S. Ten Brink. Extrinsic information transfer functions: model and erasure channel properties. *Information Theory, IEEE Transactions on*, 50(11):2657–2673, Nov. 2004.

[13] T. Avgeris, E. Lithopoulos, and N. Tzannes. Application of the mutual information principle to spectral density estimation. *Information Theory, IEEE Transactions on*, 26(2):184 – 188, Mar. 1980.

[14] Babak Azimi-Sadjadi, Aggelos Kiayias, Alejandra Mercado, and Bulent Yener. Robust key generation from signal envelopes in wireless networks. In *Proceedings of the 14th ACM Conference on Computer and Communications Security*, CCS '07, pages 401–410, New York, NY, USA, 2007. ACM.

[15] K. E. Baddour and N. C. Beaulieu. Autoregressive modeling for fading channel simulation. *Wireless Communications, IEEE Transactions on*, 4(4):1650–1662, July 2005.

[16] J. Baltersee, G. Fock, and H. Meyr. An information theoretic foundation of synchronized detection. *Communications, IEEE Transactions on*, 49(12):2115–2123, Dec. 2001.

[17] J. Barros and M.R.D. Rodrigues. Secrecy capacity of wireless channels. In *Information Theory, 2006 IEEE International Symposium on*, pages 356 –360, July 2006.

[18] C.H. Bennett, G. Brassard, C. Crepeau, and U. M. Maurer. Generalized privacy amplification. *Information Theory, IEEE Transactions on*, 41(6):1915 –1923, Nov. 1995.

[19] Charles Bennett, Franois Bessette, Gilles Brassard, Louis Salvail, and John Smolin. Experimental quantum cryptography. *Journal of Cryptology*, 5(1):3–28, 1992.

[20] D.S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)*. Princeton reference. Princeton University Press, 2009.

[21] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1. In *Communications, 1993. ICC 93. Geneva. Technical Program, Conference Record, IEEE International Conference on*, volume 2, pages 1064–1070 vol.2, 1993.

[22] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: information-theoretic and communications aspects. *Information Theory, IEEE Transactions on*, 44(6):2619 –2692, Oct. 1998.

[23] R.E. Blahut. *Principles and Practice of Information Theory*. Addison-Wesley Series in Electrical & Computer Engineering. Addison-Wesley, 1987.

[24] M. Bloch and J. Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[25] H. Bolcskei and A.J. Paulraj. Space-frequency coded broadband OFDM systems. In *Wireless Communications and Networking Confernce, 2000. WCNC. 2000 IEEE*, volume 1, pages 1–6 vol.1, 2000.

[26] F. Brannstrom, L.K. Rasmussen, and A.J. Grant. Convergence analysis and optimal scheduling for multiple concatenated codes. *Information Theory, IEEE Transactions on*, 51(9):3354–3364, Sep. 2005.

[27] G. Brassard and L. Salvail. Secret key reconciliation by public discussion. In *Advances in Cryptology - Proc. Eurocrypt'94*, pages 410–423, 1994.

[28] W.R. Braun and U. Dersch. A physical mobile radio channel model. *Vehicular Technology, IEEE Transactions on*, 40(2):472 –482, May 1991.

[29] S. Ten Brink. Convergence behavior of iteratively decoded parallel concatenated codes. *Communications, IEEE Transactions on*, 49(10):1727–1737, 2001.

[30] J.V. Candy. *Model-Based Signal Processing.* Adaptive and Learning Systems for Signal Processing, Communications and Control Series. John Wiley & Sons, 2005.

[31] J.K. Cavers. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *Vehicular Technology, IEEE Transactions on*, 40(4):686–693, Nov. 1991.

[32] J.K. Cavers. Pilot symbol assisted modulation and differential detection in fading and delay spread. *Communications, IEEE Transactions on*, 43(7):2206–2212, July 1995.

[33] Yi-Sheng Chen. Semiblind channel estimation for MIMO single carrier with frequency-domain equalization systems. *Vehicular Technology, IEEE Transactions on*, 59(1):53–62, 2010.

