UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Applied Science

School of Electronics and Computer Science

Long-Range Prediction of Wireless Channels and Transmit Preprocessing Techniques

by

Wei Liu

BEng, MEng

A thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy at the University of Southampton

September 2007

Supervisor: Professor Lajos Hanzo

Dipl Ing, MSc, PhD, FIEEE, FREnd, DSc

Chair of the Communications Research Group

School of Electronics and Computer Science

University of Southampton

Southampton SO17 1BJ

United Kingdom

Supervisor: Doctor Lie-Liang Yang

BEng, MEng, PhD, Senior Member IEEE, Member IET

Reader of the Communications Research Group

School of Electronics and Computer Science

University of Southampton

Southampton SO17 1BJ

United Kingdom

Dedicated

to my beloved wife, Chuan Li, and my precious daughter, Chu-Yi.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

Long-Range Prediction of Wireless Channels and Transmit Preprocessing Techniques

by Wei Liu

The advanced channel quality-aware adaptive modulation and coding techniques employed in both existing and future wireless communication systems are capable of substantially improving the achievable system performance. Furthermore, other novel techniques employed by the base-station (BS), such as transmit preprocessing, pre-equalization and so on can be used for simplifying the design of the receiver. All these techniques require accurate channel state information (CSI) at the transmitter side. In these high-rate broad-band wireless systems the carrier frequency has to be high. Furthermore, practical transmit preprocessing techniques have to rely on CSI feedback, where the outdated CSI estimated by the remote receiver based on the past data quantized and then signalled back to the base station's (BS) transmitter may not be sufficiently accurate. Hence an improved CSI accuracy has to be sought, which may be achieved with the aid of predicting the CSI for the future instant of the next transmission burst on the basis of previously receive CSI, which is achievable, since the CSI is typically correlated.

Therefore, in this thesis, novel channel prediction techniques are investigated in the context of both narrowband and wideband communication systems. Furthermore, a range of transmit preprocessing techniques are designed for multiple input multiple output (MIMO) systems operating in both single-user and multiple-user scenarios, which benefit from the employment of channel prediction.

Both minimum mean square error (MMSE) channel predictors and Kalman-filtering assisted channel predictors are investigated in the context of narrowband channels. Then, for wideband channel, two-dimensional (2D) channel estimation and prediction was considered, which was capable of predicting both the frequency-domain (FD) and time-domain (TD) fluctuation of wideband channels.

An eigenmode transmission based single-user MIMO system was also investigated, which re-

quires the computation of the singular vectors. These can be determined from the singular value decomposition (SVD) of the channel's impulse response (CIR) matrix, which has to be carried out at the regular instants and hence imposes a high computational complexity. However, instead of the periodic estimation of the CIR matrix and its regular SVD, it is possible to directly track the output of the SVD, namely the singular vectors without performing the above-mentioned channel estimation and SVD.

As far as MIMO aided multi-user systems are concerned, both zero-forcing and MMSE BS pre-processing techniques were investigated, which aim for simplifying the design of the mobile station's (MS) receiver. Again, channel prediction was invoked for acquiring the CSI required for transmit preprocessing.

Furthermore, a SVD based transmit preprocessing algorithm was proposed for both uplink (UL) and downlink (DL) transmissions in the context of a MIMO system supporting multiple users and different power allocation schemes are designed for both UL and DL transmissions.

Finally, the thesis was concluded with the investigation of recurrent neural network (RNN) based nonlinear channel prediction.

Acknowledgement

First of all, I would like to thank my supervisors, Professor Lajos Hanzo and Dr. Lie-Liang Yang, for their invaluable guidance and inspiration during my work. I consider myself very fortunate to be a student of these learned men and have greatly benefited from their vast experience, meticulous methods of research and their sincerity towards pursuit of knowledge.

Many thanks are also due to the academic staff of the Communications Group, especially Dr. Soon Xin Ng, for their useful guidance in my research. I also owe many personal thanks to all my colleagues of the Communications Group, particularly to Dr. Jos Akhtman, Dr. Ming Jiang, Dr. Bin Hu, Dr. Xiang Liu, Dr. Jin Wang, Dr. Hua Wei and Dr. Andreas Wolfgang, too many to mention here, for their help.

Finally I owe special gratitude to my beloved wife, Chuan Li, who has supported me exceptionally well during these years and sacrificed much of her career to be with me in England. I also want to thank my precious daughter, Chu-Yi, for coming to be a beautiful part of my life.

List of Publications

1) **W. Liu, L. L. Yang and L. Hanzo** "Wideband channel estimation and prediction in single-carrier wireless systems", Proceedings of the IEEE Vehicular Technology Conference 2005 Spring, Stockholm, Sweden, 30 May-1 June 2005, vol.1, pp 543 - 547.

2) **W. Liu, L. L. Yang and L. Hanzo** "Recurrent Neural Network Based Narrowband Channel Prediction", Proceedings of the IEEE Vehicular Technology Conference 2006 Spring, Melbourne, Australia, 7-10 May 2006, Vol.5, pp 2173 - 2177.

3) **B. Hu, W. Liu, L. L. Yang and L. Hanzo** "Multiuser Decorrelating Based Long-Range Frequency-Domain Channel Transfer Function Prediction in Multicarrier DS-CDMA Systems", Proceedings of IEEE International Symposium on Spread Spectrum Techniques (ISSSTA) 2006, Manaus-Amazonia, Brazil, 28-31 August 2006, pp 163 - 167.

4) **W. Liu, S. X. Ng, L. L. Yang and L. Hanzo** "Joint Channel Prediction Aided Differentially Encoded TTCM and BICM-ID Assisted Eigen-Beamforming", IEE Electronics Letters, Vol. 43, No. 4, pp 232 - 234, February, 2007.

5) **W. Liu, L. L. Yang and L. Hanzo** "Subspace Tracking Based Blind MIMO Transmit Preprocessing", Proceedings of the IEEE Vehicular Technology Conference 2007 Spring, Dublin, Ireland, 23 - 25 April 2007, pp 2228 - 2232.

6) **S. X. Ng, W. Liu, J. Wang, M. Tao, L. L. Yang and L. Hanzo** "Performance Analysis of Iteratively Decoded Variable-Length Space-Time Coded Modulation", accepted by IEEE International Conference on Communications 2007, Glasgow, Scotland, UK, 24 - 28 June 2007.

7) **S. X. Ng, W. Liu, L. L. Yang and L. Hanzo** "Channel Prediction Aided Coded Modulation

Assisted Eigen-Beamforming", Proceedings of the IEEE Vehicular Technology Conference 2007 Spring, Dublin, Ireland, 23 - 25 April 2007, pp 1742 - 174.

8) **A. Ahreans, W. Liu, S. X. Ng, V. Kuehn, L. L. Yang and L. Hanzo** "SVD-Aided, Iteratively Detected Spatial Division Multiplexing Using Long-Range Channel Prediction", accepted by IEEE Workshop on Signal Processing Systems, Shanghai, China, 17 - 19 October 2007.

9) **W. Liu, L. L. Yang and L. Hanzo** "SVD Assisted Joint Multiuser Transmitter and Multiuser Detector Design for MIMO Systems", submitted to IEEE Journal on Selected Areas in Communications, June, 2007.

10) **W. Liu, L. L. Yang and L. Hanzo** "Channel Prediction Aided Multiuser Transmission in SDMA", submitted to Proceedings of the IEEE Vehicular Technology Conference 2008 Spring, Marina Bay, Singapore, 11 - 14 May 2008.

11) **W. Liu, L. L. Yang and L. Hanzo** "SVD Aided Joint Transmitter and Receiver Design for the Uplink of Multiuser Detection Assisted MIMO Systems", submitted to IEEE International Conference on Communications, 2008, Beijing, China, 19 - 23 May 2008.

12) **W. Liu, L. L. Yang and L. Hanzo** "SVD Assisted Joint Transmitter and Receiver Design for the Downlink of MIMO Systems", submitted to IEEE Wireless Communications and Networking Conference, 2008, Las Vegas, Nevada, USA, 31 March - 3 April 2008.

Contents

Chapter 1

Introduction

Transmitter pre-processing techniques, such as adaptive-rate transmission [4–12] are capable of substantially improving the achievable capacity of wireless systems, provided that the relevant channel state information (CSI) *about to be experienced* by the transmitted signal can be obtained by the transmitter before transmissions. In high-rate broad-band wireless communications systems the carrier-frequency is usually high, which results in a relatively high Doppler frequency associated with fast fading. For wireless communications systems experiencing fast fading, the outdated CSI estimated based on previously received CSI may not be sufficiently accurate for achieving a high transmitter pre-processing gain [13]. However, an improved accuracy may be achieved with the aid of long range **channel prediction** [14–16].

The literatures of channel prediction was briefly reviewed in Tables 1.1 to 1.3. In the context of channel prediction various sum-of-sinusoids [17] based channel modelling algorithms have been investigated [18–25]. The ROOT-MUSIC (ROOT-Multiple Signal Classification [26]) [18] and the modified ESPRIT (Estimation of Signal Parameters via Rotational Invariance Technique [27]) [19] were invoked for estimating the frequencies of sinusoids, followed by determining their amplitudes. Then the future complex-valued fading channel envelope can be linearly predicted. In [20] the employment of the ESPRIT algorithm was extended to wideband communication systems. By exploiting the wideband channel transfer function's (CTF) correlation accross the FD and jointly applying the ESPRIT algorithm for predicting the TD evolution of the wideband CTF at different frequencies achieved a better performance than that of the scheme invoking it at each frequency individually [20].

Furthermore, in [21] the 2-D Unitary-ESPRIT algorithm and in [22] the 1-D Unitary-ESPRIT algorithm have been employed for estimating the CTF. Once the CTF's parameters are determined, the CTF can be predicted not only in the time domain (TD), but also in the frequency domain (FD). Sim-

ilar principles were extended to Orthogonal Frequency Division Multiplexing (OFDM) based communication systems in [24] and were shown to outperform the linear Burg predictor [28]. In [23], the regularized Linear Minimum Mean Square Error (LMMSE) algorithm was proposed for calculating the complex-valued amplitude of a sinusoidal channel model used in a single-input multiple-output (SIMO) communication system, assuming the knowledge of the sinusoidal model's freqeuncies. This scheme performed better than the linear predictor of [29]. Moreover, the above-mentioned LMMSE based channel predictror was further investigated in conjunction with predicting and then taking into account the frequency estimation error of the sum of sinusoids channel model [30]. In [31] a Joint Moving Average and Sinusoidal (JMAS) prediction model was proposed and the LS algorithm was used for determining the CIR. Furthermore, in [25], a 2-D frequency estimation algorithm was used for determining the FD CTF in the context of an OFDM system.

Another alternative is to model the narrowband channel by an autoregressive (AR) process [5,32], which can be accurately predicted by a linear predictor [29, 33–40]. In [29, 33–35] the MMSE principle was employed for predicting the CSI. The most important characteristic of this algorithm is that the sampling rate is usually significiantly lower than the data rate, which facilitates the prediction of the channel for the duration of numerous data bits, hence we refer to this approach as long range channel prediction. Adaptive channel prediction was investigated based on the RLS and LMS techniques [39].

The concept of long range prediction was also employed in OFDM systems based predictors using various optimization criteria such as the MMSE [36], normalized LMS (NLMS) and RLS [40, 41] criteria as well as an LMS fuzzy logic controller (FLC-LMS) [42]. An innovative so-called robust predictor design was provided in [5, 43], which was designed to guarantee a certain level of performance, even if the channel's Doppler power spectral density deviated from that used in its design. Then this MMSE based robust channel predictor was investigated in the context of Multiple Input Multiple Output (MIMO) schemes [37] and in OFDM systems [38], respectively.

In [44] several algorithms based on the above-mentioned sum of sinusoids channel model and on an AR model were investiged. The achievable performances of all these algorithms were tested and compared in a range of different contexts, such as Jakes' model [17], a statistical model proposed by Dersch [45] and also for measured data [44], respectively. The correspondig results of [44] showed that the performances of the algorithms based on the sinusoidal model are better than those based on the AR model, when using data generated by Jakes' model [17]. However, when either data generated by the statistical model of Dersch [45] or measured data [44] were used, the algorithms based on the AR model performed better than those based on the sinusoidal model [44].

The prediction algorithms based on the AR channel exploited only the linear correlation of channel coefficients [29]. However, the correlation of the channel coefficients may be more accurately modelled by a nonlinear process [46–51]. Consequently, nonlinear prediction algorithms can also be used for predicting the fading channels' profile [49–54]. In [49], the multivariate adaptive regression splines (MARS) based model was used for predicting both the power and the complex-valued taps of the wideband fading channels, which outperformed the classic finite impulse response (FIR) models [49]. Furthermore, the nonlinear predictors based on a so-called chaotic attractor [50] and on the recurrent least squares support vector machine [51] outperformed the AR model predictor. In [52–54] neural network based algorithms have been investigated in order to predict fading channels. It was shown that the performance of the neural network based algorithms may become better than that of linear predictors [52,53] under specific propagation conditions. Moveover, in [55] a MLP based channel predictor was used for predicting realistic channel measurements and was shown to outperform the classic linear channel predictor.

A range of other prediction algorithms can also be found in [56–59]. In [56] the novel technique of particle filtering was invoked for blind channel prediction, while in [57] Markov model based channel envelope prediction was investigated in the context of an OFDMA system. In [58], the so-called minimum energy bandlimited channel predictor was proposed.

It is well known that multiple antennas employed both at the transmitter and receiver side are capable of substantially improving the achievable system performance [60–62]. Furthermore, transmit preprocessing [63] is a key technique of future MIMO-aided wireless communication systems, which has the potential of simplifying the receiver design. The premise of transmit preprocessing is that the acurate CSI must be avaible at transitter side, which can be obtained by channel estimation at receiver and then fedback from the receiver to the transmitter. However, the estimated CSI may become outdated by the time it was conveyed to the BS for transmit preprocessing, resulting in a performance degration. Hence channel prediction constitues an ideal candidate for circumventing this problem, since it has the potential of generating the required future CSI based on its past values with the latency of signalling it from the remote receiver.

In the case of a single-user MIMO system, one of the transmit preprocessing algorithms is constituted by the so-called eigenmode transmission technique [64], which requires the knowledge of the singular eigen vectors of the MIMO channel's CIR matrix at both the transmitter and receiver side. The singular eigen vector can be obtained by estimating the MIMO channel's matrix and subjecting it to singular value decomposion (SVD) for every transmission burst, which imposes a high computational complexity. Since only the knowledge of the non-zero singular eigen vectors is required for

eigenmode transmission, it is intuively reasonable to use an algorithm that is capable of tracking the non-zero singular eigen vectors of the MIMO channel's CIR matrix without estimating the channel itself for every transmission burst. Hence this subspace-tracking approach is capable of substantially reducing the computational complexity imposed.

In the context of MIMO-aided multiple user scenarios, zero forcing and MMSE transmit preprocessing are widely used for downlink transmission. However, in realistic communication systems, only imperfect CSI is available for transmit preprocessing. Hence, the performance of both zero forcing and MMSE preprocessing techniques has to be investigated in conjunction with imperfect CSI.

Besides downlink transmit preprocessing, uplink transmit preprocessing also constitutes a promising area of investigation, since it is potentially capable of simplifying the receiver design at the BS.

Against this background, the novel contributions of this thesis are as follows:

- Both MMSE and Kalman filter assisted long range channel prediction is investigated in the context of narrowband wireless channels [65–68].

- Channel estimation is carried out in both the TD as well as the FD and Kalman filtering assisted long range prediction is invoked for wideband single-carrier communication systems [69, 70].

- Projection approximation subspace tracking with deflation (PASTD) is combined with MIMO-aided eigenmode transmission and investigated in the context of a single user system, where the MIMO channel matrix does not have to be estimated for every transmission burst. Its periodically repeated SVD is also avoided, since the non-zero singular eigen vectors are directly tracked, hence we are capable of reducing the computational complexity imposed. Furthermore, differential BPSK modulation is invoked for removing the phase ambiguity imposed by the SVD of the MIMO channel matrix [71], albeit this is achieved at a commensurate performance degradation.

- Both zero forcing and MMSE downlink transmit preprocessing techniques were designed investigated for multiple user MIMO systems and their benefits were investigated in terms of the effect of the number of antennas both at the BS and the MS. Furthermore, a Kalman filtering assisted channel predictor was invoked for generating the predicted CSI at the transmitter and the effects of channel prediction were also investigated [72].

- SVD based transmit preprocessing was proposed for both the downlink and uplink in the context of a MIMO-aided multiple user system. Specifically, two different power allocation

schemes were investigated, which are capable of either maintaining the maximum information rate or maintaining the maximum attainable SNR in the context of the proposed SVD based transmit preprocessing algorithm.

- Recurrent neural network (RNN) based long range channel prediction was investigated [73] in the context of narrowband channels, in order to explore the potential benefits of non-linear prediction.

The outline of the thesis is as follows. In Chapter 2, the characteristics of wireless channels are briefly discussed, leading to the classification of channels. Following this, the statistical model of the wireless channel is described. In Chapter 3, MMSE based long range prediction of narrowband channels is investigated, followed by the study of Kalman filtering assisted long range prediction. In Chapter 4, wideband single carrier systems are considered. We commence our discourse by investigating two-dimensional channel estimation relying on both TD and FD techniques. Then, based on the results of the channel estimation techniques of Section 4.4, Kalman filtering assisted long range prediction is extended to wideband channels. Our simulation results characterize the performance of the proposed algorithm. In Chapter 5, PASTD subspace tracking based MIMO eigenmode transmission is investigated, which is capable of tracking the non-zero singular eigen vectors without estimating the MIMO channel matrix for each transmission burst followed by its SVD. Hence again, this technique is capable of reducing the computational complexity imposed. Furthermore, DPSK modulation is used in order to avoid the phase ambiguity imposed by SVD. In Chapter 6, Kalman filtering assisted channel prediction is applied into a MIMO-aided multiple user system employing either a zero forcing or MMSE downlink transmit preprocessing. The effects of number of antennas at both the BS and the MS as well as normalized maximum Doppler frequency are investigated in conjunction with the predicted CSI. In Chapter 7, a SVD based transmit preprocessing tecnique capable of both downlink and uplink transmissions is proposed. In particular, the maximum-information-rate and the maximum-SNR based power allocation schemes are investigated in conjunction with the proposed transmit preprocesing algorithm. In Chapter 8, based on the assumption that wireless channels may be modelled by a nonlinear process, RNN based channel prediction is investigated. Specifically, two different types of activation functions are considered, namely split- and full-activation. Furthermore, three different tranining algorithms, namely the real-time recurrent learning (RTRL), the global extended Kalman filter (GEKF) and the decoupled extended Kalman filter (DEKF) are investigated in the context of channels contaminated by either additive white Gaussian noise (AWGN) or impulsive noise. Simulation results are also provided for characterizing the performance of RNN based channel

predictors. Finally, our conclusions and future work ideas are provided in Chapter 9.

Year	Author(s)	Contribution
1991	Cavers. [74]	The optimum linear channel estimation was investigated in the context of the flat fading channel.
1997	Gao *et al.* [53]	The channel envelope's magnitude prediction was investigated using nonlinear neural network.
1998	Hwang and Winters [18]	Based on the sum of sinusoids model of the channel [17], the Doppler frequency of the channel model was estimated using the Root-MUSIC (ROOT Multiple Signal Classification [26]) algorithm. Then the complex-valued fading envelope was determined by the least-square (LS) algorithm.
1999	Andesen *et al.* [19]	Based on the sum of sinusoids channel model, the Doppler frequency was estimated using the modified ESPRIT (Estimation of Signal Parameters via Rotational Invariance Technique [27]) algorithm. Then the complex-valued fading envelope was determined by the LS estimation algorithm.
1999	Ekman and Kubin [49]	Narrowband channel prediction was investigated using the nonlinear MARS (multivariate adaptive regression splines) model.
2000	Duel-Hallen *et al.* [29]	The channel was modelled by an autogressive (AR) process, and the AR coefficients were determined based on the MMSE criterion.
2001	Dong *et al.* [20]	Based on the sum of sinusoids channel model, the Doppler frequency was estimated using the modified ESPRIT algorithm in the context of wideband channels. By exploiting the wideband channel transfer function's (CTF) FD correlation and jointly applying the ESPRIT algorithm for predicting the wideband CTF at different frequencies achieved a better performance than that of the scheme invoking it at each frequency individually.
2002	Semmelrodt and Kattenbach [21,22]	Based on the sum of sinusoids channel model, the Doppler frequency of a wideband channel model was estimated by the 2-D Unitary-ESPRIT and 1-D Unitary-ESPRIT algorithm.
2002	Ekamn *et al.* [75]	The channel's power was predicted based on an AR model and an algorithm was proposed for compensating its bias.
2003	Chen *et al.* [39]	Long range channel prediction was investigated using both the recursive least square (RLS) and the least mean square (LMS) algorithms, which can adaptively predict the channel without any knowledge of the channel's statistical properties.
2003	Semmelrodt and Kattenbach [44]	The performance of the sum of sinusoids and the AR model based channel prediction algorithms were compared.
2004	Chen *et al.* [23]	Based on the sum of sinusoids channel model, the Doppler frequency was estimated by the Unitary-ESPRIT algorithm. Then the regularized Linear MMSE (LMMSE) algorithm was used for determining the complex-valued amplitudes, which avoided the noise enhencement encountered by the LS algorithm.

Table 1.1: Major contributions on channel prediction

Year	Author(s)	Contribution
2004	Wong *et al.* [36]	For OFDM systems, the prediction of the channel's transfer function (CTF) is typically carried out in three different ways, namely by determining the predictor's coefficients seperately for all subcarriers or for the specific subcarriers only corresponding to pilot symbols and then reusing them for the surrounding other subcarriers. Alternatively we may perform channel impulse response (CIR) prediction in the TD by transferring the FD CTF to CIR in TD and then transforming the predicted CIR from the TD back to the CTF in the FD. The MMSE performance, the computational complexity and the memory requirements were compared for these three different algorithms.
2004	Sun *et al.* [50, 51]	The nonlinear correlation exhibited by the channel coefficients was investigated. Furthermore, recurrent LS Support Vector Machines (SVM) were invoked for exploiting the nonlinear correlation of the channel coefficients encountered in certain propagation enviroments, which was shown to outperform the AR model based linear MMSE algorithm.
2004	Cao *et al.* [76]	Wideband CIR prediction was transformed into CTF prediction by predicting CTF in the FD and then transforming it back to CIR in the TD.
2004	Castro *et al.* [56]	Tomlinson-Harashima pre-equalization was investigated with the aid of channel prediction, which was carried out by Kalman filtering or particle filtering. Both techniques showed a similar performance. However, particle filtering has an advantage over Kalman filtering, since it constitutes a blind algorithm.
2004	Luo *et al.* [37]	So-called robust MMSE channel estimation and prediction [5, 43] were investigated in the context of a MIMO system, where the predictor was designed by taking the channel's worst-case power-delay profile into account.
2004	Akhtman and Hanzo [38]	Robust MMSE channel prediction [5, 43] was investigated in the context of OFDM and MC-CDMA systems, where the predictor was designed by taking the channel's worst-case power-delay profile into account.
2005	Wong and Evans [24]	An OFDM system was investigated and its FD CTF was modeled at a given frequency by the sum of sinusoids, where the FD CTF parameters were estimated by the ESPRIT algorithm. Then the complex-valued channel magnitudes were determined. The future CTF at a given frequency was generated by extrapolation.
2005	Schafhuber and Matz [40]	An MMSE channel predictor was designed for OFDM systems, which exploited the TD and FD correlation of the CTF. Furthermore, by assuming that the transmitted data symbols had a zero mean and were independent identically distributed, a reduced-complexity MMSE predictor was proposed. Moreover, the CTF prediction was transformed into CIR prediction, which further reduced the complexity. Adaptive channel prediction was carried out using the normalized least mean square (NLMS) and recursive least squares (RLS) algorithms.

Table 1.2: Major contributions on channel prediction continued

Year	Author(s)	Contribution
2005	Chen *et al.* [30]	The regularized LMMSE channel prediction of [23] was further improved by predicting the discrepancy between the observed and predicted channel, in other words, by taking the FD estimation error into account.
2005	Klaue and Aguiar [55]	A multi-layer perception (MLP) based predictor was invoked for predicting the realistic propagation measurement based channel instead of a simulated channel, where the comparisions showed that it outperformed the classic linear predictor.
2006	Xie *et al.* [41]	Both an MMSE and simplified MMSE predictor desigened for OFDM systems were considered, which exploited both the TD and FD correlation of the CTF. Furthermore, the NLMS algorithm based channel prediction technique was investigated in both the TD and FD.
2006	Wen *et al.* [42]	The prediction of the CTF in an OFDM system was transformed into the prediction of the CIR and a fuzzy-logic controlled LMS (FLC-LMS) predictor was used for predicting each CIR tap.
2006	Liu and Liu [25]	An OFDM system was considered and the CTF of a given subcarrier frequency was modelled by the sum of sinusoids. Then the parameters were estimated by a joint 2-D FD CTF estimation algorithm, which was shown to outperform the 1-D frequency estimation based algorithm of [24].
2006	Baddour and Beaulieu [77]	The CIR coefficients were estimated for narrowband channel prediction by a smoothening filter in order to mitigate the effects of noise. Then the estimated channel coefficients were used for predicting the future CIR taps. Furthermore, this scheme was also investigated in the context of a scenario dispensing with any statistical knowledge concerning the channel.
2006	Chee *et al.* [57]	The finite state Markov channel (FSMC) model of the narrowband channel was reviewed and channel prediction was investigated in the context of an OFDMA system.
2006	Zemen *et al.* [58]	Based on the principle that the narrowband channel may be approximated by a linear combination of orthogonal basis functions, the basis functions corresponding to future time instants were determined and used for approximating the future CIR taps.
2007	Chen *et al.* [31]	Based on the algorithms provided in [23, 30] a joint moving average and sinusoidal (JMAS) prediction model was proposed and an LS algorithm based predictor was investigated.

Table 1.3: Major contributions on channel prediction continued

Chapter 2

Overview of Wireless Channel Models

2.1 Introduction

In wireless communication, modelling of the wireless channel constitues an important research field, which has attracted substantial research insterests [17, 78, 79]. Since this thesis concentrates on wireless channel prediction, in this chapter we provide an overview of the various channel models. The chapter is organized as follows. In Section 2.2, a brief of characteriztion of the wireless channel is provided. The classification of wireless channels is considered in Section 2.3, while in Section 2.4 their statistical model is provided. Finally, the simulation of wireless channels is discussed in Section 2.5 and our conclusions are offered in Section 2.6.

2.2 Characterization of Fading Channels

In wireless communications systems, the received signal experiences diverse channel impariments imposed by multipath propagation and Doppler frequency shift. More specifically, the signal received by the receiver at any point in space may be the superposition of a large number of plane waves having randomly distributed amplitudes, phases, angles of arrival (AOA) and different propogation delays [17, 80]. These multipath components combine vectorially at the receiver antenna and hence the received signal fluctuates quite dramatically as a function of time. Furthermore, due to the relative motion of the transmitter and the receiver, each multipath component may exprience an apparent shift in frequency, which is referred to as the Doppler frequency shift. Fig.2.1 depicts a horizontal $x-y$ plane, portraying a mobile station (MS) moving at velocity v along the x-axis [17]. Let the nth plane wave arrive at the MS's antenna at an angle of incidence θ_n, as shown in Fig.2.1. Then, the MS's

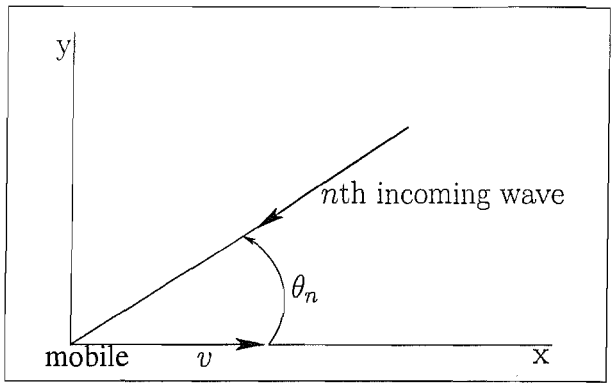


Figure 2.1: Illustration of Doppler effect

motion imposes a Doppler frequency shift upon the incident plane wave, which is given by [17]

$$f_n = f_{dm} \cos \theta_n, \tag{2.1}$$

where f_{dm} represents the maximum Doppler shift, which can be expressed as

$$f_{dm} = \frac{v}{\lambda_c} = f_c \frac{v}{c}, \tag{2.2}$$

where $\lambda_c = c/f_c$ is the wavelength of the arriving plane wave, f_c is the carrier frequency, c is the speed of light and v is the velocity of the MS.

Consider the transmission of a band-pass signal $s(t)$, which is expressed as

$$s(t) = \Re \left\{ s_l(t) e^{j2\pi f_c t} \right\}, \tag{2.3}$$

where $s_l(t)$ is the complex low-pass signal, f_c is the carrier frequency and $\Re\{\cdot\}$ denotes the real part of the argument. If the channel is comprised of N paths, then the received band-pass signal $x_n(t)$ of the nth path may be expressed as [80, 81]

$$\begin{aligned} x_n(t) &= \Re \left\{ \alpha_n(t) s_l(t - \tau_n(t)) e^{j2\pi [f_c + f_n][t - \tau_n(t)]} \right\} \\ &= \Re \left\{ \alpha_n(t) s_l(t - \tau_n(t)) e^{-j2\pi [(f_c + f_n)\tau_n(t) - f_n t]} e^{j2\pi f_c t} \right\}, \end{aligned} \tag{2.4}$$

where $\alpha_n(t)$ is the amplitude attenuation factor, $\tau_n(t)$ is the propagation delay and f_n represents the Doppler frequency shift associated with the nth path.

The band-pass signal $x(t)$ received via N distinct propogation paths can be expressed as

$$x(t) = \Re\left\{r(t)e^{j2\pi f_c t}\right\}, \tag{2.5}$$

where the complex low-pass signal $r(t)$ can be expressed as

$$\begin{aligned} r(t) &= \sum_{n=1}^{N} \alpha_n(t) e^{-j2\pi[(f_c+f_n)\tau_n(t)-f_n t]} s_l(t-\tau_n(t)) \\ &= \sum_{n=1}^{N} \alpha_n(t) e^{j\phi_n(t)} s_l(t-\tau_n(t)), \end{aligned} \tag{2.6}$$

while the phase of

$$\phi_n(t) = -2\pi\{[(f_c+f_n)\tau_n(t) - f_n t]\} \tag{2.7}$$

is associated with the nth path. Based on (2.6), the channel can be modeled by a time-variant linear filter having a complex low-pass channel impulse response (CIR) expressed as

$$c(\tau, t) = \sum_{n=1}^{N} \alpha_n(t) e^{j\phi_n(t)} \delta(\tau - \tau_n(t)). \tag{2.8}$$

Given (2.8), let us now discuss the characteristics of the wireless channels.