[34] Yi-Sheng Chen and Jyu-Han Song. A universal decoding algorithm for lattice codes. *Colloq. GRETSI*, 14:611–614, Sep. 1993.

[35] Yi-Sheng Chen and Jyu-Han Song. Semiblind channel estimation for MIMO-OFDM systems. *EURASIP J. Adv. Sig. Proc.*, 2012:212, Oct. 2012.

[36] Jinho Choi and Jeongseok Ha. On the achievable rate for wideband channels with estimated CSI. *Journal of Signal Processing Systems*, 66(1):75–86, Jan. 2012.

[37] Tzu-Han Chou, S.C. Draper, and A.M. Sayeed. Impact of channel sparsity and correlated eavesdropping on secret key generation from multipath channel randomness. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 2518 –2522, June 2010.

[38] K. Chugg, A. Anastasopoulos, and X. Chen. *Iterative Detection: Adaptivity, Complexity Reduction, and Applications.* The Springer International Series in Engineering and Computer Science. Springer US, 2001.

[39] T. Clevorn, S. Godtmann, and P. Vary. BER prediction using EXIT charts for BICM with iterative decoding. *Communications Letters, IEEE*, 10(1):49–51, Jan. 2006.

[40] R. Couillet and M. Debbah. *Random Matrix Methods for Wireless Communications*. Random Matrix Methods for Wireless Communications. Cambridge University Press, 2011.

[41] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. John Wiley & Sons, 2006.

[42] S.N. Crozier, D.D. Falconer, and S.A. Mahmoud. Least sum of squared errors (LSSE) channel estimation. *Radar and Signal Processing, IEE Proceedings F*, 138(4):371–378, Aug. 1991.

[43] I. Csiszar and J. Korner. Broadcast channels with confidential messages. *Information Theory, IEEE Transactions on*, 24(3):339 – 348, May 1978.

[44] E. Dall'Anese, A. Assalini, and S. Pupolin. On the effect of imperfect channel estimation upon the capacity of correlated MIMO fading channels. In *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pages 1 –5, Apr. 2009.

[45] R.C. Daniels, J.N. Murdock, T.S. Rappaport, and R.W. Heath. 60 GHz wireless: Up close and personal. *Microwave Magazine, IEEE*, 11(7):44–50, Dec. 2010.

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

[47] G.K. Dietl. *Linear Estimation and Detection in Krylov Subspaces*. Foundations in Signal Processing, Communications and Networking, 1. Deutsches MAB-Nationalkomitee beim Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, 2007.

[48] W. Diffie and M. Hellman. New directions in cryptography. *Information Theory, IEEE Transactions on*, 22(6):644 – 654, Nov. 1976.

[49] Zhi Ding, R.A. Kennedy, B. D O Anderson, and C.R. Johnson. Ill-convergence of Godard blind equalizers in data communication systems. *Communications, IEEE Transactions on*, 39(9):1313–1327, Sep. 1991.

[50] Dobrusin. General formulation of Shannon's main theorem in information theory. *Usp. Mat. Nauk.*, 14:3–104, 1959.

[51] A. Dogandzic and B. Zhang. Estimating Jakes' Doppler power spectrum parameters using the Whittle approximation. *Signal Processing, IEEE Transactions on*, 53(3):987 – 1005, Mar. 2005.

[52] Min Dong and Lang Tong. Optimal design and placement of pilot symbols for channel estimation. *Signal Processing, IEEE Transactions on*, 50(12):3055–3069, Dec. 2002.

[53] M. Dorpinghaus, A. Ispas, and H. Meyr. On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels. *Information Theory, IEEE Transactions on*, 58(5):2963 –2982, May 2012.

[54] Catherine Douillard, Michel Jzquel, Claude Berrou, Dpartement Electronique, Annie Picart, Pierre Didier, and Alain Glavieux. Iterative correction of intersymbol interference: Turbo-equalization. *European Transactions on Telecommunications*, 6(5):507–511, 1995.

[55] D. Elkouss, A. Leverrier, R. Alleaume, and J.J. Boutros. Efficient reconciliation protocol for discrete-variable quantum key distribution. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pages 1879 –1883, July 2009.

[56] B. Farhang-Boroujeny. Experimental study of semi-blind channel identification/equalization through pilot signals. In *Signal Processing, 1996., 3rd International Conference on*, volume 1, pages 618–621 vol.1, Oct. 1996.

[57] R. G. Gallager. *Low-Density Parity-Check Codes.* PhD thesis, Cambridge, MA:MIT Press, 1963.

[58] Matthew S Gast. *802.11 Wireless Networks: The Definitive Guide, Second Edition.* O'Reilly Media, Inc., 2005.

[59] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami. Channel estimation and symbol detection for block transmission using data-dependent superimposed training. *Signal Processing Letters, IEEE*, 12(3):226–229, Mar. 2005.

[60] G. B. Giannakis and C. Tepedelenlioglu. Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. *Proceedings of the IEEE*, 86(10):1969–1986, 1998.

[61] D. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *Communications, IEEE Transactions on*, 28(11):1867–1875, Nov. 1980.

[62] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[63] N. Gura, A. Patel, H. Eberle Wander, and S. C. Shantz. Elliptic curve cryptography and RSA on 8-bit CPUs. In *6th International Workshop on Cryptographic Hardware and Embedded Systems*, Aug. 2004.

[64] S. Gururaja. *Joint Optimal Pilot Placement and Power Allocation with Space Frequency Code Design and Adaptive Modulation for Video Transmission Over MIMO-OFDM Systems*. State University of New York at Buffalo, 2007.

[65] R.A. Haddad and T.W. Parsons. *Digital Signal Processing: Theory, Applications, and Hardware*. Electrical Engineering, Communications and Signal Processing Series. Freeman Company, 1991.

[66] Joachim Hagenauer. The EXIT chart - introduction to extrinsic information transfer. In *in Iterative Processing, In Proc. 12th Europ. Signal Proc. Conf (EUSIPCO)*, pages 1541–1548, 2004.

[67] J.R. Hampton. *Introduction to MIMO Communications*. Cambridge University Press, 2013.

[68] Stephen Hanly and David N. Tse. The multi-access fading channel: Shannon and delay limited capacities. In *in Proc. 33rd Allerton Conf*, pages 786–795, 1995.

[69] L.L. Hanzo, R.G. Maunder, J. Wang, and L.L. Yang. *Near-Capacity Variable-Length Coding: Regular and EXIT-Chart-Aided Irregular Designs*. Wiley - IEEE. Wiley, 2011.

[70] Amer A. Hassan, Wayne E. Stark, John E. Hershey, and Sandeep Chennakeshu. Cryptographic key agreement for mobile radio. *Digital Signal Processing*, 6(4):207 – 212, 1996.

[71] B. Hassibi and B.M. Hochwald. How much training is needed in multiple-antenna wireless links? *Information Theory, IEEE Transactions on*, 49(4):951–963, Apr. 2003.

[72] Lanlan He, Shaodan Ma, Yik-Chung Wu, and Tung-Sang Ng. IQ imbalance compensation: A semi-blind method for OFDM systems in fast fading channels. In *Circuits and Systems (APCCAS), 2010 IEEE Asia Pacific Conference on*, pages 362 –365, Dec. 2010.

[73] C. Hermosilla and L. Szczecinski. Performance evaluation of linear turbo receivers using analytical EXIT functions. In *Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004. 15th IEEE International Symposium on*, volume 2, pages 1307–1311 Vol.2, Sep. 2004.

[74] J.E. Hershey, A.A. Hassan, and R. Yarlagadda. Unconventional cryptographic keying variable management. *Communications, IEEE Transactions on*, 43(1):3 –6, Jan. 1995.