2.3 Classification of the fading channel

The mobile channel is a time varying multipath medium. The fading channel may undergo different types of fading according to the specific transmitted signal characteristics and channel parameters. Based on the amount of multipath-induced time domain spreading, the fading channel may be classified as frequency-nonselective or frequency-selective. By contrast, according to the Doppler spread encountered, fading channels may be classified as rapidly fading channels or slowly fading channels. The terms frequency-nonselective fading and frequency-selective fading can be used to describe the characteristics of signals in the frequency-domain (FD). By contrast, the terminology of fast fading and slow fading channels can be used for describing the characteristics of signals in the time-domain (TD). Let us first consider the family of frequency-nonselective and frequency-selective fading channels.

2.3.1 Frequency-Nonselective Fading versus Frequency Selective Fading

The autocorrelation function of the CIR seen in (2.8) can be expressed as

$$\phi_c(\tau_1, \tau_2; \Delta t) = \frac{1}{2}E[c^*(\tau_1; t)c(\tau_2; t + \Delta t)]. \tag{2.9}$$

Assuming a wide sense stationary uncorrelated scattering (WSSUS) channel, which implies that the attenuations and phase shifts of the channel associated with different path delays are uncorrelated, (2.9) can then be written as [81]

$$\frac{1}{2}E[c^*(\tau_1; t)c(\tau_2; t + \Delta t)] = \phi_c(\tau_1; \Delta t)\delta(\tau_1 - \tau_2). \tag{2.10}$$

If we let $\Delta t = 0$, the resultant autocorrelation fucntion $\phi_c(\tau_1; 0) \equiv \phi_c(\tau)$ is simply the average output power of the channel as a function of the time delay τ. The range of τ values over which $\phi_c(\tau)$ is essentially non-zero is referred to as the delay spread of the channel, which is denoted by T_m.

By taking the Fourier transform of $c(\tau, t)$ with respect to τ, we obtain the time-variant channel transfer fucntion (CTF) $C(f; t)$, which is given by [81]

$$C(f; t) = \int_{-\infty}^{+\infty} c(\tau; t)e^{-j2\pi f\tau}\, d\tau. \tag{2.11}$$

Assuming that the channel is wide sense stationary (WSS), the frequency domain autocorrelation function (ACF) $\phi_C(f_1, f_2; \Delta t)$ is given by

$$\phi_C(f_1, f_2; \Delta t) = \frac{1}{2}E[C^*(f_1; t)C(f_2; t + \Delta t)]. \tag{2.12}$$

Upon substituting Eq.(2.11) into Eq.(2.12), we obtain [81]

$$\begin{aligned}
\phi_C(f_1, f_2; \Delta t) &= \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} E[c^*(\tau_1; t)c(\tau_2; t + \Delta t)]e^{j2\pi(f_1\tau_1 - f_2\tau_2)}\, d\tau_1 d\tau_2 \\
&= \int_{-\infty}^{+\infty} \phi_c(\tau_1; \Delta t)e^{j2\pi(f_1 - f_2)\tau_1}\, d\tau_1 \\
&= \int_{-\infty}^{+\infty} \phi_c(\tau_1; \Delta t)e^{-j2\pi(\Delta f)\tau_1}\, d\tau_1 \equiv \phi_C(\Delta f; \Delta t),
\end{aligned} \tag{2.13}$$

where $\Delta f = f_2 - f_1$ and $\phi_C(\Delta f; \Delta t)$ is termed as the spaced-frequency, spaced-time correlation

function of the channel. Upon assuming $\Delta t = 0$, we have

$$\phi_C(\Delta f; 0) \equiv \phi_C(\Delta f) = \int_{-\infty}^{+\infty} \phi_c(\tau) e^{-j2\pi(\Delta f)\tau}\, d\tau. \tag{2.14}$$

As a result of the Fourier transform based relationship between $\phi_C(\Delta f)$ and $\phi_c(\tau)$, the coherence bandwidth B_c of the channel is given by [81]

$$B_c \approx \frac{1}{T_m}. \tag{2.15}$$

Based on the above analysis, we may state that if the channel's coherence bandwidth B_c is higher than the bandwidth of the transmitted signal, all the FD components of the transmitted signal will experience similar fading. Hence, this type of fading channels is said to be a frequency-nonselective fading channel or frequency-flat fading channel. Otherwise, the channel is said to be frequency-selective, namely when the channel's coherence bandwidth B_c is lower than the bandwidth of the transmitted signal. This is because in this type of channel two frequency components having a FD spacing higher than the coherence bandwidth may experience significantly different fading. The frequency-nonselective fading channel is also referred to as a narrowband fading channel, while the frequency-selective fading channel is also referred to as a wideband fading channel.

2.3.2 Slow Fading versus Fast Fading

The Fourier transform of $\phi_C(\Delta f; \Delta t)$ with respect to the variable Δt can be expressed as

$$S_C(\Delta f; \lambda) = \int_{-\infty}^{+\infty} \phi_C(\Delta f; \Delta t) e^{-j2\pi\lambda\Delta t}\, d\Delta t. \tag{2.16}$$

When Δf is set to zero and $S_C(0; \lambda) \equiv S_C(\lambda)$, the relationship in (2.16) becomes

$$S_C(\lambda) = \int_{-\infty}^{+\infty} \phi_C(0; \Delta t) e^{-j2\pi\lambda\Delta t}\, d\Delta t. \tag{2.17}$$

The function $S_C(\lambda)$ is termed as the Doppler power spectrum of the channel, which characterizes the signal intensity as a function of the Doppler frequency λ. The range of the values of λ over which $S_C(\lambda)$ is essentially nonzero is referred to as the Doppler spread B_d of the channel. Since $S_C(\lambda)$ is related to $\phi_C(\Delta t)$ by the Fourier transform, the coherence time $(\Delta t)_c$ of the channel is given by [81]

$$(\Delta t)_c \approx \frac{1}{B_d}. \tag{2.18}$$

Therefore, if the channel's coherence time $(\Delta t)_c$ is lower than the symbol period of the transmitted signal, the channel is said to be a fast fading channel. Otherwise, it is termed as a slowly fading channel.

In wireless channel simulations one of the most widely used flat fading channel model is Jakes' model, which is given by [17, 81]

$$C(0;t) \equiv c(t) = \sum_{n=1}^{N} \alpha_n(t) e^{j(2\pi f_{dm} t \cos\theta_n + \phi_n)}, \tag{2.19}$$

where $\alpha_n(t)$ is the amplitude attenuation factor, f_{dm} is the maximum Doppler frequency shift and ϕ_n is a random phase. In this model, when assuming $E[|c(t)|^2] = 1$, the autocorrelation of the time-variant fading envelope $C(f;t)$ is given by [17, 81]

$$\begin{aligned} \phi_C(\Delta t) &= \frac{1}{2} E[C^*(0;t) C(0;t+\Delta t)] \\ &= \frac{1}{2} J_0(2\pi f_{dm} \Delta t), \end{aligned} \tag{2.20}$$

where $J_0(\cdot)$ is the zero-order modified Bessel function of the first kind [80], which is depicted in Fig.2.2. Accordingly, the Fourier transform of Eq.(2.19) yields the Doppler power spectrum [81] of

$$\begin{aligned} S_C(f) &= \int_{-\infty}^{+\infty} \phi_C(\Delta t) e^{-j2\pi f \Delta t}\, d\Delta t \\ &= \int_{-\infty}^{+\infty} \frac{1}{2} J_0(2\pi f_{dm} \Delta t)\, d\Delta t \\ &= \begin{cases} \frac{1}{2\pi f_{dm} \sqrt{1-(\frac{f}{f_{dm}})^2}} & \text{if } |f| < f_{dm} \\ 0 & \text{if } |f| > f_{dm}. \end{cases} \end{aligned} \tag{2.21}$$

The normalized Doppler power spectrum is plotted in Figure.2.3.

2.4 Statistical Model of the Fading Channel

Let us consider the transmission of an unmodulated carrier having a frequency of f_c. Then we have $s_l(t) = 1$ for all t and the received complex low-pass signal seen in (2.6) can now be expressed as

$$\begin{aligned} r(t) &= \sum_{n=1}^{N} \alpha_n(t) e^{j\phi_n(t)} \\ &= r_I(t) + j r_Q(t), \end{aligned} \tag{2.22}$$

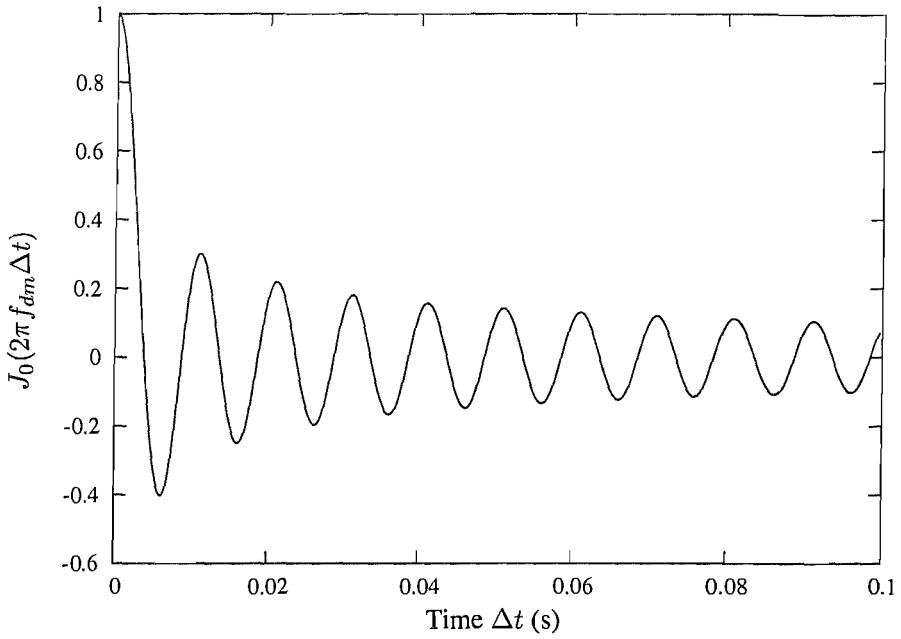


Figure 2.2: The zero-order Bessel-function of the first kind, $J_0(2\pi f_{dm}\Delta t)$

where $r_I(t)$ and $r_Q(t)$ are the real part and imagnary part of $r(t)$, respectively. Below a range of statistical models are summarized.

2.4.1 Rayleigh Distribution

According to the central limit theoem, when the value of N is sufficiently high, $r(t)$ of (2.22) can be modeled as a complex Gaussian process. When $r_I(t)$ and $r_Q(t)$ have a mean of zero and a common variance of σ^2, $r(t)$ is modeled as a zero-mean complex-valued Gaussian process. Then, the complex-valued received signal envelope of

$$X(t) = |r(t)| = \sqrt{r_I^2 + r_Q^2} \tag{2.23}$$

obeys the Rayleigh distribution at any time instant t, which has a probability density function (PDF) expressed as [80]

$$p_X(x) = \frac{x}{\sigma^2}\exp\left\{-\frac{x^2}{2\sigma^2}\right\} \quad x \geq 0. \tag{2.24}$$

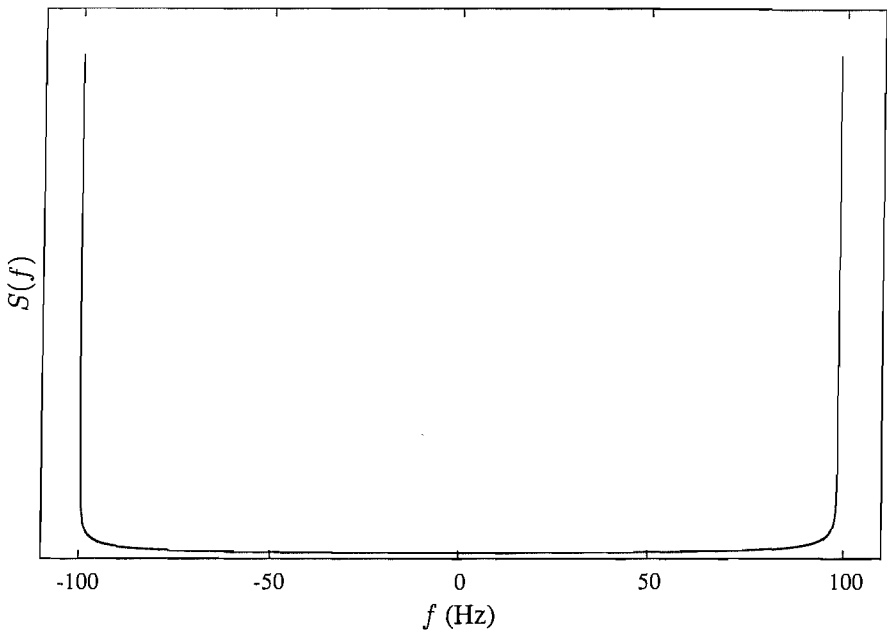


Figure 2.3: Normalized Doppler power spectrum $S(f)$, when assuming that the maximum Doppler frequency shift is $f_{dm} = 100Hz$

A channel having an amplitude obeying the Rayleigh distribution of (2.24) is referred to as a Rayleigh fading channel.

2.4.2 Ricean Distribution

When $r_I(t)$ and $r_Q(t)$ have the non-zero means of m_I and m_Q, respectively, then the complex-valued received signal envelope $X(t)$ obeys a Ricean distribution at any time instant t, which has a PDF expressed as [80]

$$p_X(x) = \frac{x}{\sigma^2}\exp\left\{-\frac{x^2+s^2}{2\sigma^2}\right\} I_0(\frac{xs}{\sigma^2}) \quad x \geq 0, \tag{2.25}$$

where $s^2 = m_I^2 + m_Q^2$ is the non-centrality parameter and $I_0(\cdot)$ is the 0th order modified Bessel function of the first kind. A channel having an amplitude obeying the Ricean distribution of (2.25) is hence termed as a Ricean fading channel.

2.4.3 Nakagami-m Distribution

Another popular distribution often used for modelling wireless channels is the Nakagami-m distribution. In the context of the Nakagami fading channel the fading amplitude $X(t)$ obeys the PDF of [80]

$$p_X(x) = \frac{2m^m x^{2m-1}}{\Gamma(m)\Omega^m} \exp\left\{-\frac{mx^2}{\Omega}\right\} \quad x \geq 0, \tag{2.26}$$

where $\Gamma(\cdot)$ is the gamma function, $\Omega = E[X^2]$ is the average envelope power of the channel and m is the Nakagami fading parameter, given by

$$m = \frac{\Omega^2}{E[(X^2 - \Omega^2)]}, \quad m \geq \frac{1}{2}. \tag{2.27}$$

2.5 Simulation of Wireless Channels

The simulation of bandlimited Rayleigh fading channels attracted substantial research attention [17, 78, 79, 82–84]. In this section, the IDFT based [82] Jakes' technique [17] and the autoregessive (AR) simulator [78, 79] of the flat Rayleigh fading channel are considered.

2.5.1 IDFT Simulator

The IDFT-based simulator was first proposed in [82] and can be implemented using the following steps [85]:

1) Specify the number of frequency domain points, N, used for representing $\sqrt{S_C(f)}$ and the maximum Doppler frequency shift f_{dm}. The value of N is usually set to be an integer power of 2;

2) Compute the frequency spacing between the adjacent spectral lines as $\Delta f = 2f_{dm}/(N-1)$. This defines the time duration of a fading waveform, $T = 1/\Delta f$;

3) Generate complex-valued Gaussian random variables for each of the $N/2$ positive frequency components;

4) Construct the negative frequency components of the noise source by conjugating the positive frequency values and assigning these values to the mirror-symmetric negative frequency components. This conjugate-complex symmetry ensures that the corresponding TD fading envelope is real-valued;

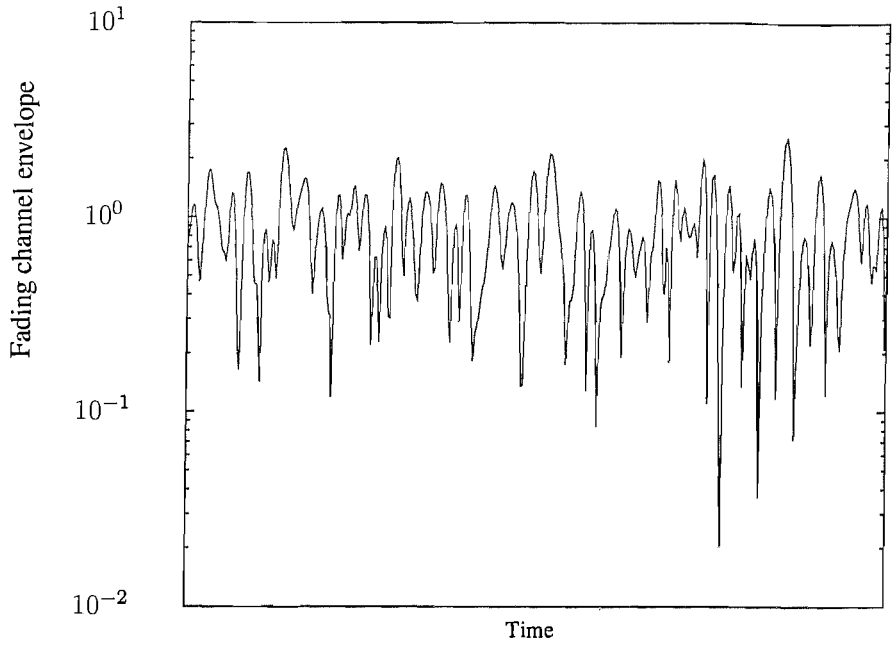


Figure 2.4: The envelope of the fading channel generated using the IDFT-based simulator

5) For each frequency component, multiply the in-phase and quadrature noise sources by the Doppler spectrum $\sqrt{S_C(f)}$;

6) Perform an IFFT on the resultant FD signals representing both the inphase and the quadrature components to generate two length-N time series and create an N-point time series of the fading magnitude according to (2.23).

As an example, the envelope of the fading channel generated by the IDFT-based Rayleigh simulator is plotted in 2.4

The IDFT-bsed channel simulator uses a complex-valued Gaussian random noise generator to produce a baseband line spectrum. The maximum frequency component of the line spectrum is f_{dm}. Again, since the fading magnitude is real-valued, the negative frequency components are constructed by simply conjugating the complex Gaussian values obtained for the positive frequencies. The resultant line spectrum is then multiplied by the discrete frequency representation of $\sqrt{S_C(f)}$ having the same number of points as the complex-valued noise source.

The IDFT-based technique is known to be an efficient fading channel generator. Howerver, when using the IDFT-based simulator, all random samples have to be generated using a single FFT operation and stored in memory. Hence, when a large number of values is required, the storage requirements

associated with this approach can make it unattractive.

2.5.2 Jakes Simulator

An alternative channel simulator has been suggested by Jakes [17], whihc has been widely used in the simulation of wireless channels. The principles behind Jakes' simulator became explicit in (2.19), which is repeated here for convienence:

$$r(t) = \sum_{n=1}^{N} e^{j(2\pi f_{dm} t \cos\theta_n + \phi_n)}, \tag{2.28}$$

where it is assumed that we have $\alpha_n(t) = 1$. We assume that the N number of arrival angles in (2.19) are uniformly distributed in $[0, 2\pi]$, that we have

$$\theta_n = \frac{2\pi n}{N}, \qquad n = 1, 2, \cdots N. \tag{2.29}$$

Furthermore, if $N/2$ is an odd integer, then as shown in [17, 80], a typical Rayleigh faded envelope can be generated by using M low-frequency oscillators having frequencies of $f_n = f_{dm}\cos(2\pi n/N)$, $n = 1, 2, \cdots, M$, and an oscillator having the frequency of f_{dm}, where we have

$$M = \frac{1}{2}\left(\frac{N}{2} - 1\right). \tag{2.30}$$

The fading channel's envelope generated by the Jakes model using $M = 8$ and $f_{dm} = 100Hz$ is shown in Fig.2.5.

Let the normalized autocorrelation function be defined as

$$\phi_{rr}(\tau) = \frac{E[r^*(t)r(t+\tau)]}{E[|r(t)|^2]}, \tag{2.31}$$

which is plotted in Fig. 2.6. From Fig.2.6 we can see that when the time delay of τ is low, for example when $\tau < 0.045s$, the Jakes model is capable of closely approximating the correlation function of (2.20). However, when the time delay is high, for example $\tau \geq 0.045s$, the autocorrelation function of the fading envelope generated by Jakes' simulator does not closely agree with the zeroth order Bessel function of first kind seen in (2.20), which is likely to be a consequence of using a low number of sinusoids for modelling the channel.

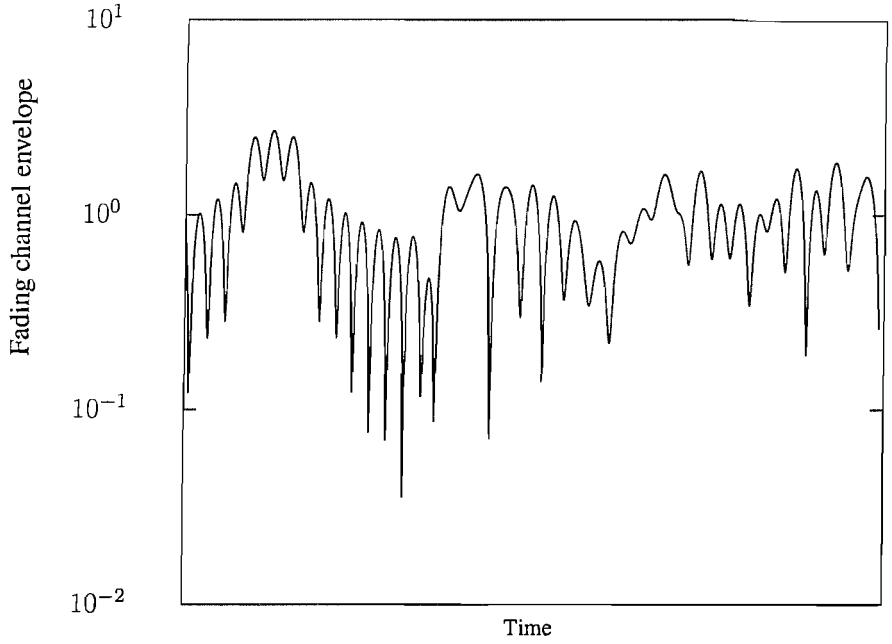


Figure 2.5: The envelope of the fading channel generated by Jakes' simulator with $M = 8$ and the maximum Doppler frequency shift of $f_{dm} = 100Hz$

2.5.3 AR Simulator

In [78, 79] an AR simulator has been proposed. Specifically, the AR simulator describes the flat Rayleigh fading channel as a complex-valued AR process of order p, which can be expressed as

$$c_n = \sum_{k=1}^{p} a_k c_{n-k} + w_n, \tag{2.32}$$

where c_n is the complex fading envelope $c(t)$ sampled at the time instants of $t = nT$, T is the sampling period, while w_n is a complex white Gaussian noise process at time instant $t = nT$ having zero mean and a variance of $\sigma^2_{w_n}$ and $(a_1, a_2, \cdots, a_p)$ are the AR process coefficients. Let us define a p-dimensional AR coefficient vector $\mathbf{a}$ as

$$\mathbf{a} = [a_1, a_2, \cdots, a_p]^T, \tag{2.33}$$

and a p-dimensional channel complex coefficient vector $\mathbf{c}_n$ as

$$\mathbf{c}_{n-1} = [c_{n-1}, c_{n-2}, \cdots, c_{n-p}]^T. \tag{2.34}$$

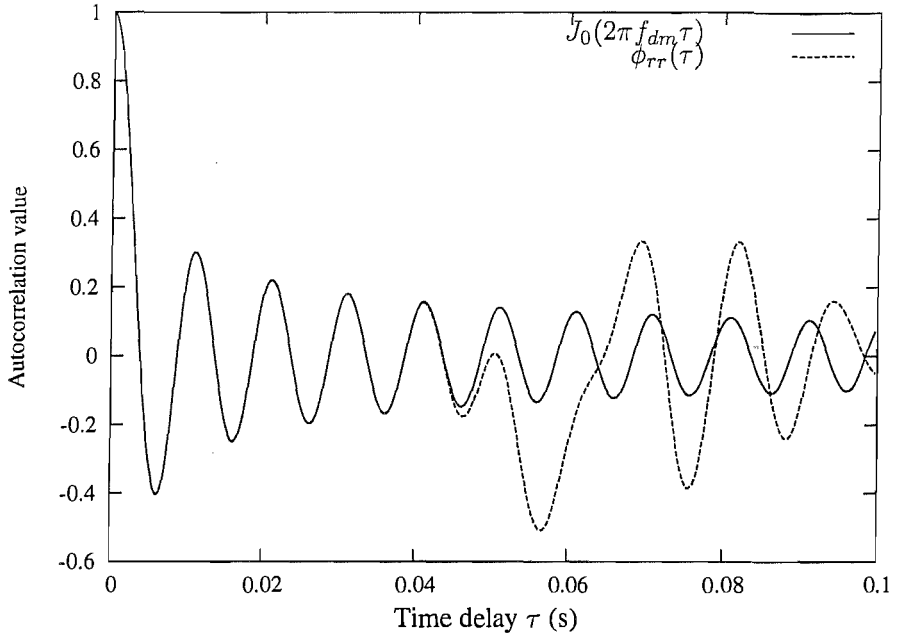


Figure 2.6: The zeroth order Bessel function of first kind $J_0(2\pi f_{dm}\tau)$ and the normalized autocorrelation function $\phi_{rr}(\tau)$ estimated from the Jakes model when $M = 8$ and assuming that the maximum Doppler frequency shift is $f_{dm} = 100Hz$

Thus, a can be formulated as [78]

$$\mathbf{a} = \mathbf{R}_{cc}^{-1}\mathbf{r}_c, \tag{2.35}$$

where $\mathbf{R}_{cc}$ is a $(p \times p)$-dimensional autocorrelation matrix having coefficients of $\mathbf{R}_{cc_{ij}} = E[c_{n-j}c_{n-i}^*]$, which is given by

$$\mathbf{R}_{cc} = E[\mathbf{c}_{n-1}^*\mathbf{c}_{n-1}^T] = \begin{bmatrix} r(0) & r^*(1) & \cdots & r^*(p-1) \\ r(1) & r(0) & \cdots & r^*(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix} \tag{2.36}$$

and $\mathbf{r}_c$ is a p-dimensional cross correlation vector with coefficients $\mathbf{r}_{c_j} = E[c_n c_{n-j}^*]$, which are given by

$$\mathbf{r}_c = E[c_n\mathbf{c}_{n-1}^*] = [r(1), r(2), \cdots, r(p)]^T, \tag{2.37}$$

where $r(l) = E[c_n c_{n-l}^*]$, $r^*(l) = E[c_n c_{n-l}^*]^*$. Accordingly, we have [78, 79]

$$\sigma_{w_n}^2 = r(0) - \sum_{k=1}^{p} a_k r(-k). \tag{2.38}$$

The AR simulator may accurately approximate theoretical statistics of a wireless channel by adjusting the order p of the model for practical finite-length implementations. A disadvantage of the AR model is that the autocorrelation matrix R_{cc} maybe singular when calculating the AR coefficients. In this case, a very small positive value has to be added to the main diagonal of the autocorrelation matrix in order to solve this problem [78]. In this report, 10^{-9} is added to the main diagonal of the autocorrelation matrix R_{cc} to make it non-singular.

2.6 Conclusion

In this chapter, a rudimentary characterization of fading channels was provided. Then the classification of wireless channels was discussed, introducing the concepts of frequency-nonselective fading, frequency-selective fading, slow fading and fast fading. Furthermore, statistical channel models were discussed and various flat Rayleigh fading channel simulators were investigated.

Chapter 3

Long-Range Linear Prediction of Narrowband Fading Channels

3.1 Introduction

The aim of the long range channel prediction is to forecast the future values of the channel coefficients as far ahead as possible. The AR model based long-range prediction (LRP) algorithm of [29] was designed for flat Rayleigh fading channels. Based on this model it was demonstrated in Section 3.2 that the minimum mean square error (MMSE) of channel prediction can be achieved, despite sampling the channel's profile at a rate lower than the symbol rate [29,33–35]. Then the channel coefficients to be used at symbol rate can be obtained by interpolation. Inspired by this principle, a LRP algorithm assisted by Kalman filtering will be designed in this chapter for flat Rayleigh fading channels and its performance will be investigated.

The objectives of LRP are multifold:

- These techniques may be used for simply improving the so-called zero-order channel prediction based on the currently received pilots at a receiver, which are assumed to be still valid at the current instant. This application becomes particularly important in the context of HSDPA-style near instantaneously adaptive modems.

- In the context of transmit preprocessing techniques the plausible rationale is that the unique, user-specific spatio-temporal CIR (ST-CIR) of each user may be employed for differentiating the transmitted signals destined for a specific user already at the transmitter, provided that this ST-CIR can be accurately predicted for the future instant of reception at the MS. This scenario

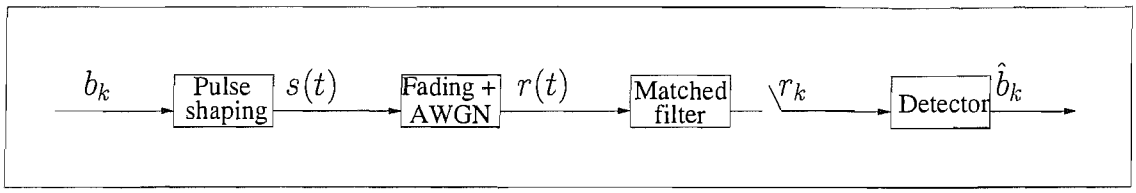


Figure 3.1: Basic schematic of communication systems

constitutes another high-significance LRP application scenario, because without its employment the transmitter would only have access to the outdated CTFs signalled by the remote receiver to the preprocessing-aided transmitter in the past, which would degrade the achievable preprocessing performance. It is worth noting that the more sophisticated the specific transmit preprocessor, the more accurate LRP is required for avoiding the erosion of its performance gain.

The rest of this chapter has the following structure. In Section 3.2 the principle of MMSE based LRP invoked for predicting narrowband fading channels is highlighted and its performance is characterized. As a more sophisticated design alternative, in Section 3.3, a Kalman filtering assisted LRP scheme designed for narrowband fading channels is advocated and characterized. Finally, our conclusions are offered in Section 3.4.