[75] F. Hlawatsch and G. Matz. *Wireless Communications Over Rapidly Time-Varying Channels*. Elsevier Science, 2011.

[76] Franois Horlin and Andr Bourdoux. *Digital Compensation for Analog Front-Ends: A New Approach to Wireless Transceiver Design*. John Wiley & Sons, Ltd, 2008.

[77] Jingyu Hua, Limin Meng, Xiaojian Xu, Dongming Wang, and Xiaohu You. Novel scheme for joint estimation of SNR, Doppler, and carrier frequency offset in double-selective wireless channels. *Vehicular Technology, IEEE Transactions on*, 58(3):1204 –1217, Mar. 2009.

[78] Jingyu Hua, Zhijiang Xu, Jin Li, Limin Meng, and Xiaohu You. Doppler shift estimator with MMSE parameter optimization for very low SNR environment in wireless communications. *Aerospace and Electronic Systems, IEEE Transactions on*, 44(3):1228 –1233, July 2008.

[79] Xiaozhou Huang and Hsiao-Chun Wu. Robust and efficient intercarrier interference mitigation for OFDM systems in time-varying fading channels. *Vehicular Technology, IEEE Transactions on*, 56(5):2517 –2528, Sept. 2007.

[80] Yuheng Huang and J.A. Ritcey. EXIT chart analysis of BICM-ID with imperfect channel state information. *Communications Letters, IEEE*, 7(9):434–436, Sep. 2003.

[81] W.C. Jakes. *Microwave mobile communications*. IEEE Press classic reissue. IEEE Press, 1974.

[82] J. Jiang. *Large Sample Techniques for Statistics*. Springer Texts in Statistics. Springer, 2010.

[83] V. Kafedziski and D. Morrell. Optimal adaptive equalization of multipath fading channels. In *Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on*, volume 2, pages 1443–1447 vol.2, 1994.

[84] K. D Kammeyer, V. Kuhn, and T. Petermann. Blind and nonblind turbo estimation for fast fading GSM channels. *Selected Areas in Communications, IEEE Journal on*, 19(9):1718–1728, Sep. 2001.

[85] A.P. Kannu and P. Schniter. Capacity analysis of MMSE pilot-aided transmission for doubly selective channels. In *Signal Processing Advances in Wireless Communications, 2005 IEEE 6th Workshop on*, pages 801 – 805, June 2005.

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

[87] A.P. Kannu and P. Schniter. On the spectral efficiency of noncoherent doubly selective block-fading channels. *Information Theory, IEEE Transactions on*, 56(6):2829 –2844, June 2010.

[88] K. Kansanen and T. Matsumoto. An analytical method for MMSE MIMO turbo equalizer EXIT chart computation. *Wireless Communications, IEEE Transactions on*, 6(1):59–63, Jan. 2007.

[89] M. Kavehrad and P. Mclane. Spread spectrum for indoor digital radio. *Communications Magazine, IEEE*, 25(6):32 –40, June 1987.

[90] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Number v. 1 in Fundamentals of Statistical Signal Processing. Prentice-Hall PTR, 1998.

[91] Tim Kelly and Michael Minges. Maximizing mobile - new world bank report points to human and economic development opportunities. Technical report, The World Bank, 2012.

[92] Hyosung Kim and J. K. Tugnait. Turbo equalization for doubly-selective fading channels using nonlinear Kalman filtering and basis expansion models. *Wireless Communications, IEEE Transactions on*, 9(6):2076–2087, June 2010.

[93] Hyosung Kim and J.K. Tugnait. Doubly-selective MIMO channel estimation using exponential basis models and subblock tracking. In *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, pages 1258–1261, Mar. 2008.

[94] Kyeong Jin Kim, T.A. Tsiftsis, and R. Schober. Semiblind iterative receiver for coded MIMO-OFDM systems. *Vehicular Technology, IEEE Transactions on*, 60(7):3156–3168, Sep. 2011.

[95] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel. Multi-input multi-output fading channel tracking and equalization using Kalman estimation. *Signal Processing, IEEE Transactions on*, 50(5):1065–1076, 2002.

[96] H. Koorapaty, A Hassan, and S. Chennakeshu. Secure information transmission for mobile radio. In *Information Theory, 1998. Proceedings. 1998 IEEE International Symposium on*, page 381, Aug. 1998.

[97] Christophe Laot, A. Glavieux, and J. Labat. Turbo equalization: adaptive equalization and channel decoding jointly optimized. *Selected Areas in Communications, IEEE Journal on*, 19(9):1744–1752, Sep. 2001.

[98] K.F. Lee and D.B. Williams. A space-time coded transmitter diversity technique for frequency selective fading channels. In *Sensor Array and Multichannel Signal Processing Workshop. 2000. Proceedings of the 2000 IEEE*, pages 149–152, 2000.

[99] Seok-Jun Lee, A.C. Singer, and N.R. Shanbhag. Linear turbo equalization analysis via BER transfer and EXIT charts. *Signal Processing, IEEE Transactions on*, 53(8):2883–2897, Aug. 2005.

[100] Kai Li and Xiaodong Wang. EXIT chart analysis of turbo multiuser detection. *Wireless Communications, IEEE Transactions on*, 4(1):300–311, Jan. 2005.

[101] Xin Li and Tan F. Wong. Turbo equalization with nonlinear Kalman filtering for time-varying frequency-selective fading channels. *Wireless Communications, IEEE Transactions on*, 6(2):691–700, 2007.

[102] D.N. Liu and M.P. Fitz. Joint turbo channel estimation and data recovery in fast fading mobile coded OFDM. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1 –6, Sep. 2008.

[103] R. Liu and W. Trappe. *Securing Wireless Communications at the Physical Layer*. Springer US, 2009.

[104] Yanpei Liu, S.C. Draper, and A.M. Sayeed. Secret key generation through OFDM multipath channel. In *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, pages 1 –6, Mar. 2011.

[105] Xiaoli Ma, G. B. Giannakis, and S. Ohno. Optimal training for block transmissions over doubly selective wireless fading channels. *Signal Processing, IEEE Transactions on*, 51(5):1351–1366, 2003.

[106] M. G. Madiseh. *Wireless Secret Key Generation Versus Capable Adversaries*. PhD thesis, University of Victoria, 2011.

[107] D. Markovic and R.W. Brodersen. *DSP Architecture Design Essentials*. Electrical Engineering Essentials. Springer, 2012.

[108] S. L. Marple. *Digital Spectral Analysis With Applications*. Prentice Hall, Australia, Sydney, 1987.

[109] Thomas L. Marzetta. BLAST training: Estimating channel characteristics for high capacity space-time wireless. In *Proc. 37th Annual Allerton Conference on Communications, Control, and Computing*, pages 958–966, 1999.

[110] Robert Maunder. Matlab EXIT charts. Accessed on `http://users.ecs.soton.ac.uk/rm/resources/matlabexit`.

[111] U. M. Maurer. Secret key agreement by public discussion from common information. *Information Theory, IEEE Transactions on*, 39(3):733 –742, May 1993.

[112] Michael McGuire, Alireza Movahedian, and Mihai Sima. Zero phase smoothing of radio channel estimates. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1608 –1612, Aug. 2012.

[113] M. Medard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *Information Theory, IEEE Transactions on*, 46(3):933 –946, May 2000.

[114] A. Milewski. Periodic sequences with optimal properties for channel estimation and fast start-up equalization. *IBM Journal of Research and Development*, 27(5):426–431, Sep. 1983.

[115] A.F. Molisch. *Wireless Communications*. Wiley - IEEE. Wiley, 2010.

[116] A.F. Molisch, Orlik P., Sahinoglu Z., and Zhang J. *UWB-based Sensor Networks and the IEEE 802.15.4a Standard - A Tutorial*. Mitsubishi Electric Research Laboratories, Inc., 2008.