3.2 MMSE Assisted Long-Range Prediction

The basic schematic of the communication system considered is depicted in Figure. 3.1, where $\{b_k\}$ is the data symbol sequence, $s(t) = \sum_k b_k g(t - kT_b)$ is the transmitted complex-valued low pass signal, $g(t)$ is the transmitter pulse shape and T_b is the data symbol duration. The complex low pass received signal $r(t)$ seen in Figure. 3.1 is given by

$$r(t) = c(t)s(t) + n(t), \tag{3.1}$$

where $c(t)$ is the non-dispersive fading coefficient, which was given in (2.19) and repeated here for convenience

$$c(t) = \sum_{n=1}^{N} \alpha_n e^{j(2\pi f_n t + \phi_n)}, \tag{3.2}$$

where N represents the number of scatterers, α_n, f_n and ϕ_n are the amplitude, Doppler frequency shift and phase associated with the nth path, respectively, while $n(t)$ is the AWGN process having a zero mean and a variance of $N_0/2$ per dimension. Additionally, without loss of generality, it is assumed that the average channel power obeys $E[|c(t)|^2] = 1$. Furthermore, by sampling the output of the matched filter (MF) at the symbol rate, the communication system can be modeled by a discrete-time system given by

$$r_k = c_k b_k + n_k, \tag{3.3}$$

where r_k, c_k and n_k are obtained from the complex low pass received signal $r(t)$, from the complex-valued fading profile $c(t)$ and the AWGN $n(t)$ by sampling them at the time instants of $t = kT_b$, respectively. The output $\hat{b}_k$ of the detector represents the estimate of b_k. An AR model based linear LRP algorithm wasn proposed in [29,33–35]. The principles behind it may be described as follows.

3.2.1 Principles of MMSE Assisted Long-Range Prediction

Let us assume that a sequence of p previous samples of the complex-valued fading channel profile is obtained at the time instant $t = nT_s$, where T_s is the sampling interval duration, in contrast to the data symbol duration. The MMSE prediction of the future channel sample $\hat{c}_n$ based on p previous samples, namely $c_{n-1}, \cdots, c_{n-p}$ is given by

$$\hat{c}_n = \sum_{j=1}^{p} d_j c_{n-j}, \tag{3.4}$$

where p is the order of the AR model and $\hat{c}_n$ is the estimate of c_n. Let

$$\mathbf{c}(n-1) = [c_{n-1}, c_{n-2}, \cdots, c_{n-p}]^T, \tag{3.5}$$

be a p-dimensional vector containing the p number of observation samples. Let furthermore

$$\mathbf{d} = [d_1, \cdots, d_p]^T, \tag{3.6}$$

be the long-range predictor's weight vector. Then, the optimal predictor's weight vector $\mathbf{d}_o = [d_{o1}, \cdots, d_{op}]$ generated using the minimum mean square error (MMSE) cirterion is given by [1]

$$\mathbf{d}_o = \mathbf{R}^{-1}\mathbf{r}, \tag{3.7}$$

where $\mathbf{R}$ is the $(p \times p)$-dimensional autocorrelation matrix of the observation sample vector $\mathbf{c}(n-1)$, which is given by

$$\mathbf{R} = E[\mathbf{c}(n-1)\mathbf{c}^H(n-1)] = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r^*(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(p-1) & r^*(p-2) & \cdots & r(0) \end{bmatrix}, \tag{3.8}$$

where $r(l) = E[c_n c^*_{n-l}]$, $r^*(l) = E[c_n c^*_{n-l}]^*$. It is explicit that we have $r(-l) = r^*(l)$. Furthermore, in (3.7) $\mathbf{r}$ is the p-dimensional autocorrelation vector of the observation sample vector $\mathbf{c}(n-1)$ and of the desired channel sample c_n to be predicted, which is given by

$$\mathbf{r} = E[\mathbf{c}(n-1)c_n^*] = E\begin{bmatrix} c_{n-1}c_n^* \\ c_{n-2}c_n^* \\ \vdots \\ c_{n-p}c_n^* \end{bmatrix} = \begin{bmatrix} r^*(1) \\ r^*(2) \\ \vdots \\ r^*(p) \end{bmatrix} = \begin{bmatrix} r(-1) \\ r(-2) \\ \vdots \\ r(-p)) \end{bmatrix}, \tag{3.9}$$

where $\mathbf{r}(-j) = E[c_{n-j}c_n^*]$.

Note that the samples in Eq.(3.4) have to be taken at least at the Nyquist rate given by twice the maximum Doppler frequency f_{dm}. In LRP, it can be shown [29, 33–35] that the sampling rate can be chosen to be close to the Nyquist rate, which is usually much lower than the data symbol rate, as seen in (3.3). After the LRP, interpolation between the predicted samples can used in order to provide estimates of the fading profile at the data symbol rate [33], which generates a channel estimate for each data symbol.

In order to show that low-rate sampling may nonetheless result in sufficiently accurate LRP, when the filter length p in (3.4) is fixed, we extend the one-step prediction of (3.4) to a general channel prediction problem as follows. More explicitly, in this general channel prediction approach, our objective is to find the MMSE estimate of a future sample $c(\tau)$ at the time instant τ, where we have $\tau > 0$, by observing p previous samples collected both at and prior to time zero at the sampling rate of $f_s = 1/T_s$. Let

$$\mathbf{d}_f = [\ d_0 \quad d_1 \quad d_2 \quad \cdots \quad d_{p-1}]^T \tag{3.10}$$

be the long-range predictor's weight vector consisting of the p number of AR coefficients. Let fur-

thermore

$$\mathbf{c}_f = [\ c_0 \quad c_{-1} \quad c_{-2} \quad \cdots \quad c_{-(p-1)}\ ]^T \tag{3.11}$$

be a p-dimensional vector consisting of p previous data symbols. The predicted value to $c(\tau)$ at the time instant $\tau > 0$ can be expressed as

$$\hat{c}(\tau) = \sum_{k=0}^{p-1} d_k^* c_{-k} = \mathbf{d}_f^H \mathbf{c}_f, \tag{3.12}$$

where $\hat{c}(\tau)$ is the estimate of $c(\tau)$. Additionally, the estimation error $e(\tau)$ is given by

$$e(\tau) = c(\tau) - \hat{c}(\tau). \tag{3.13}$$

Let the predictor-optimization cost function J be defined by

$$\begin{aligned} J &= E[e(\tau)e^*(\tau)] \\ &= E[|e(\tau)|^2]. \end{aligned} \tag{3.14}$$

Then the MMSE predictor is obtained by choosing the weight vector $\mathbf{d}$, which minimizes the cost function J.

Accordingly, the optimum weight vector of the MMSE predictor can be expressed as

$$\mathbf{d}_{of} = \mathbf{R}^{-1}\mathbf{r}, \tag{3.15}$$

where the optimum coefficient vector $\mathbf{d}_{of}$ is given by

$$\mathbf{d}_{of} = [\ d_{o0} \quad d_{o1} \quad d_{02} \quad \cdots \quad d_{o(p-1)}]^T\ . \tag{3.16}$$

Furthermore, in (3.15) $\mathbf{R}$ is a $(p \times p)$-dimensonal autocorrelation matrix of observation channel profile

vector $\mathbf{c}_f$, which is expressed as

$$\mathbf{R} = E[\mathbf{c}_f \mathbf{c}_f^H] = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r^*(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(p-1) & r^*(p-2) & \cdots & r(0) \end{bmatrix}, \tag{3.17}$$

while $\mathbf{r}$ is the p-dimensional cross-correlation vector of the observation channel profile vector $\mathbf{c}_f$ and of the desired channel sample τ, which is expressed as

$$\mathbf{r} = E[\mathbf{c}_f c^*(\tau)] = E\begin{bmatrix} c_0 c^*(\tau) \\ c_{-1} c^*(\tau) \\ \vdots \\ c_{-(p-1)} c^*(\tau) \end{bmatrix} = \begin{bmatrix} r^*(\tau) \\ r^*(\tau+1) \\ \vdots \\ r^*(\tau+p-1) \end{bmatrix} = \begin{bmatrix} r(-\tau) \\ r(-(\tau+1)) \\ \vdots \\ r(-(\tau+p-1)) \end{bmatrix}. \tag{3.18}$$

In this chapter, we assume $r(\tau) = J_0(2\pi f_{dm}\tau)$ [17]. Hence, the autocorrelation function $\mathbf{R}$ of (3.17) can be simplified to

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix}. \tag{3.19}$$

and the cross-correlation function $\mathbf{r}$ seen in (3.18) can be expressed as

$$\mathbf{r} = \begin{bmatrix} r(\tau) \\ r(\tau+1) \\ \vdots \\ r(\tau+p-1) \end{bmatrix} \tag{3.20}$$

Finally, the MMSE estimate of the channel profile sample $c(\tau)$ at time instant of $\tau > 0$ can be expressed as

$$\begin{aligned} \hat{c}_o(\tau) &= \sum_{k=0}^{p-1} d_{ok}^* c_{-k} \\ &= \mathbf{d}_{of}^H \mathbf{c}_f. \end{aligned} \tag{3.21}$$

Order of AR model p	20
Maximum dopper f_{dm}	100Hz
Data symbol rate f_b	25kHz

Table 3.1: Parameters of the MMSE predictor

Since we assumed that $\mathbf{c}_f$ has a zero mean, the predicted value $\hat{c}(\tau)$ also has a zero mean. Furthermore, the variance of $\hat{c}(\tau)$ can be evaluated using (3.21), which yields

$$\begin{aligned}\sigma^2_{\hat{c}_o(\tau)} &= E[\mathbf{d}_{of}^H \mathbf{c}_f \mathbf{c}_f^H \mathbf{d}_{of}] \\ &= \mathbf{d}_{of}^H E[\mathbf{c}_f \mathbf{c}_f^H] \mathbf{d}_{of} \\ &= \mathbf{d}_{of}^H \mathbf{R} \mathbf{d}_{of} \\ &= \mathbf{r}^H \mathbf{R}^{-1} \mathbf{r}. \end{aligned} \tag{3.22}$$

The minimum MSE can be expressed as [1]

$$J_{min} = \sigma^2_{c(\tau)} - \mathbf{r}^H \mathbf{R}^{-1} \mathbf{r}, \tag{3.23}$$

where $\sigma^2_{c(\tau)}$ is the variance of $c(\tau)$.

3.2.2 Performance Analysis of the Long-Range MMSE Predictor

In this section, we use a number of examples to characterize the estimation performance of the long-range predictor. Note that the effects of the channel noise on the predictor can be incorporated into the autocorrelation matrix $\mathbf{R}$ by including $1/SNR * \mathbf{I}$ in (3.17), where $\mathbf{I}$ is a $(p \times p)$-dimensional identity matrix. Note that the autocorrelation matrix $\mathbf{R}$ can be singular in the noiseless case, but $\mathbf{R}$ of (3.15) is usually invertible , when additive noise is present. Furthermore, when assuming $E(|c_k|^2) = 1$, (3.23) can be rewritten as

$$\begin{aligned} J_{min} = E[|e(\tau)|^2] &= \sigma^2_{c(\tau)} - \mathbf{r}^H \mathbf{d}_{of} \\ &= 1 - \sum_{k=0}^{p-1} d_{ok} r(\tau + k). \end{aligned} \tag{3.24}$$

Let us first consider a predictor having a fixed order of p, but a variable sampling rate of f_s. It can be shown that as f_s increases, the portion of the autocorrelation function spanned by the samples

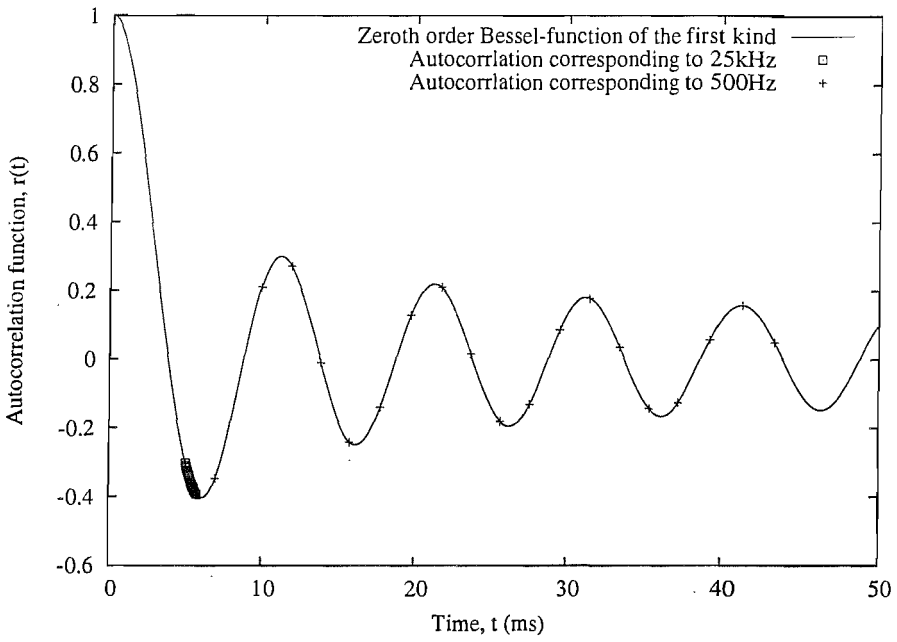


Figure 3.2: Theoretical autocorrelation function of the Rayleigh fading channel and the observation interval spaned by the autocorrelation when assuming a prediction range of $\tau = 5ms$ and two different sampling frequncies of $f_{s1} = 25kHz$ and $f_{s2} = 500Hz$, respectively. The remaining parameters are assumed to be as in Table 3.1.

$r(\tau + k)$ in (3.24) decreases, which can be seen in Figure. 3.2. As shown in Figure. 3.2, when a higher sampling frequency of $f_s = 25KHz$ is adopted, the observation interval spanned by the autocorrelation function samples is only about $0.76ms$. The autocorrelation value range spanned by these autocorrelation function samples along this interval is small. Consequently, when we aim for predicting the channel far ahead, i.e., when τ is large, based on these samples, the correlation between the desired channel sample and the observed samples becomes low, and the corresponding MMSE generated in (3.24) increases. However, when a lower sampling frequency of $f_s = 500Hz$ is adopted, the observed samples are spaced far apart, which results in an observation interval duration of 38ms. In this case, as shown in Figure. 3.2, the autocorrelation sample values vary significantly for any realistic prediction range. Due to the high sidelobes of the autocorrelation function, some of these autocorrelation samples are sufficiently large to prevent the MMSE in (3.24) from becoming very large [29].

The MMSE performance of LRP associated with various sampling rates against the prediction range τ is portrayed in Figure 3.3, when assuming a SNR higher than 100dB. The other paramers were the same as those shown in Table 3.1. The reason for us to select a high SNR value is to com-

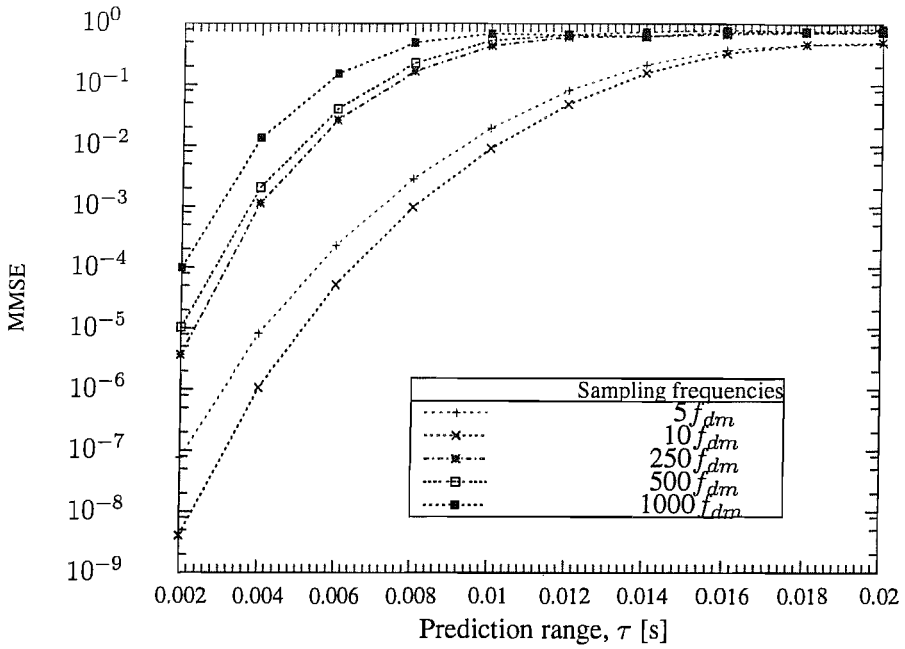


Figure 3.3: MMSE performance of LRP for the various values of sampling rate, f_s, against the prediction range τ when assuming a SNR higher than 100dB. The remaining paramers are assumed to as in Table 3.1.

pare the achievable performance, when various sampling rates were considered. In our forthcoming discussions, different p and SNR values will be considered. In Figure 3.3, the MMSE curves were computed from (3.24) for a given prediction range of τ. For example, when we have $\tau = 0.002s$, which corresponds to a 50-data-symbol look-ahead, when a sampling frequency of $f_s = 25Hz$ is adopted. However, this scenario implies using one-step prediction, when setting the sampling frequeny to $f_s = 500Hz$. As seen from Figure 3.3, the channel profile can be predicted at a given time instant of τ in the future, when using different sampling rates. However, the prediction becomes more accurate, when using relatively low sampling rates [29]. Hence, in the context of LRP, the channel profile's sampling rate can be significantly lower than the data rate, provided that the order p of prediction filter remains the same. Hence, the LRP of the channel profile becomes feasible.

The effect of the sampling rate on the MMSE performance achievable for the prediction range of $\tau = 4ms$, at a SNR higher than 140dB is shown in Figure 3.4, when using various predictor orders p. The remaining parameters are shown in Table 3.1. We observe from Figure 3.4, that for a given predictor order p, there exists an optimal sampling rate, which minimizes the corresponding MMSE. As shown in Fig.3.4, the optimum sampling rate is approximately 1KHz for a moderate to a high order of p, ranging from $p = 10$ to $p = 100$.

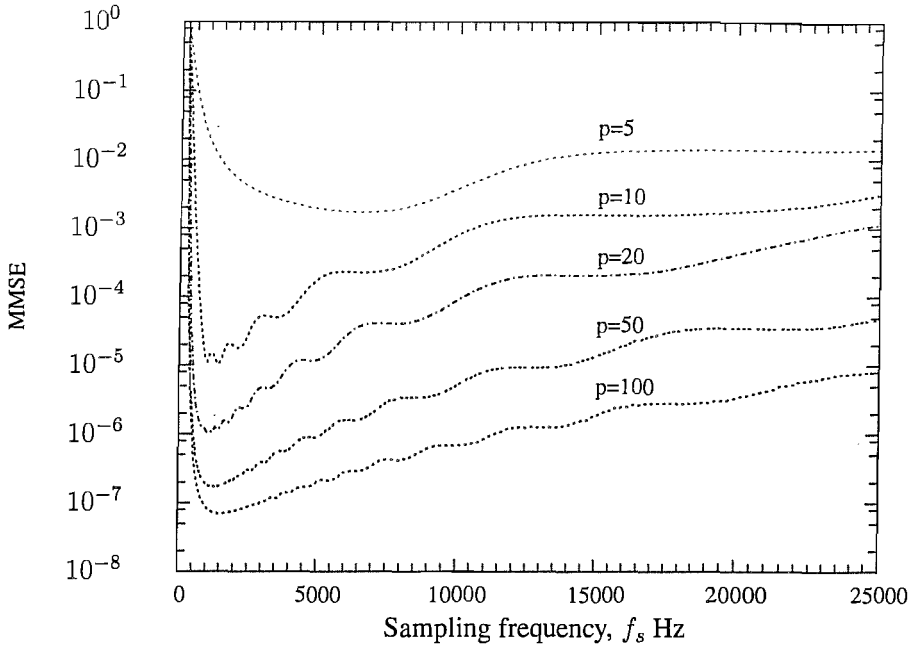


Figure 3.4: MMSE of the long-range predictor versus sampling rate, f_s performance for a prediction range $\tau = 4ms$, a SNR higher than 100dB when using various predictor orders p. The remaining parameters are assumed to be the same as in Table 3.1.

In Figuure 3.5 the MMSE performance versus the predictor order p is plotted for different values of the SNR, when assuming that the sampling rate was $f_s = 500Hz$ and the prediction range was $\tau = 2ms$. The remaining parameters are assumed to be as in Table 3.1. As shown in Figure 3.5, when the predictor order p increases, the MMSE curves approach a floor value. Furthermore, for any a given SNR value and for a low predictor order p, the MMSE value decreases near-linearly upon increasing the predictor order p. However, after the predictor order p reaches a certain value, the MMSE approaches a floor and remains near-constant, when increasing the value of p. Since the predictor order p is closely related to the prediction complexity and the complexity is usually on the order of $O(p^2)$, the affordable complexity predetermines the LRP order along with the remaining parameters, such as f_{dm}, τ, SNR, etc. Additionally, we can see from Figure 3.5 that when the SNR value increases, usually a long-range predictor having a relatively higher order of p is required, in order to achieve an improved prediction performance. In other words, when the SNR value increases, the MMSE floor decreases as a function of the predictor order.

Finally, in Figure 3.6 the prediction accuracy of the LRP algorithm is characterized, when the channel's profile is generated by Jakes' simulator using nine oscillators. We assumed that the SNR

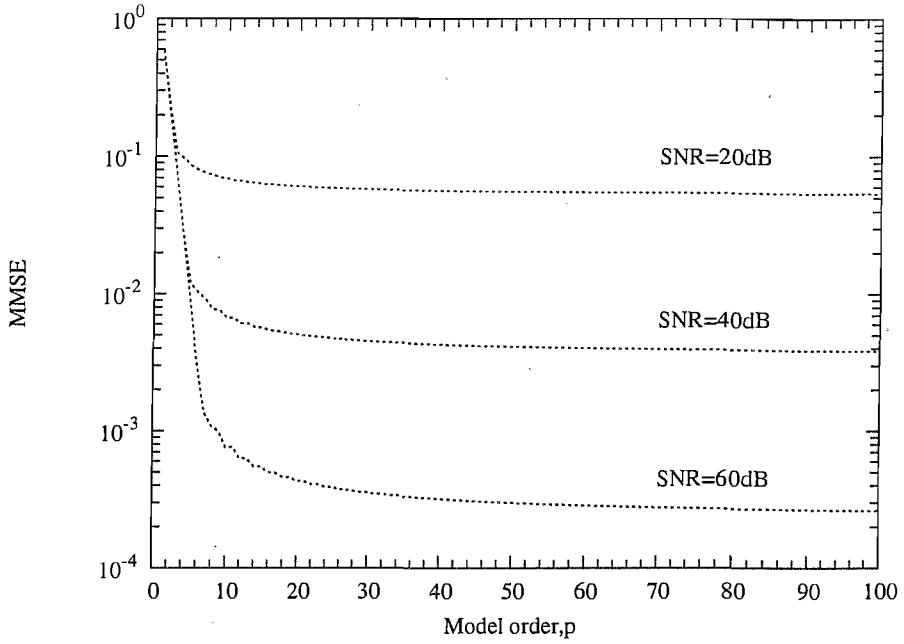


Figure 3.5: MMSE versus predictor order preformance for different values of the SNR, when assuming a sampling rate of $f_s = 500Hz$, and a prediction range of $\tau = 2ms$. The remaining parameters are assumed to be the same as in Table 3.1.

was infinity, the order of the long-range predictor was $p = 50$, the sampling frequency was $f_s = 5f_{dm}$ and the prediction range was $\tau = 0.002s$. All other parameters are shown in Table 3.1. We observe from the results of Figure 3.6 that the values obtained using the long-range predictor closely agree with the actual channel envelope, provided that the SNR value is sufficiently high. In practice, since the SNR value cannot be extremely high, it can be expected that the predicted value generated using LRP may be slightly different from the actual value of the channel envelope.

3.3 Kalman Filter Assisted Long-Range Prediction

In the last section the MMSE-assisted long-range predictor used only p past observations in order to predict the future complex-valued channel coefficient. However, in practice all the past observations are available, provided that predictor has sufficient memory. Consequently, if we can use all the past observation samples for prediction, it can be expected that a more precise prediction of the future channel profile can be achieved. It is well-known the Kalman filters are capable of exploiting all the available observed samples in order to predict the future states. Based on the Kalman filtering principles, prediction is carried out recursively, which usually results in low-complexity prediction.

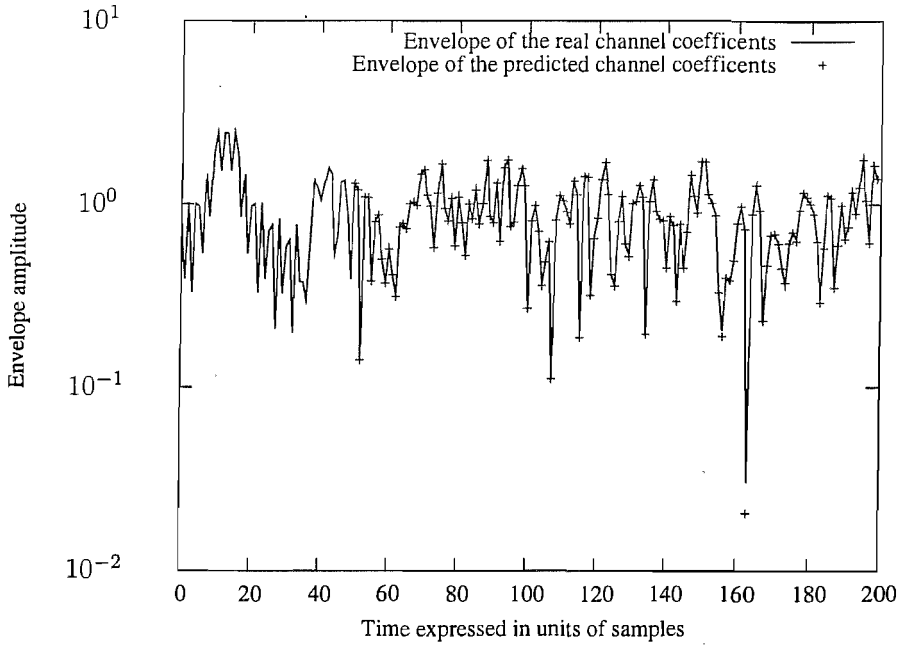


Figure 3.6: Long-range prediction of the envelope of a Rayleigh fading channel profile generated by Jakes' model associated with 9 oscillators when assuming that the SNR was infinity, the order of the predictor was $p = 50$, the sampling frequency was $f_s = 500Hz$ and the prediction range was $\tau = 0.002s$. All other parameters are assumed to the same as in Table 3.1

Hence, in this section we extend the LRP philosophy discussed in the previous section by invoking the classic Kalman filtering principles [1–3].

3.3.1 Kalman Filtering Principles [1–3]

According to the classic principles of Kalman filtering [1–3], the discrete vector-based Kalman filter has a process equation, which can be expressed as [1]

$$\mathbf{c}_n = \mathbf{F}_{n-1}\mathbf{c}_{n-1} + \mathbf{w}_n, \tag{3.25}$$

where the p-dimensional vector $\mathbf{c}_n$ represents the state vector, the $(p \times p)$-dimensional matrix $\mathbf{F}_{n-1}$ is the transition matrix, while the p-dimensional vector $\mathbf{w}_n$ represents the process noise vector, which is modeled as a white sequence, where all members of the set of sequences have zero means and are

mutually uncorrelated with each other [2]. The correlation matirx of $\mathbf{w}_n$ is defined by

$$E[\mathbf{w}_n \mathbf{w}_k^H] = \begin{cases} \mathbf{Q}_{\mathbf{w}_n}, & n = k \\ \mathbf{0}, & n \neq k\,. \end{cases} \tag{3.26}$$

The measurement equation for the discrete vector-based Kalman filter is assumed to have the form of

$$\mathbf{r}_n = \mathbf{H}_n \mathbf{c}_n + \mathbf{v}_n, \tag{3.27}$$

where the N-dimensional vector $\mathbf{r}_n$ represents the observation vector, $\mathbf{H}_n$ is a known $(N \times p)$-dimensional measurement matirx, while the N-dimensional vector $\mathbf{v}_n$ is the measurement noise vector. The measurement noise vector is modeled as a zero-mean, white-noise process having correlation matrix of

$$E[\mathbf{v}_n \mathbf{v}_k^H] = \begin{cases} \mathbf{Q}_{\mathbf{v}_n}, & n = k \\ \mathbf{0}, & n \neq k\,. \end{cases} \tag{3.28}$$

Furthermore, it is assumed that the process noise vector $\mathbf{w}_n$ and the measurement noise vector $\mathbf{v}_n$ are statistically independent, yielding

$$E[\mathbf{w}_n \mathbf{v}_k^H] = 0, \quad \text{for all } n \text{ and } k. \tag{3.29}$$

Let the innovation process α_n be defined as

$$\alpha_n = \mathbf{r}_n - \hat{\mathbf{r}}_{[n|n-1]}, \quad n = 1, 2, \cdots, \tag{3.30}$$

where the vector $\hat{\mathbf{r}}_{[n|n-1]}$ denotes the MMSE prediction of the observed data $\mathbf{r}_n$ at time n, based on all the past observed data. According to (3.25), the process equation can be written as

$$\begin{aligned} \mathbf{c}_k &= \mathbf{F}_{k-1}\mathbf{F}_{k-2}\mathbf{F}_{k-3}\cdots\mathbf{F}_0\mathbf{c}_0 + \mathbf{F}_{k-1}\mathbf{F}_{k-2}\mathbf{F}_{k-3}\cdots\mathbf{F}_1\mathbf{w}_1 \\ &\quad + \mathbf{F}_{k-1}\mathbf{F}_{k-2}\mathbf{F}_{k-3}\cdots\mathbf{F}_2\mathbf{w}_2 + \cdots + \mathbf{F}_{k-1}\mathbf{w}_{k-1} + \mathbf{w}_k, \end{aligned} \tag{3.31}$$

which shows that $\mathbf{c}_k$ is a linear combination of the initial state of $\mathbf{c}_0$ and the noise vectors of $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_k$. Additionally, it is assumed in our model that $\mathbf{c}_0$ is uncorrelated with both $\mathbf{w}_n$ and $\mathbf{v}_n$ for $n \geq 0$.

Since the measurement noise vector $\mathbf{v}_n$ is uncorrelated with both the initial state of $\mathbf{c}_0$ and the

process noise vector $\mathbf{w}_n$, according to (3.31), we obtain that

$$E[\mathbf{c}_k \mathbf{v}_n^H] = \mathbf{0}, \quad k, n \geq 0. \tag{3.32}$$

Additionally, from (3.27), we have

$$E[\mathbf{r}_k \mathbf{v}_n^H] = \mathbf{0}, \quad 0 \leq k \leq n-1 \tag{3.33}$$

and

$$E[\mathbf{r}_k \mathbf{w}_n^H] = \mathbf{0}, \quad 0 \leq k \leq n-1. \tag{3.34}$$

It can be seen from (3.27) that the MMSE-based prediction for the present observation of $\mathbf{r}_n$ can be expressed as

$$\hat{\mathbf{r}}_{[n|n-1]} = \mathbf{H}_n \hat{\mathbf{c}}_{[n|n-1]} + \hat{\mathbf{v}}_{[n|n-1]}, \tag{3.35}$$

where $\hat{\mathbf{c}}_{[n|n-1]}$ represents the MMSE prediction of $\mathbf{c}_n$ and $\hat{\mathbf{v}}_{[n|n-1]}$ represents the MMSE prediction of $\mathbf{v}_n$, respevtively, based on the observations $\mathbf{r}_1, \cdots, \mathbf{r}_{n-1}$. According to (3.33), $\mathbf{v}_n$ is orthogonal to the past observations $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{n-1}$, which results in $\hat{\mathbf{v}}_{[n|n-1]} = 0$. Consequently, (3.35) is reduced to

$$\hat{\mathbf{r}}_{[n|n-1]} = \mathbf{H}_n \hat{\mathbf{c}}_{[n|n-1]}. \tag{3.36}$$

Upon substituting (3.36) and (3.27) into (3.30), the innovation vector $\boldsymbol{\alpha}_n$ can be expressed as [1]

$$\begin{aligned} \boldsymbol{\alpha}_n &= \mathbf{r}_n - \mathbf{H}_n \hat{\mathbf{c}}_{[n|n-1]} \\ &= \mathbf{H}_n \boldsymbol{\varepsilon}_n + \mathbf{v}_n, \end{aligned} \tag{3.37}$$

where

$$\boldsymbol{\varepsilon}_n = \mathbf{c}_n - \hat{\mathbf{c}}_{[n|n-1]}, \tag{3.38}$$

which represents the error between the state vector $\mathbf{c}_n$ and the predicted state vector $\hat{\mathbf{c}}_{[n|n-1]}$. Furthermore, it can be shown that $\boldsymbol{\varepsilon}_n$ is orthogonal to both $\mathbf{w}_{n+1}$ and $\mathbf{v}_n$.

With the aid of classic Kalman filtering theory [1], the state vector $\hat{\mathbf{c}}_{[n+1|n]}$ can be predicted according to

$$\begin{aligned}\hat{\mathbf{c}}_{[n+1|n]} &= \sum_{k=1}^{n} E[\mathbf{c}_{n+1}\boldsymbol{\alpha}_k^H]\mathbf{R}_k^{-1}\boldsymbol{\alpha}_k \\ &= \sum_{k=1}^{n-1} E[\mathbf{c}_{n+1}\boldsymbol{\alpha}_k^H]\mathbf{R}_k^{-1}\boldsymbol{\alpha}_k + E[\mathbf{c}_{n+1}\boldsymbol{\alpha}_n^H]\mathbf{R}_n^{-1}\boldsymbol{\alpha}_n, \end{aligned} \tag{3.39}$$

where $\mathbf{R}_k$ is the autocorrelation matrix of the innovation process $\boldsymbol{\alpha}_k$, which is given by

$$\mathbf{R}_k = E[\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^H]. \tag{3.40}$$

Upon substituting (3.37) into (3.40), we obtain

$$\mathbf{R}_k = \mathbf{H}_k\mathbf{K}_k\mathbf{H}_k^H + \mathbf{Q}_{\mathbf{v}_k}, \tag{3.41}$$

where

$$\mathbf{K}_k = E[\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^H], \tag{3.42}$$

which represents the autocorrelation matrix of the predicted state error vector $\boldsymbol{\varepsilon}_k$.

From (3.34) we know that $\boldsymbol{\alpha}_k$ is orthogonal to $\mathbf{w}_{n+1}$ for $0 \le k \le n$. Then, we can obtain

$$\begin{aligned} E[\mathbf{c}_{n+1}\boldsymbol{\alpha}_k^H] &= E\{[\mathbf{F}_n\mathbf{c}_n + \mathbf{w}_{n+1}]\boldsymbol{\alpha}_k^H\} \\ &= \mathbf{F}_n E[\mathbf{c}_n\boldsymbol{\alpha}_k^H], \quad 0 \le k \le n. \end{aligned} \tag{3.43}$$

Upon substituting (3.43) into (3.39), $\hat{\mathbf{c}}_{[n+1|n]}$ can be obtained from $\hat{\mathbf{c}}_{[n|n-1]}$, with the recursive equation expressed as

$$\hat{\mathbf{c}}_{[n+1|n]} = \mathbf{F}_n\hat{\mathbf{c}}_{[n|n-1]} + \mathbf{G}_n\boldsymbol{\alpha}_n, \tag{3.44}$$

where $\mathbf{G}_n$ is the Kalman gain, which is given by

$$\begin{aligned}\mathbf{G}_n &= E[\mathbf{c}_{n+1}\boldsymbol{\alpha}_n^H]\mathbf{R}_n^{-1} \\ &= \mathbf{F}_n\mathbf{K}_n\mathbf{H}_n^H\mathbf{R}_n^{-1}, \end{aligned} \tag{3.45}$$

while $\mathbf{K}_{n+1}$ is given by

$$\begin{aligned}\mathbf{K}_{n+1} &= E[\boldsymbol{\varepsilon}_{n+1}\boldsymbol{\varepsilon}_{n+1}^H] \\ &= \mathbf{F}_n\mathbf{M}_n\mathbf{F}_n^H + \mathbf{Q}_{w_{n+1}},\end{aligned} \tag{3.46}$$

which is referred to as the Riccati difference equation [1]. Furthermore, in (3.46) $\mathbf{M}_n$ is given by [2],

$$\begin{aligned}\mathbf{M}_n &= E[\mathbf{e}_n\mathbf{e}_n^H] \\ &= E[(\mathbf{c}_n - \hat{\mathbf{c}}_{[n|n]})(\mathbf{c}_n - \hat{\mathbf{c}}_{[n|n]})^H] \\ &= \mathbf{K}_n - \mathbf{K}_n\mathbf{H}_n^H(\mathbf{H}_n\mathbf{K}_n\mathbf{H}_n^H + \mathbf{Q}_{\mathbf{v}_n})^{-1}\mathbf{H}_n\mathbf{K}_n,\end{aligned} \tag{3.47}$$

which represents the autocorrelation matrix of the estimation error vector $\mathbf{e}_n$.

3.3.2 Initialization of the Kalman Filter

The initialization described in [1] will be used in this chapter. Specifically, $\hat{\mathbf{c}}_{[0|-1]}$ and $\mathbf{K}_0$ can be initialized as

$$\hat{\mathbf{c}}_{[0|-1]} = E[\mathbf{c}_0], \tag{3.48}$$

$$\mathbf{K}_0 = E[(\mathbf{c}_0 - E[\mathbf{c}_0])(\mathbf{c}_0 - E[\mathbf{c}_0])^H], \tag{3.49}$$

respectively. Since $\mathbf{c}_n$ has a zero mean, we have

$$\hat{\mathbf{c}}_{[0|-1]} = E[\mathbf{c}_0] = \mathbf{0} \tag{3.50}$$

$$\begin{aligned}\mathbf{K}_0 &= E[(\mathbf{c}_0 - E[\mathbf{c}_0])(\mathbf{c}_0 - E[\mathbf{c}_0])^H] \\ &= E[\mathbf{c}_0\mathbf{c}_0^H].\end{aligned} \tag{3.51}$$

In summary, when using the classic Kalman filter, $\mathbf{F}_n$, $\mathbf{H}_n$, $\mathbf{Q}_{w_n}$ and $\mathbf{Q}_{v_n}$ have to be known in advance, while $\hat{\mathbf{c}}_{[0|-1]}$ and $\mathbf{K}_0$ are initialized according to (3.48) and (3.49), respectively. Then, $\mathbf{R}_n$, $\mathbf{G}_n$, $\boldsymbol{\alpha}_n$ and $\hat{\mathbf{c}}_{[n+1|n]}$ can be computed according to (3.41), (3.45), (3.37) and (3.44), while $\mathbf{K}_{n+1}$ and $\mathbf{M}_n$ are updated according to (3.46) and (3.47), respectively.

Having discussed the classic Kalman filtering principles, let us now investigate the application of Kalman filtering for the LRP of wireless channels.

3.3.3 Application of the Kalman Filter for One-Step Long-Range Prediction

As we have described in Section 2.5.3, the channel's complex-valued coefficients can be expressed as in (2.32), which is repeated here for convenience,

$$c_n = \sum_{k=1}^{p} a_k c_{n-k} + w_n, \tag{3.52}$$

where $[a_1, a_2, \cdots, a_p]$ and $\sigma_{w_n}^2$ are determined by (2.35) and (2.38), respectively. In order to invoke the vector-based Kalman filter for the LRP, we define a p-dimensional state vector $\mathbf{c}_n$ at the time instant $t = nT_s$ as

$$\mathbf{c}_n = [c_n, c_{n-1}, \cdots, c_{n-p+1}]^T. \tag{3.53}$$

Hence, we have

$$\mathbf{c}_n = \mathbf{F}_{n-1}\mathbf{c}_{n-1} + \mathbf{w}_n, \tag{3.54}$$

where $\mathbf{F}_{n-1}$ is a $(p \times p)$-dimensional transition matrix, which is given by

$$\mathbf{F}_{n-1} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{p-1} & a_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \tag{3.55}$$

while $\mathbf{w}_n$ is a p-dimensional process noise vector, which is expressed as

$$\mathbf{w}_n = [w_n, 0, \cdots, 0]^T. \tag{3.56}$$

Let the $(1 \times p)$-dimensional measurement matrix $\mathbf{H}_n$ be defined as

$$\mathbf{H}_n = [b_n, 0, 0, \cdots, 0], \tag{3.57}$$

where b_n is the data symbol after modulation. Then, according to (3.3) , the measurement equation can be written as

$$r_n = \mathbf{H}_n \mathbf{c}_n + n_n. \tag{3.58}$$

Based on the above arguments and according to the Kalman filtering principles described by (3.45), (3.46), (3.47), $\hat{\mathbf{c}}_{[n+1|n]}$ can be obtained recursively by

$$\hat{\mathbf{c}}_{[n+1|n]} = \mathbf{F}_n \hat{\mathbf{c}}_{[n|n-1]} + \mathbf{G}_n \alpha_n, \tag{3.59}$$

where

$$\mathbf{G}_n = \mathbf{F}_n \mathbf{K}_n \mathbf{H}_n^H \mathbf{R}_n^{-1}, \tag{3.60}$$

$$\mathbf{K}_{n+1} = \mathbf{F}_n \mathbf{M}_n \mathbf{F}_n^H + \mathbf{Q}_{w_{n+1}}, \tag{3.61}$$

$$\mathbf{M}_n = \mathbf{K}_n - \mathbf{K}_n \mathbf{H}_n^H (\mathbf{H}_n \mathbf{K}_n \mathbf{H}_n^H + \mathbf{Q}_{v_n})^{-1} \mathbf{H}_n \mathbf{K}_n, \tag{3.62}$$

where the first element in $\hat{\mathbf{c}}_{[n+1|n]}$, is the prediction of c_{n+1} based on $r_1, \cdots, r_n$, while the first element in $\mathbf{K}_{n+1}$ is the corresponding MMSE, when c_{n+1} is predicted by the observations $r_1, \cdots, r_n$. Since c_{n+1} is predicted based on the observations until the instant n, i.e. based on $r_1, \cdots, r_n$, the corresponding prediction is the simplest possible one-step prediction. Let us now consider our LRP problem in conjunction with an arbitrary prediction range.

3.3.4 Application of the Kalman Filter for Long-Range Prediction

In order to predict the channel's complex-valued coefficients for an arbitrary time interval τ, $\tau > 0$ ahead of the available observations, (3.62) is modified as

$$c_{n+\tau} = \sum_{k=0}^{p-1} a_{\tau k} c_{n-k} + w_{n+\tau}, \quad \tau > 0 \tag{3.63}$$

and a p-dimensional AR coefficient vector $\mathbf{a}_\tau$ is defined as

$$\mathbf{a}_\tau = [a_{\tau 0}, a_{\tau 1}, \cdots, a_{\tau(p-1)}]^T, \tag{3.64}$$

which can be expressed as [78]

$$\mathbf{a}_\tau = \mathbf{R}_{cc\tau}^{-1}\mathbf{r}_{c\tau}, \tag{3.65}$$

where $\mathbf{R}_{cc\tau}$ is the $(p \times p)$-dimensional autocorrelation matrix of the state vector $\mathbf{c}_n$. The matrix $\mathbf{R}_{cc\tau}$ is given by

$$\mathbf{R}_{cc\tau} = E[\mathbf{c}_n^* \mathbf{c}_n^T] = \begin{bmatrix} r(0) & r^*(1) & \cdots & r^*(p-1) \\ r(1) & r(0) & \cdots & r^*(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix}, \tag{3.66}$$

while $\mathbf{r}_{c\tau}$ is the p-dimensional cross correlation vector of the desired channel coefficient c_τ and of the state vector $\mathbf{c}_n$, which is given by

$$\mathbf{r}_{c\tau} = E[c_{n+\tau}\mathbf{c}_n^*] = [r(\tau), r(1+\tau), \cdots, r(p-1+\tau)]^T, \tag{3.67}$$

associated with $r(l) = E[c_n c_{n-l}^*]$, $r^*(l) = E[c_n c_{n-l}^*]^*$. Accordingly, the variance of the noise $w_{n+\tau}$ in (3.63) is given by

$$\sigma_{w_{n+\tau}}^2 = r(0) - \sum_{k=0}^{p-1} a_{\tau k} r(-k-\tau). \tag{3.68}$$

Let a $(p \times 1)$-dimensional state vector $\mathbf{c}_{n+\tau}$ be defined as

$$\mathbf{c}_{n+\tau} = [c_{n+\tau}, c_n, \cdots, c_{n-p+2}]^T. \tag{3.69}$$

Then, the process equation of the Kalman filter can be expressed as

$$\mathbf{c}_{n+\tau} = \mathbf{F}_{n\tau}\mathbf{c}_n + \mathbf{w}_{n+\tau}, \tag{3.70}$$

where $\mathrm{F}_{n\tau}$ is a $(p \times p)$-dimensional transition matrix, which can be expressed as

$$\mathbf{F}_{n\tau} = \begin{bmatrix} a_{\tau 0} & a_{\tau 1} & a_{\tau 2} & \cdots & a_{\tau(p-2)} & a_{\tau(p-1)} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \tag{3.71}$$

while $\mathbf{w}_{n+\tau}$ is a p-dimensional process noise vector, which is expressed as

$$\mathbf{w}_{n+\tau} = [w_{n+\tau}, 0, \cdots, 0]^T. \tag{3.72}$$

After $\hat{\mathbf{c}}_{[n|n-1]}$ is obtained according to (3.59), $\hat{\mathbf{c}}_{[n+\tau|n]}$ can be expressed as

$$\hat{\mathbf{c}}_{[n+\tau|n]} = \mathbf{F}_{n\tau}\hat{\mathbf{c}}_{[n|n-1]} + \mathbf{G}_{n\tau}\boldsymbol{\alpha}_n, \tag{3.73}$$

where the first element in $\hat{\mathbf{c}}_{[n+\tau|n]}$, is the prediction of $c_{n+\tau}$ based on $r_1, \cdots, r_n$ and $\mathbf{G}_{n\tau}$ is given by

$$\mathbf{G}_{n\tau} = \mathbf{F}_{n\tau}\mathbf{K}_n\mathbf{H}_n^H\mathbf{R}_n^{-1}. \tag{3.74}$$

Furthermore, we have

$$\mathbf{K}_{n+\tau} = \mathbf{F}_{n\tau}\mathbf{M}_n\mathbf{F}_{n\tau}^H + \mathbf{Q}_{w_{n+\tau}}, \tag{3.75}$$

where the first element in $\mathbf{K}_{n+\tau}$ is the corresponding MMSE. Note that when $\tau = T_s$, the LRP derived for an arbitrary prediction interval is the same as the one-step LRP.

In the next section, various examples are provided in order to show the characteristics of the LRP invoking the classic Kalman filtering principles.

3.3.5 Performance Analysis of the Kalman Filter Assisted Long-Range Prediction

In this section, the simulation results are given to illustrate the performance of the Kalmang filter assisted LRP. Furthermore, pilot symbol assisted modulation (PSAM) [74] is invoked in this section and $b_n = 1$ is assumed to be the pilot symbol, which is inserted according to the sampling frequency of f_s as shown in Figure 3.7 where pilot symbol is inserted every L symbols. Consequently, the

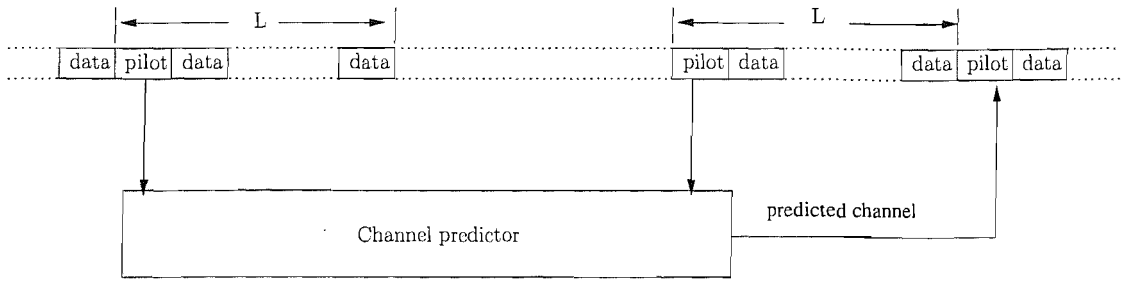


Figure 3.7: Pilot-assisted flat fading channel prediction

Maximum Doppler frequency f_{dm}	100Hz
Sampling rate f_s	500Hz
Data symbol rate f_b	25KHz
Prediction range τ	0.002s

Table 3.2: Parameters for Kalman filter assisted long range prediction

measurement matrix $\mathbf{H}_n$ in (3.57) can be expressed as

$$\mathbf{H}_n = [1, 0, \cdots, 0]. \tag{3.76}$$

In Figure 3.8, the achievable MMSE performance is evaluated versus the order p of the predictor for the Kalman filter assisted LRP with respect to different SNR values, when the 500th sample is predicted. The remaining parameters were summarized in Table 3.2. Furthermore, the prediction range of $\tau = 0.002s$ corresponds to one-step prediction in this case. As we can see from the results of Figure 3.8, for a given SNR value, initially the MMSE value decreases rapidly and near-linearly upon increasing the predictor's order p, before it reaches a MMSE floor for a moderate order p. For predictor orders in excess of this value the MMSE remains near-constant, regardless of further increasing the predictor's order p. Since the order p is related to the complexity of the prediction and the complexity is usually on the order of $O(p^3)$ for the Kalman filter based prediction, the minimum value of p used for Kalman filter assisted LRP may be found based on the specific combination of the remaining system parameters, such as f_{dm}, τ, SNR, etc. Additionally, we can see from Figure 3.8 for high SNR values, usually a Kalman filter having a slightly higher order p is required for reaching the the MMSE floor-value. This is because for higher SNRs, the precision of the AR model of the channel dominates the performance of preditor, while for lower SNRs the measurement noise dominates the

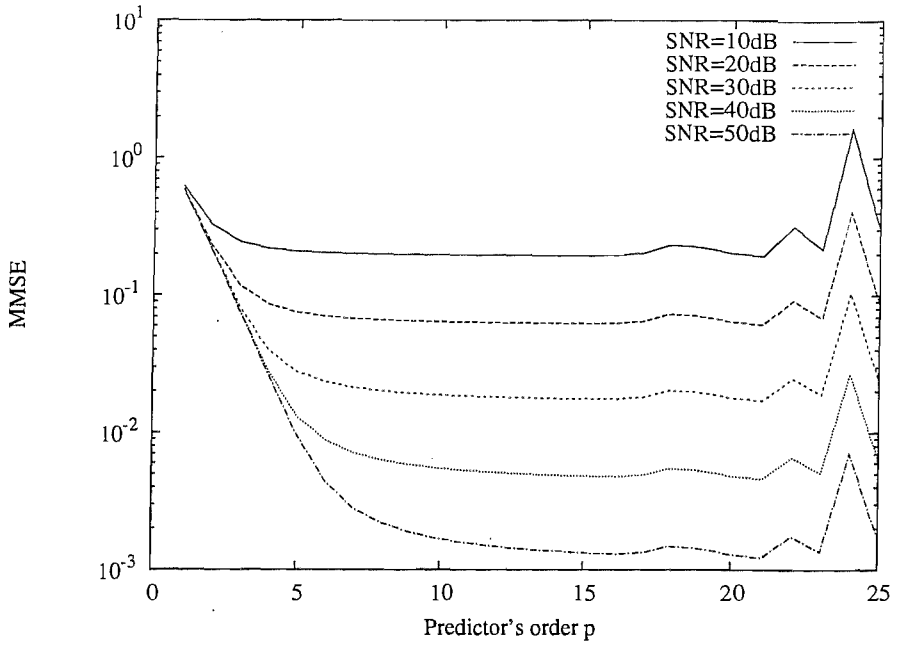


Figure 3.8: MMSE versus the predictor order p performance for the Kalman fiter assisted LRP with respect to different SNR values, when the 500th sample is predicted. The remaining parameters are summarized in Table 3.2.

performance of the predictor instead of the predictor's order p. Furthermore, when the SNR increases, the MMSE recorded for a specific predictor order is reduced, because the influence of measurement noise diminishes. Addtionally, when the predictor order p is higher than 15, there are some spikes in the MMSE curves of Figure 3.8. This is because the matrix $\mathbf{R}_{cc}$ used for calculating the AR-model's coefficients in (2.35) is ill-conditioned for these preditor orders and hence the results become unreliable [78, 79].

In Figure 3.9 we evaluate the convergence of Kalman filtering assissted LRP in terms of its MMSE versus the number pilots used for LRP for different SNR values, when the preditor's order is $p = 15$. The remaining parameters are assumed to be the same as in Table 3.2. We observe from the results of Figure 3.9 that for a given SNR value, the MMSE initially decreases near-linearly upon increasing the number of the pilots and finally reaches a constant MMSE floor value. Futhermore, Kalman filter assisted LRP converges faster for higher SNR values because the influence of the measurement noise is less dominant. Additionally, for a given number of pilots, the MMSE decreases upon increasing the SNR.

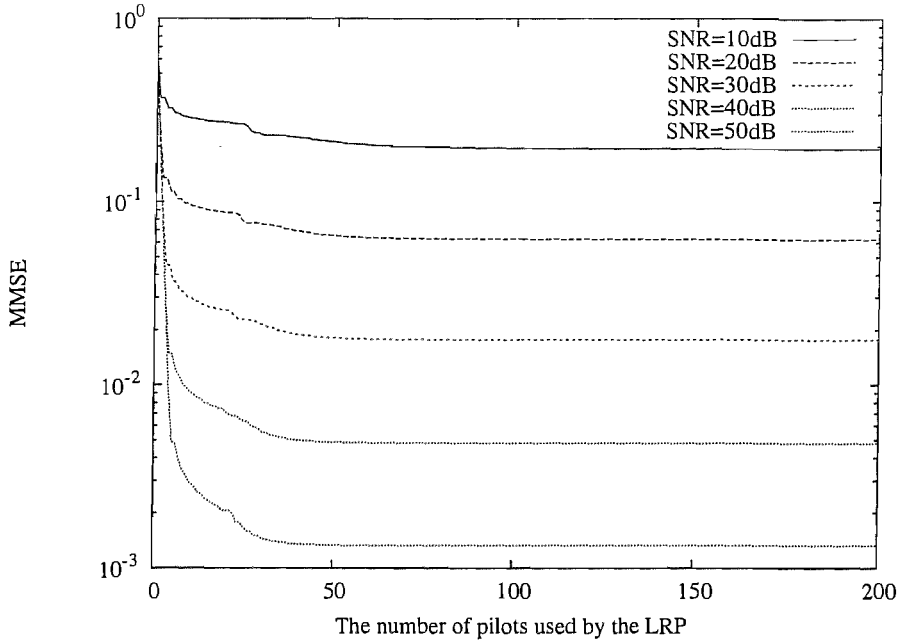


Figure 3.9: MMSE performance versus the number of pilots used by the LRP at different SNR values when assiming that the order of the predictor is $p = 15$, while the remaining parameters are assumed to be the same as in Table 3.2.

3.4 Conclusions

In Section 3.2, MMSE assisted LRP designed for narrowband fading channels has been discussed. From the results of Figures 3.4 and 3.3 we conclude that the best MMSE performance may be achieved by choosing an appropriate sampling rate f_s, which can be significantly lower than the data rate f_b. In order to further improve the achievable performance, in Section 3.3 more sophisticated Kalman filtering assisted LRP was proposed and investigated.

Chapter 4

Long-Range Linear Prediction of Wideband Fading Channels

4.1 Introduction

In the previous chapter, we have mainly discussed LRP in the context of narrowband fading channels. However, high-rate state-of-the-art wireless communications typically experience frequency-selective wideband fading rather than flat fading. In the context of wideband channel conditions, the system suffers from intersymbol interference ISI [81]. In order to cope with this problem, multicarrier transmission systems have been proposed, such as Orthogonal Frequency-Division Multiplexing (OFDM) [5], which is capable of transforming the frequency-selective fading channel into to numerous low rate parallel flat fading channels, hence avoiding the ISI. However, a specific drawback of OFDM is its high peak-to-average power ratio (PAPR). As a design alternative, single carrier frequency domain equalization (SC-FDE) has been proposed as a couterpart of OFDM, which promises a similar performance to that of OFDM, but without imposing a high PAPR [86, 87].

Recently, pre-equalization applied in the frequency domain has attracted wide attention [88–90], which is capable of simplifying the design of the receiver. Furthermore, advanced techniques such as adaptive modulation and precoding have also been investigated in the context of single carrier wideband systems [4, 91–94]. However, both pre-equalization in the frequency domain and adaptive modulation require the accurate channel state information (CSI) at the transmitter side. As a design option, the receiver can estimate the CSI and feed it back to the transmitter. However, due to the inherent delay, the estimated CSI becomes outdated at transmitter. Altervaltively, channel prediction

can be used for solving this problem. Therefore, it is important to investigate wideband channel prediction designed for supporting sophisticated transmitter preprocessing techniques, such as adaptive modulation [4–6], pre-Rake processing [95], multiuser transmission techniques [96].

In [20] the ESPRIT algorithm has been employed for wideband channel prediction. Since the wideband channel transfer function (CTF) tends to be highly correlated, in [20] the CTFs were determined during an observation interval and were then used by the ESPRIT algorithm in order to keep track of the CTF's poles. Once the poles of the CTF have been estimated, the corresponding complex CIR-tap gains can be determined by solving a set of linear equations and can also be used for predicting their future values. In contrast to [20], in [21,22] both the 1-D and 2-D Unitary-ESPRIT algorithm assisted by a so-called chirp-signal estimation scheme have been employed for estimating the CTF's parameters. Once these parameters have been determined, the CTF can be predicted both in the time-domain and frequency-domain in order to provide future estimates. With this motivation, in this chapter we investige the LRP of wideband channels.

The rest of this chapter has the following structure. The wideband channel is described in Section 4.2, while classic single carrier data transmission is discussed in Section 4.3. In Section 4.4 the philosophy of two-dimensional (2-D) channel estimation applied in a single carrier system is demonstrated, while a LRP aided single carrier wideband system is proposed in Section 4.5. Our simulation results are provided in Section 4.6. Finally, our conclusions are offered in Section 4.7.

4.2 Description of the Wideband Channel

The received signal of wideband wireless communication systems can be expressed as [97,98]

$$r(t) = \sum_{k=-\infty}^{+\infty} b_k c(t; t - kT_b) + z(t), \tag{4.1}$$

where b_k represents the discrete-time transmitted signal, T_b is the symbol period and $z(t)$ is the AWGN contaminating the received signal. In (4.1) $c(t;\tau)$ represents the combined channel impulse response (CIR), which can be expressed as

$$c(t;\tau) = g(t) \otimes g(t;\tau) \otimes g^*(-t), \tag{4.2}$$

where $g(t)$ and $g^*(-t)$ represent the pulse shaping filter at the transmitter and the corresponding matched filter at the receiver, respectively. Furthermore $g(t;\tau)$ represents the CIR of the time-varying

frequency-selective fading channel, which can be expressed as [43]

$$g(t;\tau) = \sum_{l=-\infty}^{+\infty} \alpha_l(t)\delta(\tau - \tau_l), \tag{4.3}$$

where l is the multipath component index, while τ_l and $\alpha_l(t)$ represent the delay and the complex channel gain of the lth path, respectively. Furthermore, in (4.2) the symbol $\otimes$ denotes the convolution operation.

When $r(t)$ of (4.1) is sampled at the symbol rate, the output of the matched filter can be expressed as

$$\begin{aligned} r_n \equiv r(t)|_{t=nT_b} &= \sum_{k=-\infty}^{+\infty} b_k c(nT_b; nT_b - kT_b) + z(nT_b) \\ &= \sum_{k=-\infty}^{+\infty} b_k c_{(n;n-k)} + z_n \\ &= \sum_{k=-\infty}^{+\infty} c_{(n;k)} b_{(n-k)} + z_n. \end{aligned} \tag{4.4}$$

As shown in (4.4), the impulse response $c_{(n;k)}$ generally extends to infinity. However, in practical communications applications it is common to truncate it at some order L, yielding the discrete-time model, which can be expressed as

$$r_n = \sum_{k=0}^{L-1} c_{(n;k)} b_{(n-k)} + z_n. \tag{4.5}$$

Let us now consider channel estimation and prediction in the context of a single-carrier wideband system.

4.3 Single-Carrier Block-Based Data Transmission

In wideband frequency-selective fading channels, different frequencies may experience different fading, if the frequency band is wider than the coherence bandwidth of the channel. By contrast, when the transmission bandwidth is signifficiantly lower than the coherence bandwidth of the channel, all frequencies will experience similar fading. An attractive technique of estimating or predicting wideband fading channels - in particular, if their CIR is long - is to transfer the received signal from the time-domain to the frequency-domain and then estimate or predict the channel in the frequency do-

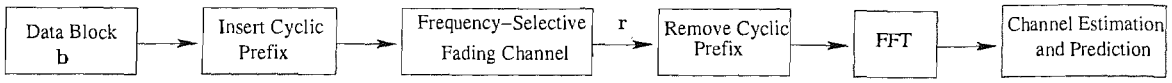


Figure 4.1: Schematic of single-carrier block-based data transmission

main. Specifically, the total frequency band can be divided into a number of subbands having a low bandwidth. Consequently, the signals associcated with each subband experience flat fading. In this case, the channel corresponding to each of the subchannels can be estimated or predicted by invoking various algorithms designed for flat fading channels. Each subchannel can be predicted, for example, using the LRP approaches that have been investigated in Chapter 3.