[117] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue. Subspace methods for the blind identification of multichannel FIR filters. *Signal Processing, IEEE Transactions on*, 43(2):516–525, Feb. 1995.

[118] A. Movahedian and M. McGuire. An iterative receiver for fast fading channels using two Kalman filters. In *Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on*, pages 122 –127, Oct. 2011.

[119] A. Movahedian and M. McGuire. Capacity of iteratively estimated channels using LMMSE estimators. In *Communications, Computers and Signal Processing (PACRIM), 2013 IEEE Pacific Rim Conference on*, pages 200–205, Aug. 2013.

[120] A. Movahedian and M. McGuire. Estimation of fast fading channels for turbo receivers with high order modulation. *Vehicular Technology, IEEE Transactions on*, 62(2), 2013.

[121] A. Movahedian and M. McGuire. Low complexity estimation of fast fading radio channels for higher order modulation. In *Communications (ICC), 2013 IEEE International Conference on*, pages 5537–5541, June 2013.

[122] A. Movahedian and M. McGuire. Efficient and accurate semiblind estimation of MIMO-OFDM doubly-selective channels. In *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th (Accepted paper)*, Sep. 2014.

[123] L. Musavian, M.R. Nakhai, M. Dohler, and A.H. Aghvami. Effect of channel uncertainty on the mutual information of MIMO fading channels. *Vehicular Technology, IEEE Transactions on*, 56(5):2798 –2806, Sept. 2007.

[124] Soon Xin Ng, O.R. Alamri, Yonghui Li, J. Kliewer, and L. Hanzo. Near-capacity turbo trellis coded modulation design based on EXIT charts and union bounds. *Communications, IEEE Transactions on*, 56(12):2030–2039, Dec. 2008.

[125] Roald Otnes and Michael Tchler. EXIT chart analysis applied to adaptive turbo equalization. In *Proc. Nordic Signal Processing Symposium*, 2002.

[126] P. Z. Peebles. *Probability, random variables, and random signal principles*. McGraw-Hill, 1993.

[127] E. Perahia, Carlos Cordeiro, Minyoung Park, and L.L. Yang. IEEE 802.11ad: Defining the next generation Multi-Gbps Wi-Fi. In *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, pages 1–5, Jan. 2010.

[128] D.M. Pozar. *Microwave engineering*. Wiley, 1997.

[129] John G. Proakis and Dimitris K. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. Pearson Education, 2007.

[130] Fengzhong Qu and Liuqing Yang. On the estimation of doubly-selective fading channels. *Wireless Communications, IEEE Transactions on*, 9(4):1261–1265, Apr. 2010.

[131] J. Ran. *Signal Processing, Channel Estimation and Link Adaptation in MIMO-OFDM Systems*. Cuvillier, 2008.

[132] T.S. Rappaport. *Wireless communications: principles and practice*. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, 2002.

[133] T.S. Rappaport, E. Ben-Dor, J.N. Murdock, and Yijun Qiao. 38 GHz and 60 GHz angle-dependent propagation for cellular & peer-to-peer wireless communications. In *Communications (ICC), 2012 IEEE International Conference on*, pages 4568–4573, June 2012.

[134] F.M. Reza. *An Introduction to Information Theory*. Dover Books on Mathematics Series. Dover Publ., 1961.

[135] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 26(1):96–99, January 1983.

[136] M. Russell and G.L. Stuber. Interchannel interference analysis of OFDM in a mobile environment. In *Vehicular Technology Conference, 1995 IEEE 45th*, volume 2, pages 820 –824 vol.2, July 1995.

[137] A. Sayeed and A. Perrig. Secure wireless communications: Secret keys through multipath. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 3013 –3016, 31 2008-April 4 2008.

[138] Louis L. Scharf. The SVD and reduced rank signal processing. *Signal Processing*, 25(2):113–133, Nov. 1991.