In [86,87,99–101], a single-carrier block-based transmission scheme designed for multipth fading channels has been discussed, which was shown in Fig.4.1 . In the context of block-based data transmissions, each block is constituted by N useful data symbols, namely by $[b_{(k;0)}, b_{(k;1)}, \cdots, b_{(k;N-1)}]$ for the kth block. In order to eliminate the interblock interference (IBI), each transmission block is also quasi-periodically extended by a length-u cyclic prefix repeating the last u symbols of the transmission block, where the value of u is determined by the CIR duration. Specifically, if the CIR is modeled by an Lth-order filter, then u should satisfy $u \geq L-1$ [99]. Consequently, a data block is comprised of $(N+u)$ data symbols expressed as $[b_{(k;N-u)}, \cdots, b_{(k;N-1)}, b_{(k;0)}, b_{(k;2)}, \cdots, b_{(k;N-1)}]$. Let T_B be the duration of a data block and T_b be the duration of a data symbol. Then we have

$$T_B = (N+u)T_b. \tag{4.6}$$

We assume that the CIR taps of the channel concerned remain constant during the transmission of a data block [86, 87, 99–101]. Furthermore, we assume that for the kth data block, the CIR vector $\mathbf{c}_k$ can be expressed as

$$\mathbf{c}_k = [c_{(k;0)}, c_{(k;1)}, \cdots, c_{(k;L-1)}]^T, \tag{4.7}$$

where

$$c_{(k;l)} = c(kT_B; lT_b) \tag{4.8}$$

represents the channel's amplitude in correspondence with the lth data symbol within the kth data block. Hence, according to (4.5) and (4.8), the received signal's samples observed during the kth data

block can be expressed as

$$r_{(k;n)} = \sum_{l=0}^{L-1} c_{(k;l)} b_{(k;n-l)} + z_{(k;n)}, \quad n = 0, 1, \cdots, N-1, \cdots, N+u-1. \tag{4.9}$$

Having received $(N+u)$ number of observed samples corresponding to a data block, the first u samples are discarded, in order to eliminate the IBI. Consequenctly, following the removal of the cyclic prefix the N received samples of the kth data block can be expressed as

$$\mathbf{r}_k = \mathbf{C}_k \mathbf{b}_k + \mathbf{z}_k, \tag{4.10}$$

where $\mathbf{r}_k$ consists of the N observation samples, which is expressed as

$$\mathbf{r}_k = [r_{(k;0)}, r_{(k;1)}, \cdots, r_{(k;N-1)}]^T, \tag{4.11}$$

$\mathbf{z}_k$ is the N-dimensional noise vector, which is expressed as

$$\mathbf{z}_k = [z_{(k;0)}, z_{(k;1)}, \cdots, z_{(k;N-1)}]^T, \tag{4.12}$$

while $\mathbf{b}_k$ is a N-dimensional vector containing the N number of transmitted data symbols, which is denoted by

$$\mathbf{b}_k = [b_{(k;0)}, b_{(k;1)}, \cdots, b_{(k;N-1)}]^T. \tag{4.13}$$

Finally, in (4.10) $\mathbf{C}_k$ is a $(N \times N)$-dimensional circulant matrix with the first coulmn containing CIR

followed by $(N-L)$ zeros [100], where $\mathbf{C}_k$ is given by

$$\mathbf{C}_k = \begin{bmatrix} c_0 & 0 & 0 & 0 & \cdots & c_{L-1} & c_{L-2} & \cdots & \cdots & c_1 \\ c_1 & c_0 & 0 & 0 & \cdots & 0 & c_{L-1} & \cdots & \cdots & c_2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & 0 & 0 & \ddots & \vdots \\ c_{L-2} & \ddots & \ddots & \ddots & c_0 & \ddots & \ddots & \ddots & 0 & c_{L-1} \\ c_{L-1} & c_{L-2} & \cdots & \cdots & \cdots & c_0 & \ddots & \ddots & 0 & 0 \\ 0 & c_{L-1} & \cdots & \cdots & \cdots & \cdots & c_0 & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & c_{L-1} & c_{L-2} & \cdots & \cdots & c_1 & c_0 \end{bmatrix}. \tag{4.14}$$

Since $\mathbf{C}_k$ of (4.14) is a circulant matrix, it can be shown that $\mathbf{C}_k$ can be expressed as [100, 102]

$$\mathbf{C}_k = \mathbf{Q}^H \Lambda_k \mathbf{Q} \tag{4.15}$$

where the superscript $\{\cdot\}^H$ represents the conjugate transpose operation, $\mathbf{Q}$ is the $(N \times N)$-dimensional orthonormal discrete Fourier transform (DFT) matrix and its (m,n)th element of $\mathbf{Q}_{(m,n)}$ is given by

$$\mathbf{Q}_{(m,n)} = \frac{1}{\sqrt{N}} e^{-j2\pi \frac{mn}{N}}, \quad 0 \le m, n \le N-1. \tag{4.16}$$

Furthermore, it can be shown that $\mathbf{Q}$ has the property of

$$\mathbf{Q}^H \mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}. \tag{4.17}$$

In (4.15) Λ_k is a diagonal matrix, whose (n,n)th element is equal to the nth DFT coefficient of $[c_0, c_1, \cdots, c_{L-1}]$, where the DFT can be expressed as

$$\breve{c}_{(k;n)} = \sum_{l=0}^{L-1} c_{(k;l)} e^{-j2\pi \frac{ln}{N}}, \quad 0 \le n \le N-1. \tag{4.18}$$

In fact, the DFT of $[c_{(k;0)}, c_{(k;1)}, \cdots, c_{(k;L-1)}]$ is the discrete CTF. In Figure 4.2, $|\breve{c}_{(k;n)}|$ represents the envelope of the discrete CTF of $\breve{c}_{(k;n)}$, which is ploted for $N = 128$, $L = 10$. Observe in

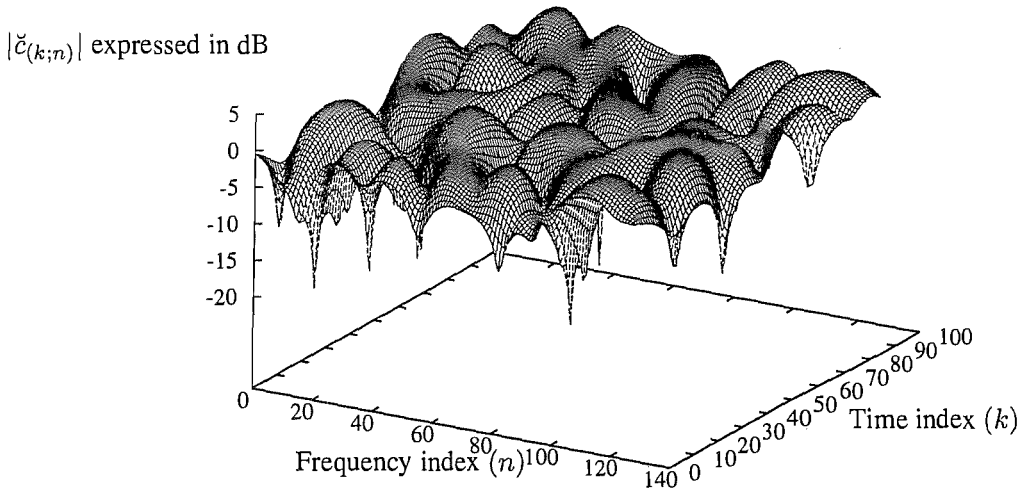


Figure 4.2: Envelope of the Discrete Channel Transfer Function (CTF) for a channel corresponding to $f_{dm} = 120Hz$, data block duration of $T_B = 0.00025s$, $L = 10$, $N = 128$. The power intensity is set to 0.1 for each path

Figure 4.2, that the CTF can be estimated or predicted either in the time-domain or in the frequency-domain, since it exhibits correlation along both axes.

Therefore, upon applying the DFT to the received samples $\mathbf{r}_k$, $\mathbf{r}_k$ can be transformed from the time-domain to the frequency domain. Specifically, after applying the DFT to $\mathbf{r}_k$ of (4.10), we arrive at the frequency domain observation vector, which can be expressed as

$$\breve{\mathbf{r}}_k = \sqrt{N}\mathbf{Q}\mathbf{r}_k. \tag{4.19}$$

When substituting (4.10) and (4.15) into (4.19) and using the property of (4.17), we obtain

$$\begin{aligned} \breve{\mathbf{r}}_k &= \sqrt{N}\mathbf{Q}\mathbf{C}_k\mathbf{b}_k + \sqrt{N}\mathbf{Q}\mathbf{z}_k \\ &= \Lambda_k\breve{\mathbf{b}}_k + \breve{\mathbf{z}}_k, \end{aligned} \tag{4.20}$$

where $\breve{\mathbf{b}}_k = \sqrt{N}\mathbf{Q}\mathbf{b}_k$ represents the DFT of the input symbol vector of $\mathbf{b}_k$ and $\breve{\mathbf{z}}_k = \sqrt{N}\mathbf{Q}\mathbf{z}_k$ is the DFT of the noise vector $\mathbf{z}_k$. If AWGN has a covariance of N_0, then, the covariance matrix of $\mathbf{z}_k$ is given by $N_0\mathbf{I}_N$. Consequently, according to (4.17), the covariance matrix of $\breve{\mathbf{z}}_k$ can be formulated as

$E[\breve{\mathbf{z}}_k \breve{\mathbf{z}}_k^H] = NN_0\mathbf{I}_N$. Hence, the noise vector of $\breve{\mathbf{z}}_k$ in the frequency domain is still a white Gaussian noise vector.

By expanding (4.20), the nth element of $\breve{\mathbf{r}}_k$ can be expressed as

$$\breve{r}_{p(k;n)} = \breve{c}_{(k;n)}\breve{b}_{p(k;n)} + \breve{z}_{p(k;n)}, \quad 0 \leq n \leq N-1, \tag{4.21}$$

where $\breve{r}_{p(k;n)}$ is the nth DFT coefficient of the received observation samples $\mathbf{r}_k = [r_{(k;0)}, r_{(k;1)}, \cdots, r_{(k;N-1)}]^T$, i.e. we have

$$\breve{r}_{p(k;n)} = \sum_{m=0}^{N-1} r_{(k;m)} e^{-j2\pi \frac{mn}{N}}, \quad 0 \leq n \leq N-1, \tag{4.22}$$

$\breve{b}_{p(k;n)}$ in (4.21) is the nth DFT coefficient of the transmitted data symbols of $\mathbf{b}_k = [b_{(k;0)}, b_{(k;1)}, \cdots, b_{(k;N-1)}]^T$, which is expressed as

$$\breve{b}_{p(k;n)} = \sum_{m=0}^{N-1} b_{(k;m)} e^{-j2\pi \frac{mn}{N}}, \quad 0 \leq n \leq N-1. \tag{4.23}$$

Finally, $\breve{z}_{p(k;n)}$ is the nth DFT coefficient of the noise samples $\mathbf{z}_k = [z_{(k;0)}, z_{(k;1)}, \cdots, z_{(k;N-1)}]^T$, which is given by

$$\breve{z}_{p(k;n)} = \sum_{m=0}^{N-1} z_{(k;m)} e^{-j2\pi \frac{mn}{N}}, \quad 0 \leq n \leq N-1. \tag{4.24}$$

Above we have briefly characterized the family of wideband wireless channels, which exhibit time-varying frequency-selective fading. Hence, it is expected that in wideband systems the channel estimation and prediction process becomes significantly more complex than in narrowband systems. Below we investigate the process of channel estimation and LRP in the context of single-carrier wideband systems.

4.4 Estimation of Wideband Channels

Since at the receiver side only noisy channel samples are available, the noise may have a significant impact on the achievable performance of the prediction, if the noisy channel samples are directly used for channel prediction without any preprocessing. In order to mitigate the effects of the noise on the channel prediction, first channel estimation is carried out.

4.4.1 Two-Dimensional Channel Estimation

Wideband channel estimation techniques may be classified as time-domain (TD) [4] or as frequency-domain (FD) [5] methods, which have different applications. The main criterion of deciding between these design options is essentially the length of the CIR expressed in terms of the number of transmitted symbols. More explicitly, given a total delay-spread of say 1us in a certain propagation enviroment, it is the transmission rate, which determines the normalized CIR length. For example, at a signalling rate of 1M Baud, a relatively short channel-sounding sequence of a few symbol-duration would provide an adequate CIR estimate, hence requiring a relatively low-order channel equalizer. By contrast, at 100M Baud the channel's despersion is on the order of 100 symbols, which may require an excessive channel sounding sequence duration and overhead as well as a complex, high-order time domain channel equalizaer. Furthermore, when employing LRP for the sake of improving the achievable channel estimation accuracy, on the order of 100 CIR taps would have to be predicted. In this scenario it may become implementationally less complex to transform the received N-symbol TD pilot block - which physically represents the estimated CIR - to the FD and carry out the LRP of the FD CTF, as we will discuss in the context of Figure 4.5.

In [103–107], a 2-D channel estimation algorithm has been proposed for wideband channel estimation aided multi-carrier systems. As shown in [103], the pilot symbols may be transmitted on different subcarriers of the OFDM symbols. At the receiver, two cascaded 1-D filters are used for estimating the channel in the frequency-domain and time-domain, respectively. This decomposed 2-D channel estimation algorithm has also been used in [104], where two cascaded 1-D filters were invoked for channel estimation in order to improve the attainable performance of MC-CDMA systems. Furthermore, in [105–107] a 2-D channel estimation algorithm using a 2-D Wiener filter was used for minimizing the MSE of 2-D channel estimation. The performance of channel estimation using two cascaded 1-D filters was shown to be similar to that of 2-D filters in [105–107]. However, the former has a lower complexity than the latter [105–107].

From now on let $\mathbf{b}'_{pk} = [b_{p(k;N-u)}, \cdots, b_{p(k;N-1)}, b_{p(k;0)}, \cdots, b_{p(k;N-1)}]$ denote the pilot symbol block, while $\mathbf{b}'_{dk} = [b_{d(k;N-u)}, \cdots, b_{d(k;N-1)}, b_{d(k;0)}, \cdots, b_{d(k;N-1)}]$ denote the data symbol block. Then the transmission of the pilot symbol block and data symbol block is illustrated in Figure 4.3. Moreover, in Figure 4.3 M denotes the interval between two pilot symbol blocks. Furthermore, let $\mathbf{r}_{pk} = [r_{p(k;0)}, \cdots, r_{p(k;N-1)}]$ denote the received signal block of (4.10) corresponding to pilot symbol block $\mathbf{b}'_{pk}$ after the removal of the cyclic prefix, while $\mathbf{r}_{dk} = [r_{d(k;0)}, \cdots, r_{d(k;N-1)}]$ denote the received signal block of (4.10) corresponding to data symbol block $\mathbf{b}'_{dk}$ after the removal of the

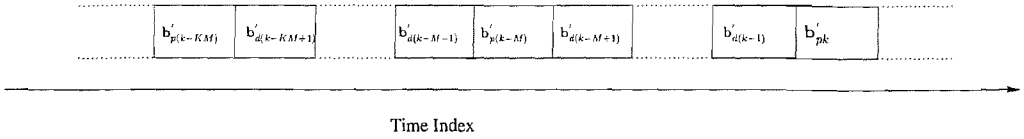


Figure 4.3: Illustration of transmitted symbol block, where the vector $\mathbf{b}'_{pk}$ denotes the pilot symbol block while the vector $\mathbf{b}'_{dk}$ denotes the data symbol block. Moreover, M denotes the interval between two pilot symbol blocks.

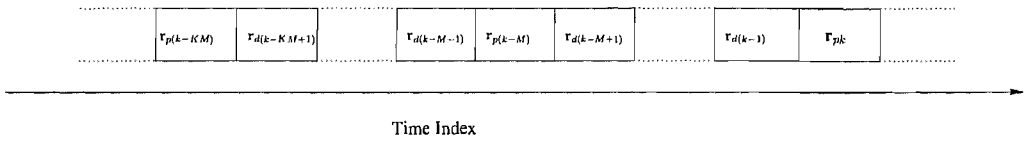


Figure 4.4: Illustration of received signal block, where the vector $\mathbf{r}_{pk}$ denotes the received singal block corresponding to pilot symbol block $\mathbf{b}'_{pk}$, while the vector $\mathbf{r}_{dk}$ denotes the received signal block corresponding to data symbol block $\mathbf{b}'_{dk}$.

cyclic prefix, as shown in Figure 4.4.

The illustration of CTF estimation and CTF prediction in the context of single-carrier systems is described in Figure 4.5 where M denotes the interval between two pilot symbol blocks.

Let $\mathbf{b}_{pk} = [b_{p(k;0)}, b_{p(k;1)}, \cdots, b_{p(k;N-1)}]^T$ denote the N-symbol pilot data vector. Then according to (4.10), the corresponding N-symbol recived signal vector $\mathbf{r}_{pk}$ is given by

$$\mathbf{r}_{pk} = \mathbf{C}_{pk}\mathbf{b}_{pk} + \mathbf{z}_{pk}, \tag{4.25}$$

where $\mathbf{C}_{pk}$ is the $N \times N$-component circulant CIR matrix of (4.14) corresponding to the pilot symbol

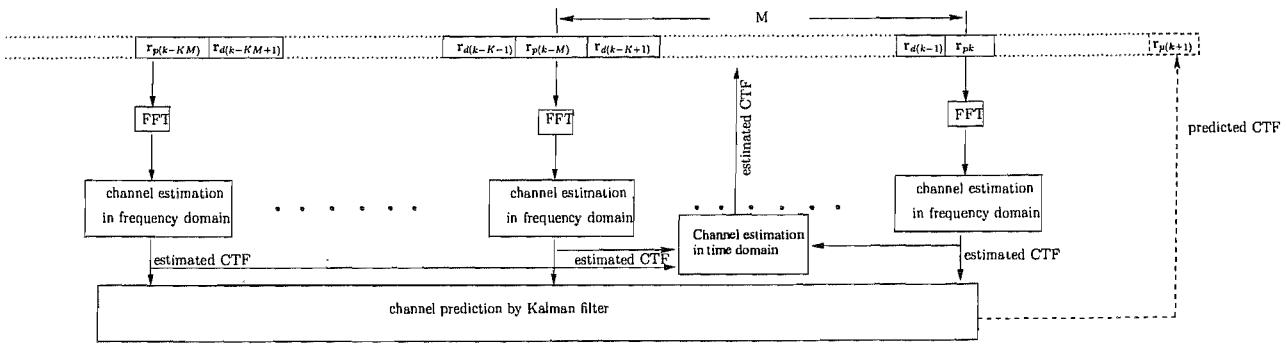


Figure 4.5: Illustration of channel estimation and LRP for single carrier systems

block and $\mathbf{z}_{pk}$ is the N-component the AWGN vector, currupting the pilot symbol block. The N-symbol time-domain pilot information can be converted to the frequency-domain with the aid of the N-point DFT of (4.20) as shwon in Figure 4.5, yielding

$$\begin{aligned} \breve{\mathbf{r}}_{pk} &= \sqrt{N}\mathbf{Q}\mathbf{C}_{pk}\mathbf{b}_{pk} + \sqrt{N}\mathbf{Q}\mathbf{z}_{pk} \\ &= \Lambda_{pk}\breve{\mathbf{b}}_{pk} + \breve{\mathbf{z}}_{pk}. \end{aligned} \tag{4.26}$$

According to (4.21), the received signal $\breve{r}_{p(k;n)}$ is expressed as

$$\breve{r}_{p(k;n)} = \breve{c}_{p(k;n)}\breve{b}_{p(k;n)} + \breve{z}_{p(k;n)}, n = 0, 1, \cdots, N-1. \tag{4.27}$$

Upon dividing both sides of (4.27) by $\breve{b}_{p(k;n)}$, the resultant estimate of $\breve{c}_{p(k;n)}$ can be expressed as

$$\tilde{c}_{p(k;n)} = \frac{\breve{r}_{p(k;n)}}{\breve{b}_{p(k;n)}} = \breve{c}_{p(k;n)} + \frac{\breve{z}_{p(k;n)}}{\breve{b}_{p(k;n)}}, n = 0, 1, \cdots, N-1, \tag{4.28}$$

where $\tilde{c}_{p(k;n)}$ represents the estimate of $\breve{c}_{p(k;n)}$ provided by the nth DFT bin on the basis of the time-domain pilots $\mathbf{b}_{pk}$. The autocorrelation of $\breve{c}_{p(k;n)}$ may be expressed as [106]

$$\begin{aligned} r[k-k';n-n'] = r[\Delta k;\Delta n] &= E[\breve{c}_{p(k;n)}\breve{c}^{*}_{p(k';n')}] \\ &= r_t[k-k']r_f[n-n'] \\ &= r_t[\Delta k]r_f[\Delta n], \end{aligned} \tag{4.29}$$

where we have $\Delta k = k - k'$, $\Delta n = n - n'$, and $r_t[\Delta k]$ is the resultant time-domain autocorrelation function, which can be expressed as [106]

$$r_t[\Delta k] = J_0(2\pi f_{dm}\Delta k T_B), \tag{4.30}$$

where again, T_B is the duration of the symbol block, $J_0(\cdot)$ is the zero order Bessel function of the first kind and f_{dm} is the maximum Doppler frequency. In (4.29) $r_f[\Delta n]$ is the autocorrelation function in the frequency-domain, which can be expressed as [108]

$$r_f[\Delta n] = \sum_{l=0}^{L-1} \sigma_l^2 e^{-j2\pi\Delta n f_0 \tau_l}, \tag{4.31}$$

where we have $f_0 = 1/(NT_b)$ [109], T_b represents symbol duration and L is the number of multipath

components. Finally, in (4.31) τ_l is the delay of the lth multipath component and σ_l^2 is the average power of the lth multipath component.

The autocorrelation function of $\tilde{c}_{p(k;n)}$ can be expeseed as

$$\begin{aligned} E[\tilde{c}_{p(k;n)}\tilde{c}^*_{p(k';n')}] &= E[(\breve{c}_{p(k;n)} + \frac{\breve{z}_{p(k;n)}}{\breve{b}_{p(k;n)}})(\breve{c}_{p(k';n')} + \frac{\breve{z}_{p(k';n')}}{\breve{b}_{p(k';n')}})^*] \\ &= r_t[k-k']r_f[n-n'] + E[(\frac{\breve{z}_{p(k;n)}}{\breve{b}_{p(k;n)}})(\frac{\breve{z}_{p(k';n')}}{\breve{b}_{p(k';n')}})^*] \\ &= r_t[\Delta k]r_f[\Delta n] + NN_0\frac{1}{\breve{b}_{p(k;n)}\breve{b}^*_{p(k';n')}}\delta(k-k')\delta(n-n'), \end{aligned} \quad (4.32)$$

where we have exploited that $\breve{c}_{p(k;n)}$ is uncorrelated with $\breve{z}_{p(k;n)}$ and that $\breve{z}_{p(k;n)}$ is the AWGN with zero mean and a variance of NN_0 [110]. Furthermore, the crosscorrelation between $\tilde{c}_{p(k;n)}$ and $\breve{c}_{p(k;n)}$ is expressed as

$$\begin{aligned} E[\tilde{c}_{p(k;n)}\breve{c}^*_{p(k';n')}] &= E[(\breve{c}_{p(k;n)} + \frac{\breve{z}_{p(k;n)}}{\breve{b}_{p(k;n)}})\breve{c}^*_{p(k';n')}] \\ &= r_t[k-k']r_f[n-n']. \end{aligned} \quad (4.33)$$

Let us now consider wideband channel estimation based on two cascaded 1-D filters. In this case channel estimation is carried out first in the frequency-domain and then in the time-domain.

4.4.1.1 Frequency-Domain Channel Estimation

In the context of FD channel estimation as shown in Figure 4.5, the channel estimates formulated in (4.28) are used. Consequently, the optimum channel estimation in the MMSE sense formulated in the frequency-domain can be expressed for the pilot data block at time instant k as

$$\begin{aligned} \hat{c}_{p(k;n)} &= \sum_{i=0}^{N-1} d^*_{(k;i)}\tilde{c}_{p(k;i)} \\ &= \mathbf{d}^H_{(k)}\tilde{\mathbf{c}}_{p(k)} \qquad 0 \le n \le N-1, \end{aligned} \quad (4.34)$$

where $\tilde{\mathbf{c}}_{p(k)} = [\tilde{c}_{p(k;0)}, \tilde{c}_{p(k;1)}, \cdots, \tilde{c}_{p(k;N-1)}]^T$ and $\hat{c}_{p(k;n)}$ represents the estimate of $\breve{c}_{p(k;n)}$ in the frequency-domain. Furthermore, in (4.34) $d_{(k;i)}$ denotes the ith filter coefficient and $\mathbf{d}_{(k)} = [d_{(k,0)}, d_{(k;1)}, \cdots, d_{(k;N-1)}]^T$ denotes the coefficient vector of the frequency-domain filter used for

estimating $\breve{c}_{p(k;n)}$, which is determined by [1],

$$\mathbf{d}_{(k)} = \mathbf{R}_f^{-1}\mathbf{r}_f, \tag{4.35}$$

where $\mathbf{R}_f$ is the $(N \times N)$-dimensisonal autocorrelation matrix of $\tilde{\mathbf{c}}_{p(k)}$, which is given according to (4.32) by:

$$\begin{aligned}\mathbf{R}_f &= E[\tilde{\mathbf{c}}_{p(k)}\tilde{\mathbf{c}}_{p(k)}^H] \\ &= \begin{bmatrix} r_t[0]r_f[0] + \frac{NN_0}{|\breve{b}_{(k;0)}|^2} & r_t[0]r_f[-1] & \cdots & r_t[0]r_f[1-N] \\ r_t[0]r_f[1] & r_t[0]r_f[0] + \frac{NN_0}{|\breve{b}_{(k;1)}|^2} & \cdots & r_t[0]r_f[2-N] \\ \vdots & \vdots & \ddots & \vdots \\ r_t[0]r_f[N-1] & r_t[0]r_f[N-2] & \cdots & r_t[0]r_f[0] + \frac{NN_0}{|\breve{b}_{(k;N-1)}|^2} \end{bmatrix}\end{aligned} \tag{4.36}$$

and $\mathbf{r}_f$ is the N-dimensional cross correlation vector between $\tilde{\mathbf{c}}_{p(k)}$ and $\breve{c}^*_{p(k;n)}$, which is given according to (4.33) by

$$\mathbf{r}_f = E[\tilde{\mathbf{c}}_{p(k)}\breve{c}^*_{p(k;n)}] = E\begin{bmatrix} \tilde{c}_{p(k;0)}\breve{c}^*_{p(k;n)} \\ \vdots \\ \tilde{c}_{p(k;n)}\breve{c}^*_{p(k;n)} \\ \vdots \\ \tilde{c}_{p(k;N-1)}\breve{c}^*_{p(k;n)} \end{bmatrix} = \begin{bmatrix} r_t[0]r_f[-n] \\ \vdots \\ r_t[0]r_f[0] \\ \vdots \\ r_t[0]r_f[N-1-n] \end{bmatrix}. \tag{4.37}$$

Based on (4.30), we have

$$r_t[0] = J_0(0) = 1. \tag{4.38}$$

Hence upon substituting Eq.(4.38) into (4.36) and (4.37), we have

$$\mathbf{R}_f = \begin{bmatrix} r_f[0] + \frac{NN_0}{|\breve{b}_{(k;0)}|^2} & r_f[-1] & \cdots & r_f[1-N] \\ r_f[1] & r_f[0] + \frac{NN_0}{|\breve{b}_{(k;1)}|^2} & \cdots & r_f[2-N] \\ \vdots & \vdots & \ddots & \vdots \\ r_f[N-1] & r_f[N-2] & \cdots & r_f[0] + \frac{NN_0}{|\breve{b}_{(k;N-1)}|^2} \end{bmatrix} \tag{4.39}$$

and

$$\mathbf{r}_f = \begin{bmatrix} r_f[-n] \\ \vdots \\ r_f[0] \\ \vdots \\ r_f[N-1-n] \end{bmatrix}. \tag{4.40}$$

After frequency-domain estimation, the minimum MSE can be derived as follows. Let $J_{f(k;n)}$ denote the MSE after frequency domain estimation. Then we have

$$J_{f(k;n)} = E[|\breve{c}_{p(k;n)} - \hat{c}_{p(k;n)}|^2]. \tag{4.41}$$

Upon substituting (4.34) into (4.41), we arrive at:

$$\begin{aligned} J_{f(k;n)} &= E[(\breve{c}_{p(k;n)} - \mathbf{d}_{(k)}^H \tilde{\mathbf{c}}_{p(k)})(\breve{c}_{p(k;n)} - \mathbf{d}_{(k)}^H \tilde{\mathbf{c}}_{p(k)})^*] \\ &= E[|\breve{c}_{p(k;n)}|^2 - \mathbf{d}_{(k)}^H \tilde{\mathbf{c}}_{p(k)} \breve{c}_{p(k;n)}^* - \mathbf{d}_{(k)}^T \tilde{\mathbf{c}}_{p(k)}^* \breve{c}_{p(k;n)} + \mathbf{d}_{(k)}^H \tilde{\mathbf{c}}_{p(k)} \tilde{\mathbf{c}}_{p(k)}^H \mathbf{d}_{(k)}] \\ &= \sigma_{\breve{c}}^2 - \mathbf{d}_{(k)}^H \mathbf{r}_f - \mathbf{d}_{(k)}^T \mathbf{r}_f^* + \mathbf{d}_{(k)}^H \mathbf{R}_f \mathbf{d}_{(k)}, \end{aligned} \tag{4.42}$$

where according to (4.29) and (4.38), we have

$$\sigma_{\breve{c}}^2 = E[|\breve{c}_{p(k;n)}|^2] = r_t[0] r_f[0] = r_f[0] = \sum_{l=0}^{L-1} \sigma_l^2, \tag{4.43}$$

which represents the average power of $\breve{c}_{p(k;n)}$ received over L multipath component channels. Furthermore, when substituting (4.35) and (4.43) into (4.42), it can be shown that the minimum MSE $J_{fo(k;n)}$ after the frequency domain filtering is given by

$$\begin{aligned} J_{fo(k;n)} &= \sigma_{\breve{c}}^2 - \mathbf{r}_f^H \mathbf{R}_f^{-1} \mathbf{r}_f \\ &= r_f[0] - \mathbf{r}_f^H \mathbf{R}_f^{-1} \mathbf{r}_f. \end{aligned} \tag{4.44}$$

Finally, after MMSE frequency-domain estimation, $\hat{c}_{p(k;n)}$ can be expressed as

$$\hat{c}_{p(k;n)} = \breve{c}_{p(k;n)} + \zeta_{(k;n)}, \tag{4.45}$$

where $\zeta_{(k;n)}$ represents the estimation error between $\hat{c}_{p(k;n)}$ and $\breve{c}_{p(k;n)}$ in the MMSE sense, which can be modeled as a zero mean process having a variance of $J_{fo(k;n)}$. Furthermore, the autocorrelation function of $\hat{c}_{p(k;n)}$ derived for a given frequency component is given by

$$\begin{aligned} E[\hat{c}_{p(k;n)}\hat{c}^*_{p(k';n)}] &= E[(\breve{c}_{p(k;n)} + \zeta_{(k;n)})(\breve{c}_{p(k';n)} + \zeta_{(k';n)})^*] \\ &= r_t[k-k']r_f[0] + J_{fo(k;n)}\delta(k-k'), \end{aligned} \tag{4.46}$$

where we have used the property that, in Wiener filtering, $\zeta_{(k;n)}$ is independent of $\hat{c}_{p(k;n)}$ and $\zeta_{(k;n)}$ is also independent of $\zeta_{(k';n)}$, when $k \neq k'$ [106, 107]. The crosscorrelation between $\hat{c}_{p(k;n)}$ and $\breve{c}_{p(k;n)}$ derived for a given frequency component can be expressed as

$$\begin{aligned} E[\hat{c}_{p(k;n)}\breve{c}^*_{p(k';n)}] &= E[(\breve{c}_{p(k;n)} + \zeta_{(k;n)})\breve{c}^*_{p(k';n)}] \\ &= r_t[k-k']r_f[0]. \end{aligned} \tag{4.47}$$

4.4.1.2 Time-Domain Channel Estimation

In section 4.4.1.1 the CTF corresponding to N-symbol pilot block was carried out in the frequency-domain in the context of a fixed time index k. Recall from Figure 4.2 that the consecutive time-domain fading samples associated with different values of k and recorded for each fixed frequency index n of the FD CTF are also correlated. Hence, following frequency-domain CTF estimation, we can carry out CTF estimation also in the time-domain for both the pilot and the data blocks as seen in Figure 4.5. Specifically, if at time instant k an N-symbol pilot block was received first, FD CTF estimation is carried out, as discussed in Section 4.4.1.1. Then, for a fixed frequency index n, K consecutive estimated FD CTFs $\hat{c}_{p(k;n)}$ of (4.34) were collected and assigned to a K-element vector $\hat{\mathbf{c}}_{p(n)}$, which is given by

$$\hat{\mathbf{c}}_{p(n)} = [\hat{c}_{p(k;n)}, \hat{c}_{p(k-M;n)}, \cdots, \hat{c}_{p(k-(K-1)M;n)}]^T. \tag{4.48}$$

Hence the optimum linear CTF estimation in the MMSE sense can be expressed in the time domain as [106]

$$\begin{aligned} \acute{c}_{(k';n)} &= \sum_{j=0}^{K-1} d^*_{(j;n)}\hat{c}_{p(k-jM;n)} \\ &= \mathbf{d}^H_{(n)}\hat{\mathbf{c}}_{p(n)}, \qquad (K-1)M \leq k' \leq k, \end{aligned} \tag{4.49}$$

where $\acute{c}_{(k';n)}$ is the estimate of $\breve{c}_{(k';n)}$ based on the samples $\hat{\mathbf{c}}_{p(n)}$ estimated in the frequency-domain, while $\mathbf{d}_{(n)} = [d_{(0;n)}, d_{(1;n)}, \cdots, d_{(K-1;n)}]^T$ denotes the coefficient vector of the time-domain filter, which is given by [1]

$$\mathbf{d}_{(n)} = \mathbf{R}_t^{-1}\mathbf{r}_t, \tag{4.50}$$

where $\mathbf{R}_t$ is the $(K \times K)$-dimensional autocorrelation matrix of $\hat{\mathbf{c}}_{p(n)}$, expressed as

$$\begin{aligned}\mathbf{R}_t &= E[\hat{\mathbf{c}}_{p(n)}\hat{\mathbf{c}}_{p(n)}^H] \\ &= \begin{bmatrix} r_t[0]r_f[0] + J_{fo(0;n)} & r_t[-M]r_f[0] & \cdots & r_t[(1-K)M]r_f[0] \\ r_t[M]r_f[0] & r_t[0]r_f[0] + J_{fo(1;n)} & \cdots & r_t[2-K]r_f[0] \\ \vdots & \vdots & \ddots & \vdots \\ r_t[(K-1)M]r_f[0] & r_t[(K-2)M]r_f[0] & \cdots & r_t[0]r_f[0] + J_{fo(K-1;n)} \end{bmatrix}.\end{aligned} \tag{4.51}$$

Furthermore, $\mathbf{r}_t$ in (4.50) is the K-dimensional cross-correlation vector of $\hat{\mathbf{c}}_{p(n)}$ and $\breve{c}^*_{(k';n)}$, which is given by

$$\mathbf{r}_t = E[\hat{\mathbf{c}}_{p(n)}\breve{c}^*_{(k';n)}] = E\begin{bmatrix} \hat{c}_{(0;n)}\breve{c}^*_{(k';n)} \\ \vdots \\ \hat{c}_{((K-1)M;n)}\breve{c}^*_{(k';n)} \end{bmatrix} = \begin{bmatrix} r_t[-k']r_f[0] \\ \vdots \\ r_t[(K-1)M - k']r_f[0] \end{bmatrix}. \tag{4.52}$$

Correspondingly, the MMSE $J_{to(k;n)}$ recorded after the time-domain channel estimation can be expressed as

$$\begin{aligned} J_{to(k';n)} &= \sigma_{\breve{c}}^2 - \mathbf{r}_t^H\mathbf{R}_t^{-1}\mathbf{r}_t \\ &= r_f[0] - \mathbf{r}_t^H\mathbf{R}_t^{-1}\mathbf{r}_t, \end{aligned} \tag{4.53}$$

After the time-domain estimation in the MMSE sense, $\acute{c}_{(k;n)}$ can be expressed as

$$\acute{c}_{(k';n)} = \breve{c}_{(k';n)} + \xi_{(k';n)} \tag{4.54}$$

where $\xi_{(k';n)}$ represents the estimation error between $\acute{c}_{(k';n)}$ and $\breve{c}_{(k';n)}$ in the MMSE sense, which is a zero-mean process with a variance of $J_{to(k';n)}$.

Let us now investigate the LRP of wideband channels based on the observation samples obtained in this section.

4.5 Long-Range Channel Prediction of Wideband Channels

Above we have demonstrated the principles of 2-D wideband channel estimation. In order to employ the narrowband based Kalman filtering assisted LRP described in Section 3.3 in the context of wideband channels, the CTF $\hat{c}_{p(k;n)}$ estimated after the first stage frequency-domain filtering can be directly used for TD CTF prediction. Hence, the CTF corresponding to the next pilot symbol block is predicted with the aid of all previous pilot blocks available from the past, as illustrated in Figure 4.5, which allow us for example to predict the resultant CTF for the future instant of reception at a mobile receiver by a transmit preprocessing scheme.

Since any of the frequency components experiences flat fading, the Kalman filtering assisted LRP derived in Section 3.3 can be directly used in the context of each frequency components.

More specifically, for any fixed frequency component n, according to (3.52), the channel's future state can be described by the AR process of [79]

$$\breve{c}_{p(k;n)} = \sum_{m=1}^{P} a_m \breve{c}_{p(k-mM;n)} + w_{(k;n)}, \tag{4.55}$$

where P is the order of the AR process. Let $\breve{\mathbf{c}}_{p(k;n)}$ be a P-dimensional CTF vector, which is written as

$$\breve{\mathbf{c}}_{p(k;n)} = [\breve{c}_{p(k;n)}, \breve{c}_{p(k-M;n)}, \cdots, \breve{c}_{p(k-(P-1)M;n)}]^T. \tag{4.56}$$

Then, the channel prediction problem can be modelled by the classic Kalman process equation of [1]

$$\breve{\mathbf{c}}_{p(k;n)} = \mathbf{F}_{(k-1;n)} \breve{\mathbf{c}}_{p(k-1;n)} + \mathbf{w}_{(k;n)}, \tag{4.57}$$

where $\mathbf{F}_{(k-1;n)}$ and $\mathbf{w}_{(k;n)}$ are defined in (3.55) and (3.56), respectively. According to (3.58), the CTF $\hat{c}_{p(k;n)}$ of (4.45) estimated in the frequency domain can be expressed by the Kalman measurement equation as

$$\hat{c}_{p(k;n)} = \mathbf{H}_{(k;n)} \breve{\mathbf{c}}_{p(k;n)} + \zeta_{(k;n)}, \tag{4.58}$$

where $\mathbf{H}_{(k;n)}$ is a $(1 \times P)$-dimensional measurement matrix, which is given by

$$\mathbf{H}_{(k;n)} = [1, 0, \cdots, 0]. \tag{4.59}$$

According to (3.59), the prediction of $\breve{\mathbf{c}}_{p(k+1;n)}$, which is the specific value of the CTF at fre-

quency index n at time instant $(k+M)$, can be expressed as [1]

$$\grave{\mathbf{c}}_{p[(k+1;n)|(k;n)]} = \mathbf{F}_{(k;n)}\grave{\mathbf{c}}_{p[(k;n)|(k-1;n)]} + \mathbf{G}_{(k;n)}\boldsymbol{\alpha}_{(k;n)}, \tag{4.60}$$

where $\grave{\mathbf{c}}_{p[(k+1;n)|(k;n)]}$ represents the prediction of $\breve{\mathbf{c}}_{p(k+1;n)}$ in the MMSE sense, based on all the past observations corresponding to all the N-symbol pilot blocks from the past until time instant k for a fixed frequency component n. Furthermore, in (4.60), $\mathbf{G}_{(k;n)}$ is the Kalman gain and $\boldsymbol{\alpha}_{(k;n)}$ is the innovation process [1], which are given by [1]

$$\mathbf{G}_{(k;n)} = \mathbf{F}_{(k;n)}\mathbf{K}_{(k;n)}\mathbf{H}^H_{(k;n)}\mathbf{R}^{-1}_{(k;n)}, \tag{4.61}$$

$$\boldsymbol{\alpha}_{(k;n)} = \hat{c}_{p(k;n)} - \mathbf{H}_{(k;n)}\grave{\mathbf{c}}_{[(k;n)|(k-1;n)]}, \tag{4.62}$$

where $\mathbf{K}_{(k+1;n)}$ is the correlation of the prediction state error vector, which is given by [1]

$$\mathbf{K}_{(k+1;n)} = \mathbf{F}_{(k;n)}\mathbf{M}_{(k;n)}\mathbf{F}^H_{(k;n)} + \mathbf{Q}_{w_{(k+1;n)}} \tag{4.63}$$

and $\mathbf{M}_{(k;n)}$ is the estimation state error vector, which is given by [1]

$$\mathbf{M}_{(k;n)} = \mathbf{K}_{(k;n)} - \mathbf{K}_{(k;n)}\mathbf{H}^H_{(k;n)}(\mathbf{H}_{(k;n)}\mathbf{K}_{(k;n)}\mathbf{H}^H_{(k;n)} + \mathbf{Q}_{\zeta_{(k;n)}})^{-1}\mathbf{H}_{(k;n)}\mathbf{K}_{(k;n)}. \tag{4.64}$$

Moreover, the first element in $\mathbf{K}_{(k+1;n)}$ represents the resultant MMSE, when $\breve{c}_{p(k+1;n)}$ is predicted by the observations $\hat{c}_{p(1;n)}, \cdots, \hat{c}_{p(k;n)}$.

Let us now provide a range of simulation results in order to characterize the achievable performance of both wideband channel estimation and prediction.

4.6 Simulation Results

For single-carrier wideband systems, the design of optimal pilot sequence has been discussed intensively in [111–116]. Since a Chu-sequence [117] has a constant magnitue in both the time domain and in the frequency domain, it avoids the high PAPR of OFDM systems and as an added benefit, it also yields identical MMSE estimates for all frequency components of the CTF [114–116]. Hence in

this section Chu-sequences [117] were adopted as pilot symbols, which are given by [117]

$$b_{p(k,n)} = \begin{cases} \exp\frac{jN'\pi n^2}{N} & \text{if N is even} \\ \exp\frac{jN'\pi n(n+1)}{N} & \text{if N is odd} \end{cases} \tag{4.65}$$

where N' and N are relatively primes [117].

Furthermore, according to (4.5), the average received SNR, γ_b, calculated before the DFT was applied to the pilot symbol block, is defined as

$$\begin{aligned} \gamma_b &= \frac{E_s}{N_0}\sum_{l=0}^{L-1} E[|c_{(k;l)}|^2] \\ &= \frac{E_s}{N_0}\sum_{l=0}^{L-1} \sigma_l^2, \end{aligned} \tag{4.66}$$

where E_s is the average energy of $b_{p(k,n)}$, N_0 is the variance of the noise and $\sigma_l^2 = E[|c_{(k;l)}|^2]$ is the average power of the lth path.

Moreover, according to (4.21), the average received SNR, $\gamma_{(a,n)}$, evaluated for the nth frequency component after the DFT operation is defined as

$$\gamma_{(a,n)} = \frac{E[|\breve{c}_{(k;n)}\breve{b}_{(k;n)}|^2]}{E[\breve{z}_{(k;n)}|^2]}. \tag{4.67}$$

According to (4.18), we have $E[|\breve{c}_{(k;n)}|^2] = \sum_{l=0}^{L-1}\sigma_l^2$ and $E[|\breve{z}_{(k;n)}|^2] = NN_0$. Hence (4.67) can be expressed as

$$\gamma_{(a,n)} = \frac{\breve{E}_s}{NN_0}\sum_{l=0}^{L-1}\sigma_l^2, \tag{4.68}$$

where $\breve{E}_s$ is the average energy of $\breve{b}_{(k;n)}$. Since the Chu sequences of (4.66) are adopted, we have $\gamma_{(a,0)} = \gamma_{(a,1)} = \cdots = \gamma_{(a,N-1)}$.

Furthermore, in order to demonstrate the benefits of CTF estimation in the frequency domain, we introduce the effective SNR $\gamma_{(e,n)}$ for employment in LRP after the FD CTF estimation was carried

Normalized Doppler frequency $f_{dm}T_B$	0.001
Number of multipath	$L = 3$
Length of data block	$N = 128$
Length of cyclic prefix is	$u = 2$
Power intensity for each multipath component σ_l^2	1.0/3.0
Predictor's order	$P = 20$
$[\tau_0, \tau_l, \tau_2]$	$[0, T_b, 2T_b]$

Table 4.1: Parameters for linear wideband channel estimation and prediction

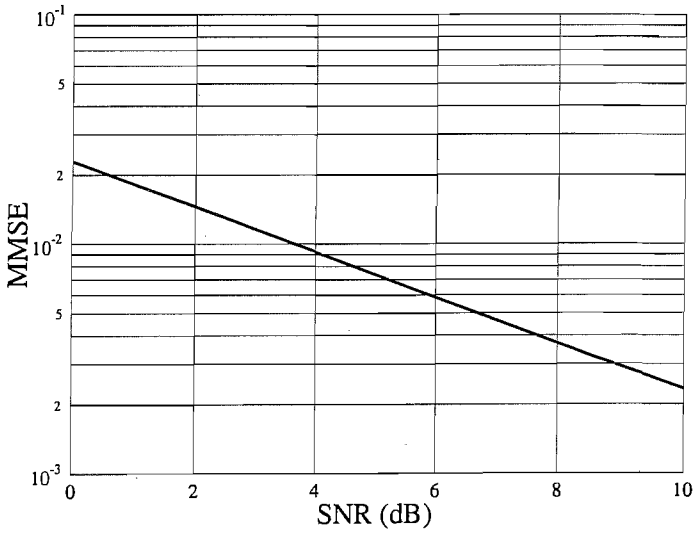


Figure 4.6: MMSE of (4.44) versus average the SNR per symbol of (4.66) performance of CTF estimation in the frequency domain. The remaining parameters are summarized in Table 4.1.

out for the nth frequency component based on (4.44), which is defined as

$$\begin{aligned} \gamma_{(e,n)} &= \frac{E[|\breve{c}_{(k;n)}|^2]}{E|\zeta_{(k;n)}|^2} \\ &= \frac{1}{J_{fo(k;n)}} \sum_{l=0}^{L-1} \sigma_l^2. \end{aligned} \tag{4.69}$$

Likewise we have $\gamma_{(e,0)} = \gamma_{(e,1)} = \cdots = \gamma_{(e,N-1)}$ since Chu sequences are invoked.

In Figure 4.6 the MMSE versus average the SNR per symbol performance of CTF estimation in

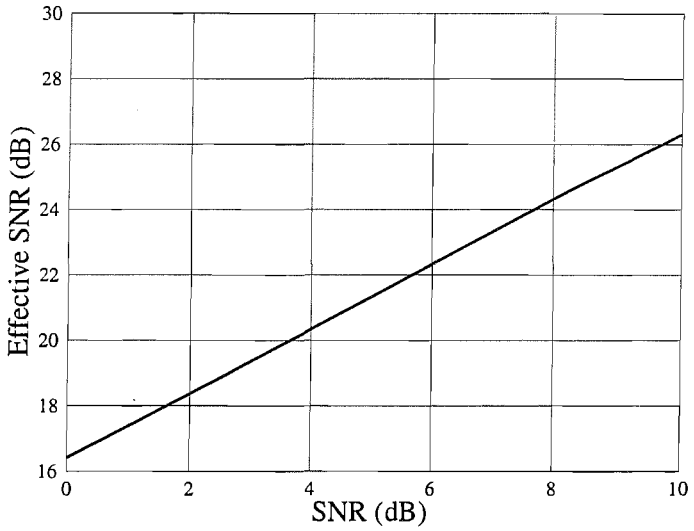


Figure 4.7: The effective SNR of (4.69) for LRP after CTF estimation in the frequency domain per symbol versus average SNR of (4.66) per symbol. The remaining parameters are summarized in Table 4.1.

SNR γ_b	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Effective SNR $\gamma_{(e,n)}$	16.4	17.4	18.4	19.4	20.3	21.3	22.3	23.3	24.3	25.3	26.3

Table 4.2: Effective SNR of (4.69) after frequency domain estimation

the frequency domain is plotted. The system parameters are summarized in Table 4.1. As we can see from Figure 4.6, MMSE decreases linearly upon increasing SNR.

In Figure 4.7 the effective SNR of (4.69) calculated for LRP after CTF estimation in the frequency domain per symbol versus the average SNR of (4.66) per symbol is plotted. The remaining parameters are summarized in Table 4.1. We can see from Figure 4.7 that the effective SNR of (4.69) increases linearly upon increasing the SNR of (4.66). This is because the MMSE of (4.44) decreases linearly upon increasing the SNR, as seen in Figure 4.6. Furthermore, the effective SNR of (4.69) calculated for LRP after CTF estimation in the frequency domain corresponding to Figure 4.7 is summarized in Table 4.2. We can see from Table 4.2 that an approximately 16.3 dB SNR gain can be achieved after frequency domain estimation.

In Figure 4.8 the CTF generated after frequency domain estimation is plotted, when the value of the SNR expression of (4.66) was γ_b=10dB. The remaining parameters are summarized in Table 4.1.

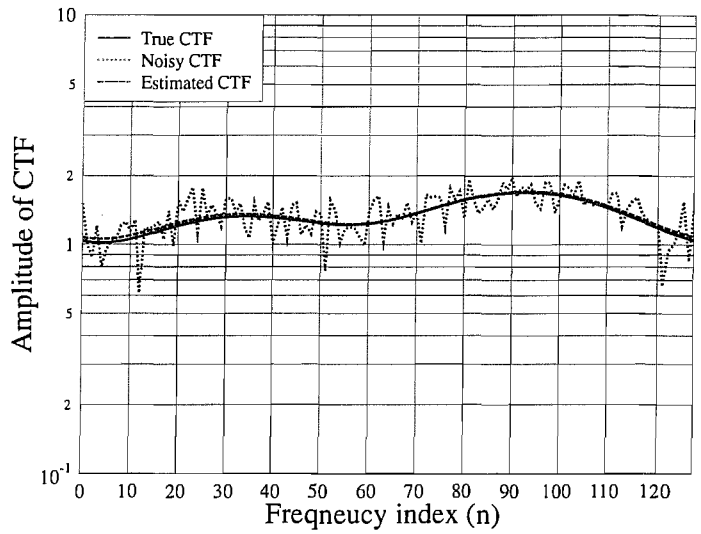


Figure 4.8: CTF after frequency domain estimation when SNR of (4.66) is γ_b=10dB. The remaining parameters are summarized in Table 4.1.

We can see from Figure 4.8 that the noisy CTF generated before frequency domain estimation substantially deviate from the true CTF owing to the effects of the AWGN. However, observe in Figure 4.8 that the effects of noise are effectively mitigated by frequency domain filtering and hencethe estimated CTF closely argees with the true CTF after frequency domain filtering.

In Figure 4.9 the MMSE versus SNR of (4.66) per symbol performance of was plotted after Kalman filtering assisted prediction, when the pilot block interval M was $M = 5$, 10, 15, 20 and the normalized maximum Doppler frequency was set to $f_{dm}T_B = 0.001$ and 0.0001, respectively. The remaining parameters are summarized in Table 4.1. We can see from Figure 4.9 that the MMSE performance corresponding to $f_{dm}T_B = 0.0001$ is always better than that corresponding to $f_{dm}T_B = 0.001$. This because the lower normalized maximum Doppler frequency corresponds to a slower TD fluctuation of the CTF and hence the Kalman filtering assisted predictor can predict the CTF more precisely. Furthermore, for a given normalized maximum Doppler frequency $f_{dm}T_B$, the MMSE is increased upon increasing the interval between pilot symbol blocks M from $M = 5$ to $M = 20$. This is because the lower the interval M between N-symbol pilot blocks, the stronger the correlation between the pilot blocks, hence the Kalman filtering assisted predictor can predict the CTF more precisely.

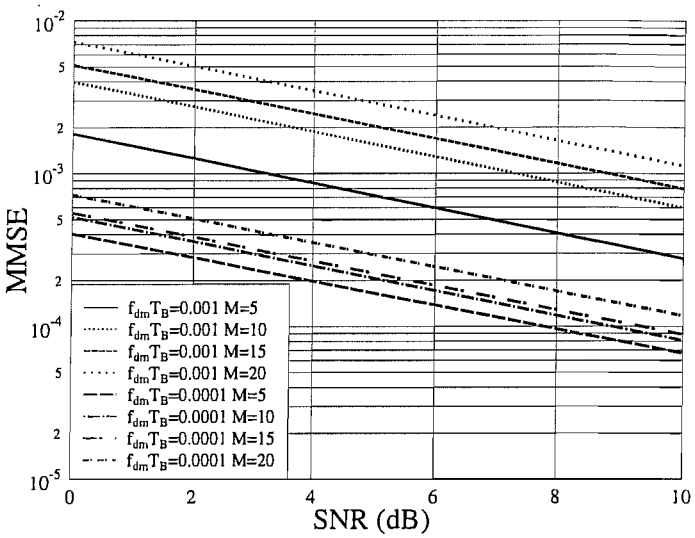


Figure 4.9: MMSE versus SNR of (4.66) per symbol performance of Kalman filtering assisted prediction when the pilot block interval M is $M = 5, 10, 15, 20$ and the normalized maximum Doppler frequency $f_{dm}T_B = 0.001$ and 0.0001, respectively. The remaining parameters are summarized in Table 4.1.

4.7 Conclusions

In this chapter, both channel estimation and prediction were investigated in the context of a single-carrier wideband system. The received signals corresponding to N-symbol pilot blocks were first processed by the DFT operation in order to transfter the CIR to the frequency domain. Then 1-D CTF estimation was carried out in the frequency domain. As shown in Table 4.2, as approximately 16.3dB SNR gain can be achieved. Then both CTF estimation and prediction can be carried out in the time domain. Since the effect of AWGN has been mitigated by frequency domain filtering as shown in Figure 4.8, a good estimation and prediction performance can be expected, as evidenced by Figure 4.9. Furthermore, as we can see from Figure 4.9 the lower the normalized maximum Doppler frequency and the lower the interval between pilot symbol blocks, the better the achievable performance.

Chapter 5

Subspace Tracking Based Blind MIMO Transmit Preprocessing

5.1 Introduction

Due to the emerging high demand for supporting novel multimedia applications, next generation wireless systems are expected to support high data rates. When employing multiple antennas at both the transmitter and receiver, multiple input multiple output (MIMO) systems have the potential achieving a high transmission rate than their traditional single input single output (SISO) systems counterparts [60].

MIMO systems have attracted intensive research interests during the last decade [60–62]. In the absence of CSI at the tranmitter, space time coding [118] or spatial multiplexing [119–121] constitue prime candidates for MIMO transmission. However, when the CSI is available at both the transmitter and the receiver, a more sophisticated technique refered to as eigenmode transmission [64] can be used for decomposing the MIMO channel into several independent SISO subchannels, which involves the singular value decomposition (SVD) of the MIMO channel matrix. In this case no joint detection is needed and the resultant single-antenna-based detection algorithm becomes rather simple.

The third-generation (3G) wireless systems support two different modes, namely frequency division duplexing (FDD) and time division duplexing (TDD) [122–125]. In the FDD mode, the uplink (UL) and downlink (DL) signals are transmitted at different carrier frequencies, which results in independently fading channels for the UL and DL. By contrast, in the TDD mode, the UL and DL transmissions ensue at the same carrier frequency. Hence the UL and DL channels tend to fade to-

gether and therefore can be considered as similar [122]. We will exploit this similarity of the UL and DL channels of the TDD mode in this chapter.

The CSI required at the transmitter can be obtained with the aid of the side-information control channel from the receiver in the FDD mode. Alternatively, it can be directly estimated on the basis of the received signal's quality and exploited by the transmitter in the TDD mode [126].

Channel estimation (CE) followed by SVD is invoked, when eigenmode transmissions are employed [127–129], which potentially imposes a high computational complexiy. Instead of estimating the entire MIMO channel matrix and then additionally implementing SVD, it was claimed in [130–139] that subspace tracking based algorithms may result in lower computational complexity in the context of eigenmode transmissions.

In the family of subspace tracking algorithms, the so-called projection approximation tracking combined with deflation (PASTD) [140] has been shown to be applicable in diverse scenarios [141, 142]. Hence, in this chapter PASTD algorithm is employed for subspace tracking in a MIMO-aided TDD system.

The chapter is structured as follows. In Secion 5.2, a MIMO system using eigenmode transmission is considered, while in Secion 5.3, a TDD-based MIMO system is discussed further. In Secion 5.4, the PASTD algorithm is are introduced. In Secion 5.5, differential coding is invoked for removing the phase ambiguity imposed by the non-unique nature [143] of the subspace considered. In Section 5.6 simulation results are provided. Finally, our conclusions are offered in Section 5.7.

5.2 MIMO Transmission Model

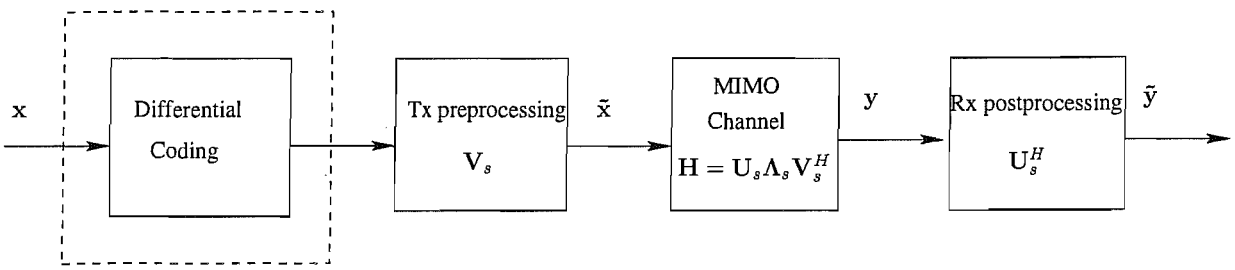


Figure 5.1: Schematic of MIMO eigenmode transmission

Consider a system having M_T transmitter and M_R receiver antennas subjected to a flat-fading channel between any pair of transmitter and receiver antennas. Then the received M_R-dimensional

symbol vector $\mathbf{y}$ can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{5.1}$$

where $\mathbf{x}$ is the M_T-dimensional transmitted symbol vector and $\mathbf{H}$ is an $(M_R \times M_T)$-dimensional complex channel matrix with the (i,j)th element being the fading channel between the ith receive and jth transmit antennas. Finally, $\mathbf{n}$ is the M_R-dimensional AWGN vector having a zero-mean and $E(\mathbf{n}\mathbf{n}^H) = \sigma_n^2 \mathbf{I}_{M_R}$. Here $\mathbf{I}_M$ is an $(M \times M)$-dimensional identity matrix.

If the rank of $\mathbf{H}$ is assumed to be q ($q \leq min(M,N)$), the SVD of the channel matrix $\mathbf{H}$ is given by

$$\begin{aligned} \mathbf{H} &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H \\ &= [\mathbf{U}_s \, \mathbf{U}_n] \begin{bmatrix} \mathbf{\Lambda}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}, \end{aligned} \tag{5.2}$$

where $\mathbf{U}$ is an $(M_R \times M_R)$-dimensional unitary matrix satisfying $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{M_R}$ and $\mathbf{V}$ is an $(M_T \times M_T)$-dimensional unitary matrix having the property of $\mathbf{V}^H\mathbf{V} = \mathbf{I}_{M_T}$, while $\mathbf{\Lambda}$ is an $(M_R \times M_T)$-dimensional matrix and $\mathbf{I}_M$ is an $(M \times M)$-dimensional identity matrix. In the second line of (5.2), $\mathbf{\Lambda}_s$ is a $(q \times q)$-dimensional diagonal matrix having diagonal elements of $\lambda_1 \geq \lambda_2 \cdots \lambda_{q-1} \geq \lambda_q$, which are the singular values of $\mathbf{H}$. Furthermore, in (5.2) we portrayed $\mathbf{U}$ and $\mathbf{V}$ in form of two components, where $\mathbf{U}_s$ is an $(M_R \times q)$-dimensional matrix constituted by the first p columns of $\mathbf{U}$, which span the column-space of $\mathbf{H}$ and $\mathbf{V}_s$ is an $(M_T \times p)$-dimensional matrix formed by the first q columns of $\mathbf{V}$, which span the row-space of $\mathbf{H}$. Still referring to (5.2), $\mathbf{U}_n$ is an $[M_R \times (M_R - q)]$-dimensional matrix, which is orthogonal to $\mathbf{U}_s$, while spanning the null space of $\mathbf{H}$ and $\mathbf{V}_n$ is an $[M_T \times (M_T - q)]$-dimensional matrix that is orthogonal to $\mathbf{V}_s$ and spans the left null space of $\mathbf{H}$.

If the channel matrix $\mathbf{H}$ is known at both the transmitter and receiver, the so-called eigenmode transmission regime of [64] shown in Fig.5.1 can be invoked to decompose the MIMO channel into orthogonal subchannels by applying $\mathbf{V}_s$ and $\mathbf{U}_s$ at the transmitter and receiver, respectively, yielding

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{U}_s^H \mathbf{y} \\ &= \mathbf{U}_s^H (\mathbf{H}\mathbf{V}_s \tilde{\mathbf{x}} + \mathbf{n}) \\ &= \mathbf{\Lambda}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \end{aligned} \tag{5.3}$$

where $\tilde{\mathbf{x}}$ is a q-dimensional transmitted symbol vector, while $\tilde{\mathbf{n}} = \mathbf{U}_s^H \mathbf{n}$ is a q-dimensional noise

vector, which has the same statistical properties as $\mathbf{n}_{(dl,k)}$, because $\mathbf{U}_s^H$ is a unitary matrix.