[139] H. Schumacher, H. Tchouankem, J. Nuckelt, T. Kurner, T. Zinchenko, A. Leschke, and L. Wolf. Vehicle-to-Vehicle IEEE 802.11p performance measurements at urban intersections. In *Communications (ICC), 2012 IEEE International Conference on*, pages 7131 –7135, June 2012.

[140] C. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal, Vol 28, pp. 656715*, Oct. 1949.

[141] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.

[142] D. Shepherd, Zhenning Shi, M. Anderson, and M.C. Reed. EXIT chart analysis of an iterative receiver with channel estimation. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 4010–4014, Nov. 2007.

[143] M. Shirvanimoghaddam, Yonghui Li, and B. Vucetic. Near-capacity adaptive analog fountain codes for wireless channels. *Communications Letters, IEEE*, 17(12):2241–2244, Dec. 2013.

[144] D. Simon. *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006.

[145] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty -V: the discrete case. *Bell System Technical Journal, The*, 57(5):1371–1430, May 1978.

[146] F. Sowell. A decomposition of block Toeplitz matrices with application to vector time series. Technical report, Tech. Rep., 1989, 2013, Accessed on http://fsowell.tepper.cmu.edu/Papers/decomposition_of_block_toeplitz_matrices.pdf.

[147] M. Stege, P. Zillmann, and G. Fettweis. MIMO channel estimation with dimension reduction. In *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on*, volume 2, pages 417–421 vol.2, Oct. 2002.

[148] Zijian Tang and G. Leus. A novel receiver architecture for single-carrier transmission over time-varying channels. *Selected Areas in Communications, IEEE Journal on*, 26(2):366–377, Feb. 2008.

[149] Vahid Tarokh, Hamid Jafarkhani, and A.R. Calderbank. Space-time block codes from orthogonal designs. *Information Theory, IEEE Transactions on*, 45(5):1456–1467, July 1999.

[150] Emre Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10:585–595, 1999.

[151] S. Ten Brink, G. Kramer, and A. Ashikhmin. Design of low-density parity-check codes for modulation and detection. *Communications, IEEE Transactions on*, 52(4):670–678, Apr. 2004.

[152] Lang Tong and S. Perreau. Multichannel blind identification: from subspace to maximum likelihood methods. *Proceedings of the IEEE*, 86(10):1951–1968, Oct. 1998.

[153] Lang Tong, Guanghan Xu, and T. Kailath. A new approach to blind identification and equalization of multipath channels. In *Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on*, 1991.

[154] Lang Tong, Guanghan Xu, and T. Kailath. Blind identification and equalization based on second-order statistics: a time domain approach. *Information Theory, IEEE Transactions on*, 40(2):340–349, Mar. 1994.

[155] W.H. Tranter. *Principles of communication systems simulation with wireless applications*. Prentice Hall communications engineering and emerging technologies series. Prentice Hall, 2004.

[156] M. K. Tsatsanis and Zhengyuan Xu. Pilot symbol assisted modulation in frequency selective fading wireless channels. *Signal Processing, IEEE Transactions on*, 48(8):2353–2365, Aug. 2000.

[157] J.K. Tugnait, Shuangchi He, and Hyosung Kim. Doubly selective channel estimation using exponential basis models and subblock tracking. *Signal Processing, IEEE Transactions on*, 58(3):1275 –1289, Mar. 2010.

[158] J.K. Tugnait and Weilin Luo. On channel estimation using superimposed training and first-order statistics. *Communications Letters, IEEE*, 7(9):413–415, Sep. 2003.

[159] G.L. Turin, W.S. Jewell, and T.L. Johnston. Simulation of urban vehicle-monitoring systems. *Vehicular Technology, IEEE Transactions on*, 21(1):9–16, Feb. 1972.

[160] M. Tuchler, R. Koetter, and A.C. Singer. Turbo equalization: principles and new results. *Communications, IEEE Transactions on*, 50(5):754–767, May 2002.

[161] M. Tuchler, A.C. Singer, and R. Koetter. Minimum mean squared error equalization using a priori information. *Signal Processing, IEEE Transactions on*, 50(3):673–683, Mar. 2002.