As an explicit benefit of using the SVD, the known channel matrix $\mathbf{H}$ is finally decomposed into q independent orthogonal subchannels, each of which has a channel gain of λ_i and this transmit preprocessing regime is referred to as eigenmode transmission [64].

As a further simplication, it was shown in [130] that high-integrity reception can be achieved, if we opt for transmitting in a limited number of p ($1 \leq p \leq q$) subchannels having channel gains of $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p$ for achieving a high throughput, while meeting the specific target BER performance.

Another potential advantage of eigenmode transmission is that only the left singular vectors of $\mathbf{U}_s$ and the right singular vectors of $\mathbf{V}_s$ are needed, as we can see in (5.3). Hence it is intuitively appealing to invoke algorithms, which estimate or update the singular vectors only [130–135] instead of estimating the entire MIMO channel matrix $\mathbf{H}$ and then additionally implementing the SVD, which would inevitably impose a high computational complexity.

5.3 TDD MIMO Transmission Model

A MIMO link may be created using either FDD or TDD mode. In this chapter, we assume employing the TDD mode. Consider a TDD system using M_T antennas at the base station (BS) and M_R antennas at the mobile station (MS), encountering a flat-fading channel between any pair of transmitter and receiver antennas. Furthermore, for simplicity, we assume that the system supports a single user. Then the M_R-dimensional received symbol vector $\mathbf{y}_{dl}(k)$ of the DL and the M_T-dimensional received symbol vector $\mathbf{y}_{ul}(k)$ of the UL can be expressed as

$$\mathbf{y}_{dl}(k) = \mathbf{H}_{dl}(k)\mathbf{x}_{dl}(k) + \mathbf{n}_{dl}(k), \tag{5.4}$$

$$\mathbf{y}_{ul}(k) = \mathbf{H}_{ul}(k)\mathbf{x}_{ul}(k) + \mathbf{n}_{ul}(k), \tag{5.5}$$

where $\mathbf{x}_{dl}(k)$ is an M_T-dimensional DL symbol vector transmitted from the BS to the MS, while $\mathbf{x}_{ul}(k)$ is an M_R-dimensional UL symbol vector transmitted from the MS to the BS. Furthermore, $\mathbf{H}_{dl}(k)$ is the DL channel matrix and $\mathbf{H}_{ul}(k)$ is the UL channel matrix. Moreover, $\mathbf{n}_{dl}(k)$ is the DL AWGN noise vector having a zero-mean and $E(\mathbf{n}_{dl}\mathbf{n}_{dl}^H) = \sigma_{\mathbf{n}_{dl}}^2 \mathbf{I}_{M_R}$, $\mathbf{n}_{ul}(k)$ is the UL AWGN noise vector having a zero-mean and $E(\mathbf{n}_{ul}\mathbf{n}_{ul}^H) = \sigma_{\mathbf{n}_{ul}}^2 \mathbf{I}_{M_T}$.

Since the UL and DL timeslots of a TDD link are transmitted on the same carrier frequency, the

UL and DL channel matrices may be assumed to be identical, provided that the Doppler frequency is sufficiently low and hence the corresponding channel impulse response (CIR) does not change dramatically during the time between the UL and DL time slot. If this is not the case,because for example FDD is used, explicit CIR signalling has to be used [144]. Hence we have

$$\mathbf{H}_{ul}(k) = \mathbf{H}_{dl}^{T}(k). \tag{5.6}$$

Upon substituting (5.6) into (5.5), we arrive at

$$\mathbf{y}_{ul}(k) = \mathbf{H}_{dl}^{T}(k)\mathbf{x}_{ul}(k) + \mathbf{n}_{ul}(k). \tag{5.7}$$

The transmitted symbol vector $\mathbf{x}_{ul}$ is conjugated before transmission, as proposed in [130]. In this case, we obtain

$$\mathbf{y}_{ul}(k) = \mathbf{H}_{dl}^{T}(k)\mathbf{x}_{ul}^{*}(k) + \mathbf{n}_{ul}(k). \tag{5.8}$$

Furthermore, the received symbol vector is conjugated as well, hence we have

$$\mathbf{y}_{ul}^{*}(k) = \mathbf{H}_{dl}^{H}(k)\mathbf{x}_{ul}(k) + \mathbf{n}_{ul}^{*}(k). \tag{5.9}$$

According to (5.2), the SVD of $\mathbf{H}_{dl}$ can be expressed as

$$\mathbf{H}_{dl} = [\mathbf{U}_{dl_s}\ \mathbf{U}_{dl_n}] \begin{bmatrix} \mathbf{\Lambda}_{dl_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{dl_s}^{H} \\ \mathbf{V}_{dl_n}^{H} \end{bmatrix}, \tag{5.10}$$

where $\mathbf{U}_{dl_s}$ is an $(M_R \times q)$-dimensional unitary matrix, while $\mathbf{V}_{dl_s}$ is an $(M_T \times q)$-dimensional unitary matrix. Furthermore, $\mathbf{\Lambda}_{dl_s}$ is a $(q \times q)$-dimensional diagonal matrix with its diagonal elements given by $\lambda_1 \geq \lambda_2 \cdots \lambda_{q-1} \geq \lambda_q$, which are the singular values of $\mathbf{H}_{dl}$. Accordingly, the SVD of $\mathbf{H}^{H}$ is given by

$$\mathbf{H}_{dl}^{H} = [\mathbf{V}_{dl_s}\ \mathbf{V}_{dl_n}] \begin{bmatrix} \mathbf{\Lambda}_{dl_s} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{U}_{dl_s}^{H} \\ \mathbf{U}_{dl_n}^{H} \end{bmatrix}. \tag{5.11}$$

When eigenmode transmission is used for the sake of avoiding interference among the transmitted data symbols, the p-dimensional transmitted symbol vectors $\tilde{\mathbf{x}}_{dl}$ and $\tilde{\mathbf{x}}_{ul}$ are multiplied by $\mathbf{V}_{dl_{sp}}$ and $\mathbf{U}_{dl_{sp}}$ given by the first p columns of $\mathbf{V}_{dl_s}$ and $\mathbf{U}_{dl_s}$, respectively, before their transmission.

According to (5.3), we obtain

$$\mathbf{y}_{dl}(k) = \mathbf{H}_{dl}(k)\mathbf{V}_{dl_{sp}}\tilde{\mathbf{x}}_{dl} + \mathbf{n}_{dl}(k), \tag{5.12}$$

$$\mathbf{y}_{ul}^{*}(k) = \mathbf{H}_{dl}^{H}(k)\mathbf{U}_{dl_{sp}}\tilde{\mathbf{x}}_{ul} + \mathbf{n}_{ul}(k). \tag{5.13}$$

The resultant received symbol vectors $\mathbf{y}_{dl}$ and $\mathbf{y}_{ul}^{*}$ are multiplied by the matrices $\mathbf{U}_{dl_{sp}}^{H}$ and $\mathbf{V}_{dl_{sp}}^{H}$, respectively, for the sake of avoiding interference among the transmitted data symbols. Finally, we obtain

$$\begin{aligned}\tilde{\mathbf{y}}_{dl}(k) &= \mathbf{U}_{dl_{sp}}^{H}(k)\mathbf{y}_{dl}(k) \\ &= \mathbf{\Lambda}_{dl_p}(k)\tilde{\mathbf{x}}_{dl}(k) + \mathbf{U}_{dl_{sp}}^{H}(k)\mathbf{n}_{dl}(k),\end{aligned} \tag{5.14}$$

$$\begin{aligned}\tilde{\mathbf{y}}_{ul}(k) &= \mathbf{V}_{dl_{sp}}^{H}(k)\mathbf{y}_{ul}^{*}(k) \\ &= \mathbf{\Lambda}_{dl_p}(k)\tilde{\mathbf{x}}_{ul}(k) + \mathbf{V}_{dl_{sp}}^{H}(k)\mathbf{n}_{ul}(k),\end{aligned} \tag{5.15}$$

where $\mathbf{\Lambda}_{dl_p}$ is a $(p \times p)$-dimensional diagonal matrix having $\lambda_1 \geq \lambda_2 \cdots \lambda_{p-1} \geq \lambda_p$ as its diagonal elements. As we can see, only the matrix $\mathbf{U}_{dl_{sp}}$ has to be known at the MS, while the matrix $\mathbf{V}_{dl_{sp}}$ is used for preprocessing at the BS.

The matrices $\mathbf{U}_{dl_{sp}}$ and $\mathbf{V}_{dl_{sp}}$ can be obtained by SVD of the channel matrix $\mathbf{H}_{dl}$. However, this requires estimating the channel matrix first, then implementing the SVD, which imposes a high computational complexity. Observe in (5.9) to (5.15) however, that only the subspace matrices $\mathbf{U}_{dl_{sp}}$ and $\mathbf{V}_{dl_{sp}}$ are required instead of the knowledge of the entire channel matrix.

Let us continue by considering the DL transmission in more detail. More explicitly, our goal is to obtain the matrices $\mathbf{U}_{dl_{sp}}$ and $\mathbf{V}_{dl_{sp}}$ without estimating the channel matrix $\mathbf{H}$ and without performing the SVD of $\mathbf{H}$. Upon substituting (5.10) into (5.12), we obtain

$$\mathbf{y}_{dl} = [\mathbf{U}_{dl_s}\ \mathbf{U}_{dl_n}] \begin{bmatrix} \mathbf{\Lambda}_{dl_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{dl_s}^{H} \\ \mathbf{V}_{dl_n}^{H} \end{bmatrix} \mathbf{V}_{dl_{sp}}\tilde{\mathbf{x}}_{dl} + \mathbf{n}_{dl}. \tag{5.16}$$

The autocorrelation matrix of the vector $\mathbf{y}_{dl}$ of received symbols is given by

$$\mathbf{R}_{\mathbf{y}_{dl}} = E[\mathbf{y}_{dl}\mathbf{y}_{dl}^H] = \mathbf{H}_{dl}\mathbf{V}_{dl_{sp}}\mathbf{R}_{\tilde{\mathbf{x}}_{dl}}\mathbf{V}_{dl_{sp}}^H\mathbf{H}_{dl}^H + \sigma_{\mathbf{n}_{dl}}^2\mathbf{I}. \tag{5.17}$$

Let the total average transmit power P be a constant and let us allocate an equal power to each nonzero subchannel in (5.16). Then we obtain the autocorrelation of the p-dimensional vector $\tilde{\mathbf{x}}_{dl}$ of transmitted symbols as follows

$$\mathbf{R}_{\tilde{\mathbf{x}}_{dl}} = E[\tilde{\mathbf{x}}_{dl}\tilde{\mathbf{x}}_{dl}^H] = \frac{P}{p}\mathbf{I}_p. \tag{5.18}$$

Hence, following a few further manipulations, (5.17) can be written as

$$\mathbf{R}_{\mathbf{y}_{dl}} = [\mathbf{U}_{dl_s}\ \mathbf{U}_{dl_n}] \begin{bmatrix} \frac{P}{p}\mathbf{\Lambda}_{dl_{sp}}^2 + \sigma_{\mathbf{n}_{dl}}^2 & \mathbf{0} \\ \mathbf{0} & \sigma_{\mathbf{n}_{dl}}^2 \end{bmatrix} \begin{bmatrix} \mathbf{U}_{dl_s}^H \\ \mathbf{U}_{dl_n}^H \end{bmatrix}, \tag{5.19}$$

where $\mathbf{U}_{dl_s}$ is constituted by q eigenvectors of $\mathbf{R}_{\mathbf{y}_{dl}}$ associated with the q largest eigenvalues $(\frac{P}{p}\lambda_1^2 + \sigma_{\mathbf{n}_{dl}}^2) \geq (\frac{P}{p}\lambda_2^2 + \sigma_{\mathbf{n}_{dl}}^2) \cdots \geq (\frac{P}{p}\lambda_q^2 + \sigma_{\mathbf{n}_{dl}}^2)$ of $\mathbf{R}_{\mathbf{y}_{dl}}$. The space spanned by the columns of $\mathbf{U}_{dl_s}$ is referred to as the signal subspace, while $\mathbf{U}_{dl_n}$ consists of $(M_R - q)$ number of eigenvectors of $\mathbf{R}_{\mathbf{y}_{dl}}$ related to $(M_R - q)$ number of eigenvalues $\{\sigma_{\mathbf{n}_{dl}}^2\}$ of $\mathbf{R}_{\mathbf{y}_{dl}}$. Finally, the space spanned by the columns of $\mathbf{U}_{dl_n}$ is termed as the noise subspace, which is orthogonal to the signal subspace [131, 141].

We can see from our discussions above that the eigenvectors in $\mathbf{U}_{dl_s}$ also consist of the orthonormal basis vector of the column-space of $\mathbf{H}_{dl}$. Moreover, when the vector $\mathbf{y}_{dl}$ of received symbols becomes available, so-called subspace tracking algorithms [140, 145, 146] can be used to track the orthonormal basis vectors of $\mathbf{U}_{dl_{sp}}$, which spans the column-space of $\mathbf{H}$.

Similarly, when the vector $\mathbf{y}_{ul}^*$ of received symbols becomes available, the eigenvectors in $\mathbf{V}_{dl_{sp}}$ can be tracked as well, which spans the row-space of $\mathbf{H}$. Upon obtaining the corresponding left and right singular vectors of $\mathbf{H}$, the eigenmode MIMO-aided transmission regime described above can be employed.

In the family of different subspace tracking algorithms, the Projection Approximation Subspace Tracking technique using deflation (PASTD) [140] stands out as one of the most popular algorithms. In the next section, the PASTD algorithm will be briefly described in the context of tracking the elements of $\mathbf{U}_{dl_{sp}}$. The same algorithm can also be used for tracking the elements of $\mathbf{V}_{dl_{sp}}$.

5.4 PASTD Subspace Tracking [140]

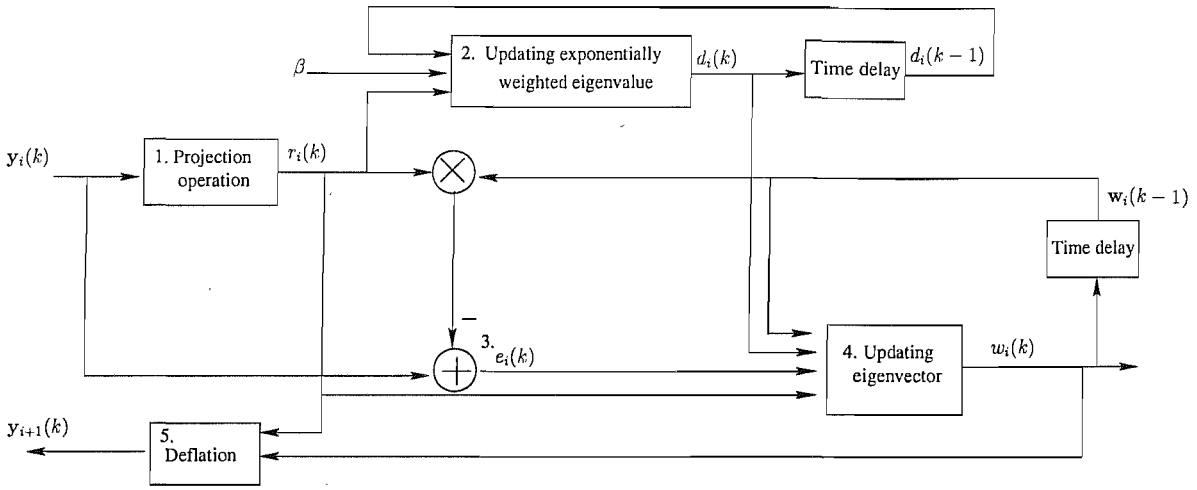


Figure 5.2: Schematic of PASTD subspace tracking

The PASTD algorithm of [140] designed for signal subspace tracking is demonstrated in Fig.5.2 and summarized in Table 5.1, where $\mathbf{y}_{dl}(k)$ is the kth M_R-dimensional received signal vector generated for DL transmission, while $d_i(k)$ represents the exponentially weighted estimate of the ith eigenvalue and $\mathbf{w}_i(k)$ denotes the estimate of the ith eigenvector at the kth time instant. Furthermore, β ($0 < \beta \leq 1$) represents the forgetting factor. Table 5.1 summarizes the operations of the PASTD algorithm, which is based on the so-called deflation technique [140] and its basic philosophy is that of the sequential estimation of the so-called principal components [140]. The most dominant eigenvector is updated first by applying the PAST algorithm at the 1st iteration [140]. Then the projection of the current signal sample vector $\mathbf{y}_{dl}(k)$ onto this eigenvector is removed from $\mathbf{y}_{dl}(k)$ itself. Now the second most dominant eigenvector becomes the most dominant one in the updated signal vector and hence can be extracted in the same way as outlined above. This procedure is applied repeatedly, until all desired eigencomponents have been estimated.

Since the deflation technique results in a strong loss of orthonormality between the singular vectors [140], the Gram-Schmidt orthonormalization [81] technique is invoked for reorthogonalizing the signal subspace after each update.

The variables $d_i(0)$ and $\mathbf{w}_i(0)$ have to be initialized, as seen in Table 5.1 . Specifically, the SVD of the first M vectors of the received symbols are used for the initialization of $d_i(0)$ and $\mathbf{w}_i(0)$ [141].

Since the singular vector generated according to (5.2) can be different up to a complex-valued

Operation procedure of PASTD algorithm
$\mathbf{y}_1(k) = \mathbf{y}_{dl}(k)$
For $i = 1, 2, \cdots, p$, Do
$r_i(k) = \mathbf{w}_i^H(k-1)\mathbf{y}_i(k)$; prejection operation
$d_i(k) = \beta d_i(k-1) + \lvert r_i \rvert^2$;
$\mathbf{e}_i(k) = \mathbf{y}_i(k) - \mathbf{w}_i(k-1)r_i(k)$;
$\mathbf{w}_i(k) = \mathbf{w}_i(k-1) + \mathbf{e}_i(k)[r_i^*(k)/d_i(t)]$; updating eigenvectors
$\mathbf{y}_{i+1}(k) = \mathbf{y}_i(k) - \mathbf{w}_i(k)r_i(k)$; deflation

Table 5.1: The PASTD algorithm designed for tracking the signal subspace components of the received signal vector $\mathbf{y}_{dl}$

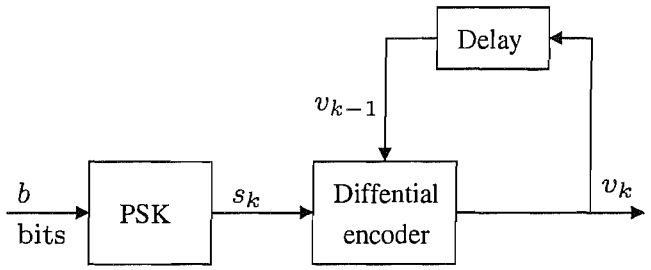


Figure 5.3: Transmitter block diagram of a classic DPSK scheme.

coefficient of unit norm [130], it may cause phase ambiguity [130], which can be resolved for example by differential encoding, leading to differential phase shift keying (DPSK) modulation [130].

5.5 Differential Encoding

In this section we consider the employment of differentil encoding. The shematic of a DPSK transmitter is shown in Figure.5.3. As seen in Figure 5.3, the symbol v_k transmitted at time instant k is obtained from $v_k = x_k v_{k-1}$, where x_k is a PSK modulated symbol and v_{k-1} is the symbol transmitted at time instant $(k-1)$ [147].

If the channel gain between the transmitter and receiver is h and the noise is n_k, the received signal can be expressed as

$$r_k = hv_k + n_k. \tag{5.20}$$

If the channel gain h may be assumed to be constant for the symbol intervals $(k-1)$ and k, the optimal estimate of x_k is to find the specific symbol of the M-PSK constellation, which meets the requirement of [148]

Number of transmitter antennas M_T	4
Number of receiver antennas M_R	4
Normalized maximum Doppler frequency $f_{dm}T_s$	0.0001
Forgetting factor β in Section 5.4	0.99

Table 5.2: Parameters for the PASTD algorithm in TDD mode for $f_{dm}T_s = 0.0001$

Number of transmitter antennas M_T	4
Number of receiver antennas M_R	4
Normalized maximum Doppler frequency $f_{dm}T_s$	0.001
Forgetting factor β in Section 5.4	0.95

Table 5.3: Parameters for the PASTD algorithm in TDD mode for $f_{dm}T_s = 0.001$

$$\widehat{x} = argmin_{x_k}|r_k r_{k-1}^* - x_k|, \tag{5.21}$$

where we have

$$\begin{aligned} r_k r_{k-1}^* &= |h|^2 v_k v_{k-1} + h v_k n_{k-1}^* + n_k h^* v_{k-1}^* + n_k n_{k-1}^* \\ &= |h|^2 x_k + N, \end{aligned} \tag{5.22}$$

with N being the Gaussian noise.

5.6 Performance of Subspace Tracking Based Blind Transmit Preprocessing

Having described the TDD system and the PASTD algorithms in Section 5.4, in this section our simulation results are provided in order to characterize the attainable performance of PASTD subspace tracking in the context of TDD system. Furthermore, differential BPSK modulation is used.

In Figures 5.4 and 5.5, the achievable BER performance against the forgetting factor β in Section 5.4 when only the largest eigenvalue is used for both the uplink and downlink is plotted, respectively, at SNRs of -10dB, -5dB and 0dB. The remaining parameters are assumed to be the same as in Table 5.2. We can see from Figures 5.4 and 5.5 that for a given SNR, the BER decreases only silghtly upon increasing the forgetting factor β. This is because for the low normalized Doppler frequency of $f_{dm}T_s = 0.0001$, the channel exhibits a high correlation for a long period, which allows us to exploit

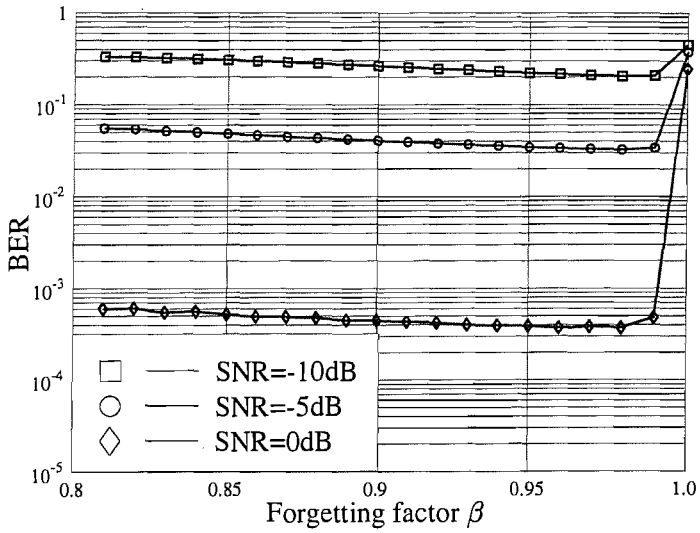


Figure 5.4: BER performance against the forgetting factor β in Section 5.4 when only the largest eigenvalue is used for uplink transmission at different values of SNR. The remaining parameters are assumed to be the same as in Table 5.2.

the channel knowledge over a longer period, resulting in a higher forgetting factor. In case of $\beta = 1$, all the past channel output samples are invoked. In case of an infinite memory, the correlation between a far distant channel sample and current one is low and therefore the effects of the noise imposed by a distant noisy sample on correlation becomes more dominant, which actually degrades the algorithm's performance. Furthermore, we observe that for differnent values of SNR, the optimum forgetting factor β may be different. For SNR=$-$10dB and $-$5dB the optimum of 0.99 is sligtly higher than the 0.98 value recorded for SNR=0dB. The reason behind this may be attributed to the observation that for lower SNRs a higher number of noisy samples may be needed to mitigate the effects of the noise and hence a higher forgetting factor is required. By contrast, for higher SNRs a lower number of noisy samples is sufficient for mitigating the effects of noise, which results in a lower forgetting factor β.

In Figures 5.6 and 5.7, the achievable BER performance is plotted against the forgetting factor β, when only the largest eigenvalue is used for both uplink and downlink transmissions , respectively, at SNRs of $-$10dB,$-$5dB and 0dB. The remaining parameters are the same as in Table 5.3. We can see from Figures 5.6 and 5.7 that for a given SNR, the BER slightly decreases upon increasing the

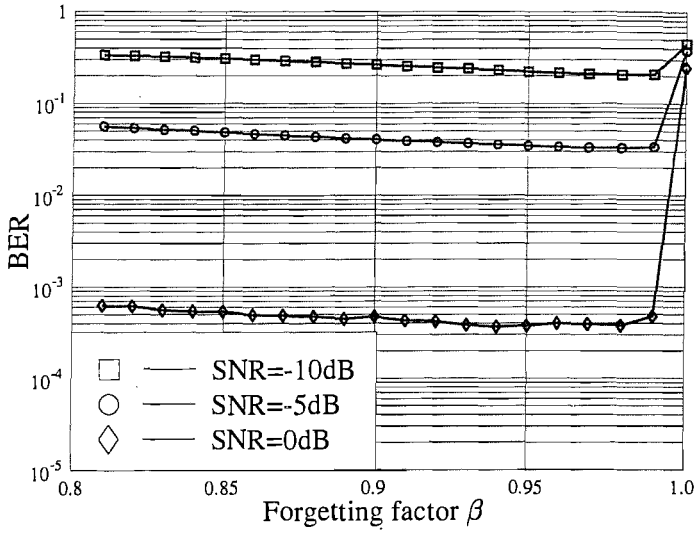


Figure 5.5: BER performance against the forgetting factor β in Section 5.4 when only the largest eigenvalue is used for downlink transmission at different values of SNR. The remaining parameters are assumed to be the same as in Table 5.2.

forgetting factor β, until an optimum point is reached, which is lower than that in the case of $f_{dm}T_s = 0.0001$. Beyond this point the BER increases relatively sharply upon increasing the forgetting factor β. This is because for the normalized Doppler frequency $f_{dm}T_s = 0.001$ the channel varies more rapidly than for $f_{dm}T_s = 0.001$. Hence the correlation between the channel samples decays faster, which implies that a lower forgetting factor β is needed. We can also see for SNR=-10dB and -5dB that the optimum forgetting factor is around $\beta = 0.95$, while for SNR=0dB it is around 0.90. The reason for this is observation the same as for the case of $f_{dm}T_s = 0.0001$. Furthermore, since a lower forgetting factor results in a faster convergence, we can opt for a lower forgetting factor under the constraint of meeting the target BER for the sake of achieving a more rapid convergence.

In Figure 5.8 the attainable BER performance is portrayed for different values of the SNR, when only the largest eigenvalue is used for uplink transmission. The remaining parameters are the same as in Table 5.2. We can see from Figure 5.8 that the achievable BER performance of PASTD subspace tracking is similar to that achieved with the aid of perfect channel knowledge. Observe, however that the BER difference between the perfect estimation based scenario and the tracked scenario becomes higher upon increasing the SNR. This is because the forgetting factor of $\beta = 0.99$ is not the optimum

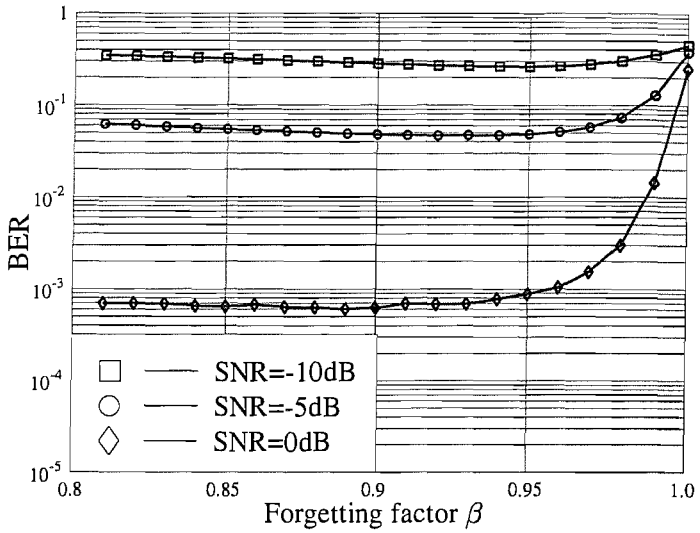


Figure 5.6: BER performance against the forgetting factor β in Section 5.4 when only the largest eigenvalue is used for uplink transmission at different values of SNR. The remaining parameters are assumed to be the same as in Table 5.3.

for higher SNRs, as seen earlier in Figure 5.4. The same performance is observed for downlink transmission because the uplink and downlink channels are similar. Furthermore, the same phenomenon is confirmed in Figure 5.9.

In Figure 5.10 the attainable mean BER performance is plotted against the forgetting factor β introduced in Section 5.4, when the first two largest eigenvalues are used for UL transmissions. The results were plotted for different values of the SNR. The remaining parameters are the same as in Table 5.2. We can see from Figure 5.10 that different optimum forgetting factors are found for the different values of the SNR. The reason for this observation is the same as that stated earlier for Figure 5.4 Furthermore, we observe the same phenomenon from Figure 5.11.

In Figure 5.12 the attainable BER performance is characterized for different values of the SNR, when the first two eigenvalues are used for uplink transmission. The remaining parameters are the same as in Table 5.2. We can see from Figure 5.12 that the performance recorded for the largest eigenvalue is the same as in Figure 5.4 and when the availability of perfect channel knowledge is assumed, while the performances achieved by the PATSD algorithm are quite different from those seen in Figure 5.4. This is because the subspace tracking algorithm is unable to eliminate the interferece

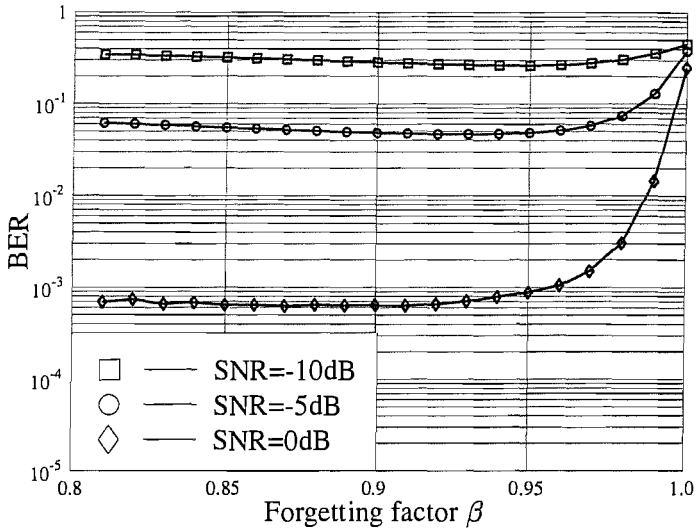


Figure 5.7: BER performance against the forgetting factor β in Section 5.4 when only the largest eigenvalue is used for downlink transmission at different values of SNR. The remaining parameters are assumed to be the same as in Table 5.3.

between the two eigenvalues, while there is no interference between the two eigenvalues in case of perfect channel knowledge. Furthermore, the BER curve exhibits a floor value upon increasing the SNR. This is because the interference between two eigenvalues becomes the dominant factor for high values of the SNR. The same phenomenon is observed in Figure 5.13.