[162] M. Tchler, S. Ten Brink, and J. Hagenauer. Measures for tracing convergence of iterative decoding algorithms. In *in Proc. 4th IEEE/ITG Conf. on Source and Channel Coding*, pages 53–60, 2002.

[163] A.R. Varma, L.L.H. Andrew, C. R N Athaudage, and J.H. Manton. Iterative algorithms for channel identification using superimposed pilots. In *Communications Theory Workshop, 2005. Proceedings. 6th Australian*, pages 195–201, Feb. 2005.

[164] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath. Optimal training for frequency-selective fading channels. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, volume 4, pages 2105–2108 vol.4, 2001.

[165] U. Wachsmann, R. F H Fischer, and J.B. Huber. Multilevel codes: theoretical concepts and practical design rules. *Information Theory, IEEE Transactions on*, 45(5):1361–1391, July 1999.

[166] Ping Wan, M. McGuire, and Xiaodai Dong. Near-optimal channel estimation for OFDM in fast-fading channels. *Vehicular Technology, IEEE Transactions on*, 60(8):3780–3791, Oct. 2011.

[167] Xiaodong Wang and H.V. Poor. Blind multiuser detection: a subspace approach. *Information Theory, IEEE Transactions on*, 44(2):677–690, Mar. 1998.

[168] P. Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50(1/2):pp. 129–134, 1963.

[169] R. Wilson, D. Tse, and R.A. Scholtz. Channel identification: Secret sharing using reciprocity in ultrawideband channels. *Information Forensics and Security, IEEE Transactions on*, 2(3):364 –375, Sept. 2007.

[170] Aaron D. Wyner. The Wire-tap Channel. *Bell Systems Technical Journal*, 54(8):1355–1387, Jan. 1975.

[171] Hua Yang, Jian Xiong, Suyue Li, and Lin Gui. Wavelet BEM based channel estimation over rapidly time-varying channels. In *Wireless Communications and Networking Conference Workshops (WCNCW), 2013 IEEE*, pages 71–75, Apr. 2013.

[172] Chunxuan Ye, A. Reznik, and Y. Shah. Extracting secrecy from jointly Gaussian random variables. In *Information Theory, 2006 IEEE International Symposium on*, pages 2593 –2597, July 2006.

[173] Chunxuan Ye, A. Reznik, G. Sternberg, and Y. Shah. On the secrecy capabilities of ITU channels. In *Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th*, pages 2030 –2034, 30 2007-Oct. 3 2007.

[174] T. Yoo and A. Goldsmith. Capacity and power allocation for fading MIMO channels with channel estimation error. *Information Theory, IEEE Transactions on*, 52(5):2203 –2214, May 2006.

[175] D. J. Young and N. C. Beaulieu. The generation of correlated Rayleigh random variates by inverse discrete Fourier transform. *Communications, IEEE Transactions on*, 48(7):1114–1127, 2000. ID: 1.

[176] Jung-Lang Yu, Biling Zhang, and Po-Ting Chen. Blind and semi-blind channel estimation with fast convergence for MIMO-OFDM systems. *Signal Processing*, 95:1–9, -02-01 2014.

[177] Yonghong Zeng and Tung-Sang Ng. A semi-blind channel estimation method for multiuser multiantenna OFDM systems. *Signal Processing, IEEE Transactions on*, 52(5):1419–1429, May 2004.

[178] W. Zhang, S. Vedantam, and U. Mitra. Joint transmission and state estimation: A constrained channel coding approach. *Information Theory, IEEE Transactions on*, 57(10):7084–7095, Oct. 2011.

[179] Yajun Zhang, Yuanyuan Gao, Shibin Su, and Guozhen Zang. The performance comparison of differential detection and coherent detection with imperfect channel estimation in cooperative communication. In *Communication Technology (ICCT), 2011 IEEE 13th International Conference on*, pages 7–11, Sep. 2011.