5.7 Summary and Conclusions

In this chapter, PASTD subspace tracking aided MIMO transmit processing techniques were investigated in the context of a TDD system. Since only the left or right singular vectors of the channel matrix are required at transmitter and receiver, respectively, for eigenmode transmission in the TDD mode, PASTD subspace tracking can be used at both the transmitter and receiver to acquire the required left and right singular vectors without estimating the entire MIMO channel matrix $\mathbf{H}$. This operation is followed by SVD of $\mathbf{H}$, which typically results in a high complexity. Furthermore, since the PASTD subspace tracking technique is a blind algorithm, it improves the achievable spectral efficiency. A specific deficiency of the family of subspace tracking algorithms is their phase ambiguity imposed by the the non-unique nature of SVD, which was resolved by employing differential encod-

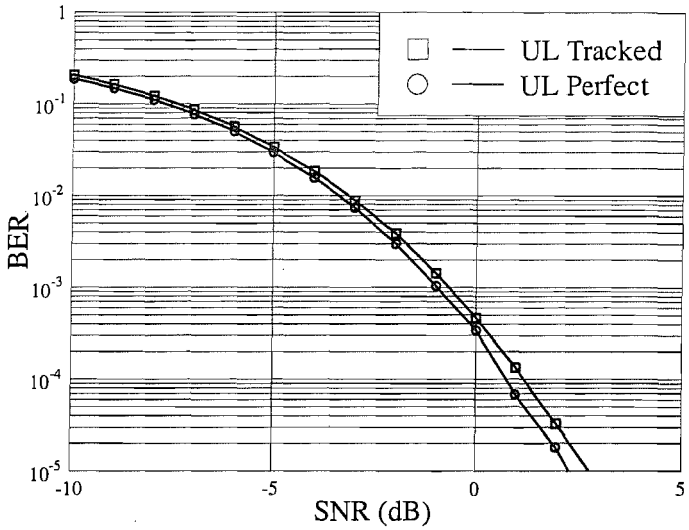


Figure 5.8: BER versus SNR performance, when only the largest eigenvalue is used for uplink transmission. The remaining parameters are be the same as in Table 5.2.

ing. Finally, the efficiency of the proposed scheme was verified by our simulations.

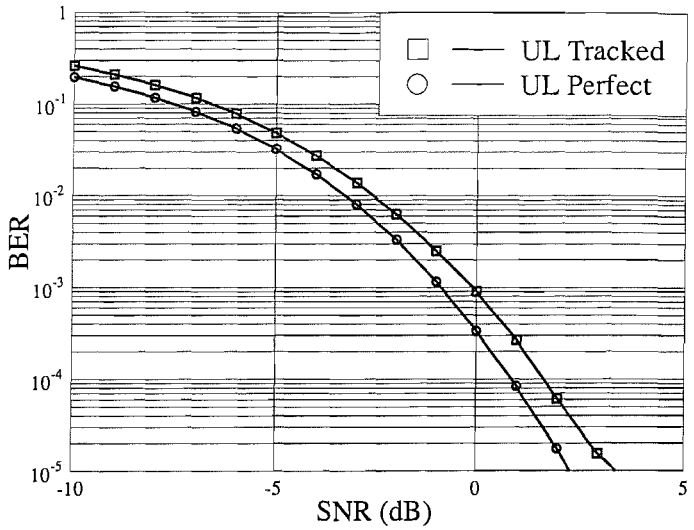


Figure 5.9: BER versus SNR performance, when only the largest eigenvalue is used for uplink transmission. The remaining parameters are the same as in Table 5.3.

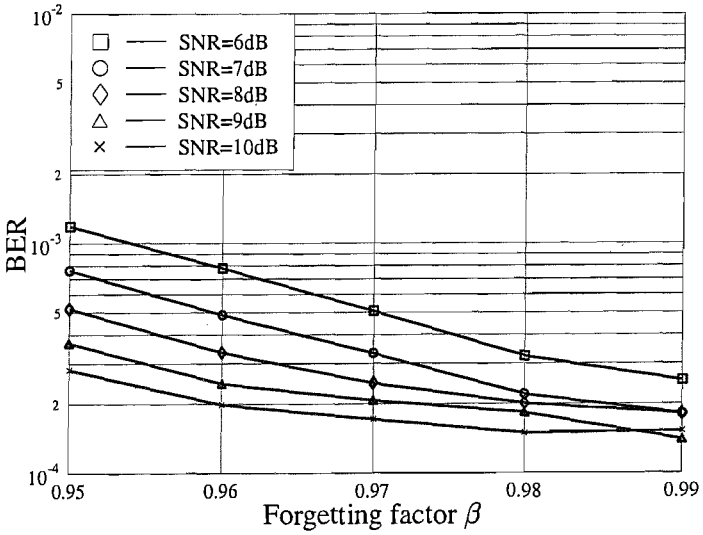


Figure 5.10: Mean BER performance against the forgetting factor β in Section 5.4, when the first two largest eigenvalues are used for uplink transmission at different SNRs. The remaining parameters are the same as in Table 5.2.

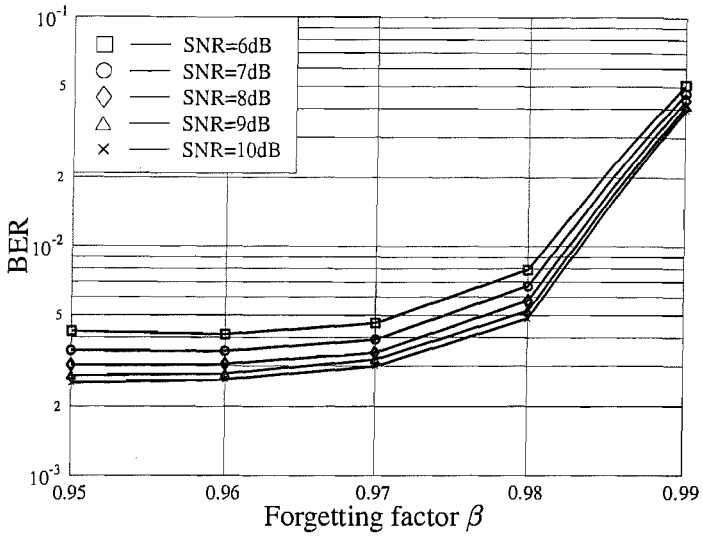


Figure 5.11: Mean BER performance against the forgetting factor β in Section 5.4, when the first two largest eigenvalues are used for uplink transmission at different SNRs. The remaining parameters are the same as in Table 5.3.

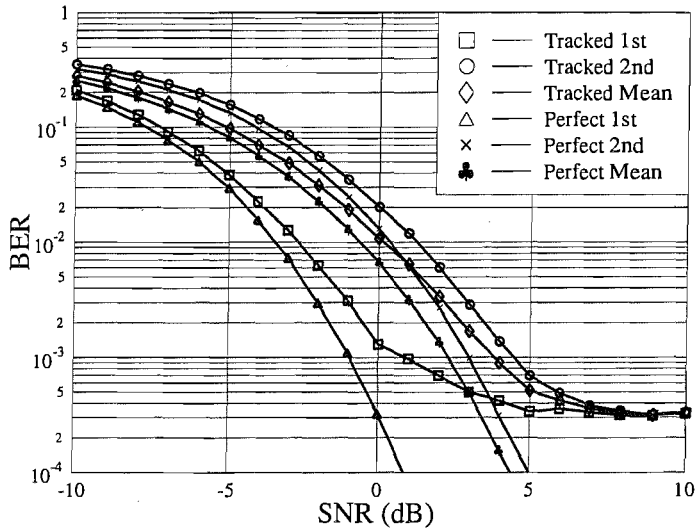


Figure 5.12: BER versus SNR performance, when the first two largest eigenvalues are used for uplink transmission. The remaining parameters are the same as in Table 5.2.

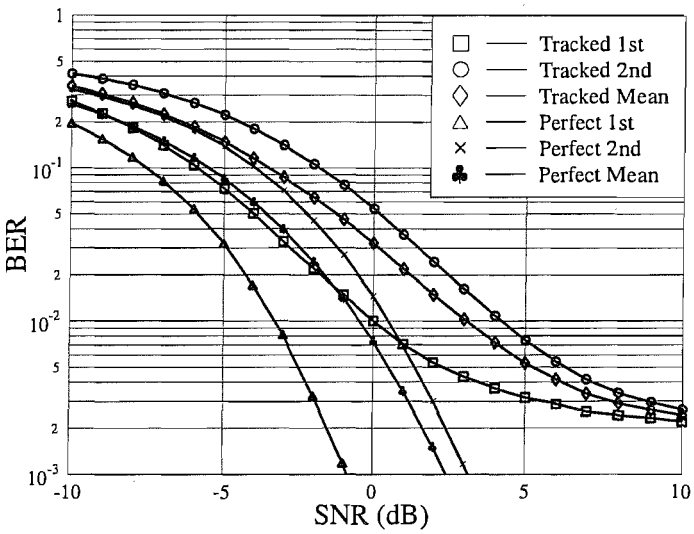


Figure 5.13: BER versus SNR performance, when the first two eigenvalues are used for uplink transmission. The remaining parameters are the same as in Table 5.3.

Chapter 6

Channel Prediction Based Multiuser Transmission in SDMA

6.1 Introduction

In the previous chapter we mainly focussed our attention on point to point communications systems, albeit practical real communications systems have to support multiple users [149, 150]. Diverse schemes have been proposed for supporting multiple users, including Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA) [94], Space Division Multiple Access (SDMA) [5] and so on. In CDMA systems, each user is assigned a unique user specific signature or spreading code, in order to differentiate them from the others, while in SDMA systems, the unique user specific spatial signature represented by the channel impulse response (CIR) acts like the unique spreading code of a CDMA system.

Recently, transmitter preprocessing techniques implemented at the base-station (BS) have received wide attention [151–155], since they require a simple receiver at the mobile station (MS). Porvided that the channel impulse response (CIR) of all the BS to mobile station (MS) links is known in advance - even before the signal's transmission - it is plausible that the different users' signals may be differentiated with the aid of their unique, user-specific downlink CIRs. Naturally, this non-causal CIR knowledge is unavailable in practice. Hence a natural design option is to estimate the CIRs at the receiver after the BS's signal was received and convey it using side-information to the BS for its future use. Naturally, the resultant CIR has to be quantized before its transmission. In addition to this quantization error, it also becomes outdated, which may seriously degrade the attainable perfor-

mance [156], and both imperfections result in an erosion of the achievable transmit preprocessing gain expressed in terms of either the attainable transmit power reduction or the number of users that may be supported. Another attractive design option is to avoid the CIR-signalling latency by invoking the previously received CIRs for predicting their future evolution using CIR-tap prediction. To elaborate a little further, explicit CIR signalling may be used in Frequency Division Duplex (FDD) systems, where the uplink and downlink operate at different carrier frequencies. By contrast, in Time Division Duplex (TDD) systems, the uplink and downlink signals are transmitted at the same carrier frequencies. Hence these signals experience similar CIRs as well as frequency domain channel transfer functions, unless their bandwidth is wider than the coherence bandwidth of the channel. The employment of a TDD mode is assumed in this chapter.

In the context of MIMO channel prediction algorithms [37, 143, 157–160], the vector Kalman filtering assisted MIMO channel predictor has been successfully employed for tracking and predicting the MIMO channel [161]. Hence this technique is adopted in this chapter, in order to facilitate BS transmitter preprocessing in the context of downlink TDD transmissons.

The outline of the chapter is as follows. In Secion 6.2, the philosophy of SDMA downlink transmission using preprocessing is described. In Section 6.3 the MMSE preprocessing criterion is investigated. In Section 6.4, SDMA uplink transmissions are reviewed briefly, while in Section 6.5 Kalman filtering assisted channel prediction is discussed. In Section 6.6, the existing third-generation TDD standard is highlighted and in Section 6.7 our simulation results are provided and analysed. Finally, our conclusions are offered in Section 6.8.

6.2 SDMA Downlink Transmission Model

Consider a system having a single BS and supporting K MSs, as shown in Fig.6.1. The BS has M transmitter antennas and the kth MS has $N_k \geq 1$ receiver antennas. Furthermore, the channel between any pair of transmitter and receiver antennas is assumed to be flat-fading. The N_k-dimensional symbol vector $\mathbf{x}_{(dl,k)}$ is transmitted from the BS to the kth MS, which can be expressed as

$$\mathbf{x}_{(dl,k)} = [x_{(dl,k1)}, x_{(dl,k2)}, \cdots, x_{(dl,kN_k)}]^T. \tag{6.1}$$

Before $\mathbf{x}_{(dl,k)}$ is transmitted, it is multiplied by the $(M \times N_k)$-dimensional transmit preprocessing matrix $\mathbf{P}_k$. Hence the M-dimensinonal preprocessed data vector $\mathbf{d}_{(dl,k)}$ is destined for the kth user,

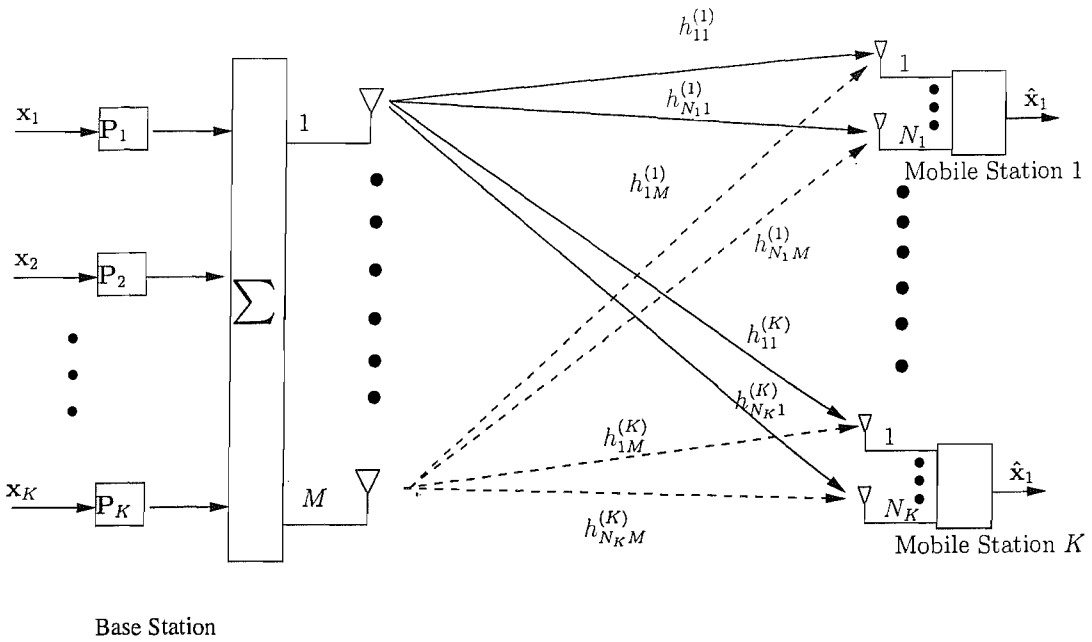


Figure 6.1: Schematic of the SDMA downlink transmitter

which is given by

$$\mathbf{d}_{(dl,k)} = \mathbf{P}_k \mathbf{x}_{(dl,k)}. \tag{6.2}$$

Hence the M-dimensional composite preprocessed data vector $\mathbf{d}_{dl}$ of the K users is given by

$$\begin{aligned} \mathbf{d}_{dl} &= \sum_{k=1}^{K} \mathbf{d}_{(dl,k)} \\ &= \sum_{k=1}^{K} \mathbf{P}_k \mathbf{x}_{(dl,k)} \\ &= \mathbf{P}\mathbf{x}_{dl}, \end{aligned} \tag{6.3}$$

where $\mathbf{P}$ is a $(M \times \mathcal{U})$-dimensional matrix and $\mathcal{U} = \sum_{k=1}^{K} N_k$, which is given by

$$\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_K], \tag{6.4}$$

while $\mathbf{x}_{dl}$ is a $\mathcal{U}$-dimensional transmitted symbol vector, which has the structure of

$$\mathbf{x}_{dl} = [\mathbf{x}_{(dl,1)}^T, \mathbf{x}_{(dl,2)}^T, \cdots, \mathbf{x}_{(dl,K)}^T]^T. \tag{6.5}$$

The received N_k-dimensional vector $\mathbf{r}_k$ of the kth user, is given by

$$\begin{aligned} \mathbf{r}_k &= \mathbf{H}_k \mathbf{d}_{dl} + \mathbf{n}_{(dl,k)} \\ &= \mathbf{H}_k \mathbf{P} \mathbf{x}_{dl} + \mathbf{n}_{(dl,k)} \\ &= \mathbf{H}_k \sum_{i=1}^{K} \mathbf{P}_i \mathbf{x}_{(dl,i)} + \mathbf{n}_{(dl,k)} \\ &= \mathbf{H}_k \mathbf{P}_k \mathbf{x}_{(dl,k)} + \mathbf{H}_k \sum_{i=1, i\neq k}^{K} \mathbf{P}_i \mathbf{x}_{(dl,i)} + \mathbf{n}_{(dl,k)}, \end{aligned} \tag{6.6}$$

where $\mathbf{n}_{(dl,k)}$ is an N_k-dimensional AWGN vector having zero mean and the autocorrelation matrix of $E[\mathbf{n}_{(dl,k)}\mathbf{n}_{(dl,k)}^H] = \sigma_{(dl,k)}^2 \mathbf{I}_{N_k}$, while $\mathbf{H}_k$ is the $(N_k \times M)$-dimensional matrix of flat-fading CIR taps, which is given by

$$\mathbf{H}_k = \begin{bmatrix} h_{11}^{(k)} & h_{12}^{(k)} & \cdots & h_{1M}^{(k)} \\ h_{21}^{(k)} & h_{22}^{(k)} & \cdots & h_{2M}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_k1}^{(k)} & h_{N_k2}^{(k)} & \cdots & h_{N_kM}^{(k)} \end{bmatrix}, \tag{6.7}$$

where $h_{ij}^{(k)}$ represents the CIR coefficients between the jth BS antenna and the ith receiver antenna of the kth MS. As we can see from (6.6), Multiple User Interference (MUI) is imposed on the kth MS by all the other DL users.

The $\mathcal{U}$-dimensional received symbol vector $\mathbf{r}$ of the K DL users can be expressed as

$$\begin{aligned} \mathbf{r} &= \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_K \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} \mathbf{P}\mathbf{x}_{dl} + \mathbf{n}_{dl}, \\ &= \mathbf{H}\mathbf{P}\mathbf{x}_{dl} + \mathbf{n}_{dl}, \end{aligned} \tag{6.8}$$

where H is a $(\mathcal{U} \times M)$-dimensional matrix, which is given by

$$\mathrm{H} = [\mathrm{H}_1^T, \mathrm{H}_2^T, \cdots, \mathrm{H}_K^T]^T, \tag{6.9}$$

while $\mathbf{n}_{dl}$ is a $\mathcal{U}$-dimensional AWGN vector, which is expressed as

$$\mathbf{n}_{dl} = [\mathbf{n}_{(dl,1)}^T, \mathbf{n}_{(dl,2)}^T, \cdots, \mathbf{n}_{(dl,K)}^T]^T. \tag{6.10}$$

The AWGN has a zero mean and an autocorrelation matrix of

$$E[\mathbf{n}_{dl}\mathbf{n}_{dl}^H] = \begin{bmatrix} \sigma_{(dl,1)}^2\mathbf{I}_{N_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{(dl,2)}^2\mathbf{I}_{N_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \sigma_{(dl,K)}^2\mathbf{I}_{N_K} \end{bmatrix}. \tag{6.11}$$

6.3 MMSE Criterion for DL Preprocessing

Numerous criteria have been proposed for designing the DL preprocessing matrix $\mathbf{P}$ [151]. In this chapter, the preprocessing matrix $\mathbf{P}$ is chosen based on the MMSE criterion, so that the MSE between the received signal vector $\mathbf{r}$ and the transmitted symbols $\mathbf{x}_{dl}$ expressed as $E[||\mathbf{r}-\mathbf{x}_{dl}||^2]$ is minimized.

6.3.1 Exploiting the Knowledge of the AWGN Variance

When the variance of the AWGN to be experienced at the MS's receiver can be correctly estimated and fed back to the BS via the uplink, the MSE between the received signal vector $\mathbf{r}$ and the transmitted symbols $\mathbf{x}_{dl}$ is expressed as [153]

$$\begin{aligned} E[||\mathbf{r}-\mathbf{x}_{dl}||^2] &= E[(\mathbf{HP}\mathbf{x}_{dl}+\mathbf{n}_{dl}-\mathbf{x}_{dl})^H(\mathbf{HP}\mathbf{x}_{dl}+\mathbf{n}_{dl}-\mathbf{x}_{dl})] \\ &= E[(\mathbf{HP}\mathbf{x}_{dl}-\mathbf{x}_{dl})^H(\mathbf{HP}\mathbf{x}_{dl}-\mathbf{x}_{dl})] + E[\mathbf{n}_{dl}^H\mathbf{n}_{dl}] \\ &= E[(\mathbf{HP}\mathbf{x}_{dl}-\mathbf{x}_{dl})^H(\mathbf{HP}\mathbf{x}_{dl}-\mathbf{x}_{dl})] + \frac{E[\mathbf{n}_{dl}^H\mathbf{n}_{dl}]}{E[||\mathbf{x}_{dl}||^2]}E[||\mathbf{P}\mathbf{x}||^2] \\ &= E[(\mathbf{HP}\mathbf{x}_{dl}-\mathbf{x}_{dl})^H(\mathbf{HP}\mathbf{x}_{dl}-\mathbf{x}_{dl})] + \alpha E[||\mathbf{P}\mathbf{x}||^2] \\ &= Trace(\mathbf{R}_{\triangle}), \end{aligned} \tag{6.12}$$

where we have applied the constraint of $E[||\mathbf{P}\mathbf{x}_{dl}||_2^2] = E[||\mathbf{x}_{dl}||_2^2]$ in the third line of (6.12), and α is given by

$$\begin{aligned} \alpha &= \frac{E[\mathbf{n}^H\mathbf{n}]}{E[|\mathbf{x}_{dl}|^2]}, \\ &= \frac{\sigma_{dl}^2}{E_s}, \end{aligned} \tag{6.13}$$

where E_s denotes the power of each transmitted DL symbol and σ_{dl}^2 represents the AWGN variance at a MS's receiver, which is assumed to be the same for all MSs. Moreover $Trace(\cdot)$ in (6.12) denotes the trace of the argument. Furthermore, $\mathbf{R}_\triangle$ is given by

$$\begin{aligned} \mathbf{R}_\triangle &= E[(\mathbf{H}\mathbf{P}\mathbf{x}_{dl} - \mathbf{x}_{dl})(\mathbf{H}\mathbf{P}\mathbf{x}_{dl} - \mathbf{x}_{dl})^H] + \alpha E[(\mathbf{P}\mathbf{x})(\mathbf{P}\mathbf{x})^H] \\ &= \mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H - \mathbf{H}\mathbf{P} + \alpha\mathbf{P}\mathbf{P}^H - \mathbf{P}^H\mathbf{H}^H + \mathbf{I}_\mathcal{U}. \end{aligned} \tag{6.14}$$

The problem of minimizing $E[||\mathbf{r} - \mathbf{x}_{dl}||^2]$ is now turned into minimizing the trace of $\mathbf{R}_\triangle$, which can be achieved by differentiating $Trace(\mathbf{R}_\triangle)$ with respect to $\mathbf{P}^*$, yielding [162]

$$\frac{\partial Trace(\mathbf{R}_\triangle)}{\partial \mathbf{P}^*} = \mathbf{H}^H\mathbf{H}\mathbf{P} + \alpha\mathbf{P} - \mathbf{H}^H. \tag{6.15}$$

Setting (6.15) to zero, we arrive at

$$\begin{aligned} \mathbf{P} &= (\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_M)^{-1}\mathbf{H}^H \\ &= \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}_\mathcal{U})^{-1}, \end{aligned} \tag{6.16}$$

where we have used the Matrix Inversion Lemma [1] in the second line of (6.16).

6.3.2 Dispensing with the Knowledge of the AWGN Variance

When the variance of the background AWGN to be experienced at the MS's receiver is unknown to the transmitter, the MSE between the received signal vector $\mathbf{r}$ and the transmitted symbols $\mathbf{x}_{dl}$ can be written as

$$\begin{aligned} E[||\mathbf{r} - \mathbf{x}_{dl}||^2] &= E[(\mathbf{H}\mathbf{P}\mathbf{x}_{dl} + \mathbf{n}_{dl} - \mathbf{x}_{dl})^H(\mathbf{H}\mathbf{P}\mathbf{x}_{dl} + \mathbf{n}_{dl} - \mathbf{x}_{dl})] \\ &= Trace(E[(\mathbf{H}\mathbf{P}\mathbf{x}_{dl} + \mathbf{n}_{dl} - \mathbf{x}_{dl}))((\mathbf{H}\mathbf{P}\mathbf{x}_{dl} + \mathbf{n}_{dl} - \mathbf{x}_{dl})^H]) \\ &= Trace(\mathbf{R}_\triangle), \end{aligned} \tag{6.17}$$

where $\mathbf{R}_{\triangle}$ is a $(\mathcal{U} \times \mathcal{U})$-dimensional matrix, which can be expressed as

$$\begin{aligned} \mathbf{R}_{\triangle} &= E[(\mathbf{HPx}_{dl} + \mathbf{n}_{dl} - \mathbf{x}_{dl}))((\mathbf{HPx}_{dl} + \mathbf{n}_{dl} - \mathbf{x}_{dl}]^H) \\ &= \mathbf{HPP}^H\mathbf{H}^H - \mathbf{HP} + \sigma_{dl}^2\mathbf{I}_{\mathcal{U}} - \mathbf{P}^H\mathbf{H}^H + \mathbf{I}_{\mathcal{U}}. \end{aligned} \tag{6.18}$$

Comparision with (6.14) shows that there is no linkage between the AWGN variance σ_{dl}^2 and the preprocessing matrix $\mathbf{P}$ in (6.18). Similarly to (6.14), the specific solution which minimizes the trace of $\mathbf{R}_{\triangle}$ also minimizes the $E[||\mathbf{r} - \mathbf{x}_{dl}||^2]$. Hence, by differentiating $Trace(\mathbf{R}_{\triangle})$ with respect to $\mathbf{P}^*$, we have [162]

$$\frac{\partial Trace(\mathbf{R}_{\triangle})}{\partial \mathbf{P}^*} = \mathbf{H}^H\mathbf{HP} - \mathbf{H}^H. \tag{6.19}$$

Setting (6.19) to zero, as in (6.15), yields

$$\mathbf{H}^H\mathbf{HP} = \mathbf{H}^H. \tag{6.20}$$

Since $\mathbf{H}^H$ is a $(M \times \mathcal{U})$-dimensional matrix, we arrive at

$$(\mathbf{H}^H)^+\mathbf{H}^H = \mathbf{I}_{\mathcal{U}}, \tag{6.21}$$

where $(\mathbf{H}^H)^+$ is the $(\mathcal{U} \times M)$-dimensional pseudo-inverse of the matrix $\mathbf{H}^H$, which is given by [1]

$$(\mathbf{H}^H)^+ = [\mathbf{HH}^H]^{-1}\mathbf{H}. \tag{6.22}$$

Upon pre-multiplying both side of (6.20) with $(\mathbf{H}^H)^+$, we have

$$(\mathbf{H}^H)^+\mathbf{H}^H\mathbf{HP} = (\mathbf{H}^H)^+\mathbf{H}^H. \tag{6.23}$$

According to (6.21), (6.23) can be rewritten as

$$\mathbf{HP} = \mathbf{I}_{\mathcal{U}}. \tag{6.24}$$

Hence $\mathbf{P}$ is given by the pseudo inverse of the matrix $\mathbf{H}$,

$$\mathbf{P} = \mathbf{H}^+, \tag{6.25}$$

where $\mathbf{H}^+$ is the $(M \times \mathcal{U})$-dimensional inverse of the matrix $\mathbf{H}$, which is given by

$$\mathbf{H}^+ = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}. \tag{6.26}$$

Upon substituting (6.26) into (6.8) we have

$$\begin{aligned} \mathbf{r} &= \mathbf{H}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{x}_{dl} + \mathbf{n}_{dl} \\ &= \mathbf{x}_{dl} + \mathbf{n}_{dl}. \end{aligned} \tag{6.27}$$

Observe from (6.27) that interestingly, we arrive at the zero-forcing preprocessing formulation despite using the MMSE criterion, when the variance of the noise is unavailable at the transmitter. Therefore, in the rest of this chapter, we refer to preprocessing dispensing with the knowledge of the background AWGN variance as zero-forcing preprocessing. By contrast, preprocessing exploiting the knowledge of the receiver's background AWGN variance is termed as MMSE preprocessing.

6.3.3 Power Control

It is a natural constraint that the transmitted power of all users should remain unchanged after preprocessing. In this case, the employment of power control has to be considered in the context of transmitter preprocessing and in fact the normalized preprocessing matrix $\mathbf{P}_o$ has to be used instead of $\mathbf{P}$ for the sake of satisfying the constraint of

$$E[||\mathbf{P}_o\mathbf{x}_{dl}||_2^2] = E[||\mathbf{x}_{dl}||_2^2] = \mathcal{U}. \tag{6.28}$$

A natural ambition is to allocate the total BS transmitter power to all the users employing the same normalized coefficient β for all the users, yielding

$$\mathbf{P}_o = \beta\mathbf{P}, \tag{6.29}$$

where β is a real-valued variable. Upon substituting (6.29) into (6.28), we have

$$\begin{aligned} E[||\mathbf{P}_o\mathbf{x}_{dl}||_2^2] &= trace(E[(\beta\mathbf{P})\mathbf{x}_{dl}\mathbf{x}_{dl}^H(\beta\mathbf{P})^H) \\ &= \beta^2 trace(\mathbf{P}\mathbf{P}^H) \\ &= \mathcal{U}. \end{aligned} \tag{6.30}$$

Hence upon using (6.30) we have

$$\beta = \sqrt{\frac{\mathcal{U}}{trace(\mathbf{P}\mathbf{P}^H)}}. \tag{6.31}$$

6.3.4 Performance Analysis

Based on (6.29), (6.27) is expressed as

$$\mathbf{r} = \mathbf{x}_{dl}\beta + \mathbf{n}_{dl}, \tag{6.32}$$

where β is determined by (6.31). For each data stream $r_{(k,j)}$ of $\mathbf{r}$, the instantaneous signal-to-noise ratio (SNR) is given by

$$\gamma = \frac{\beta^2}{\sigma_{dl}^2}. \tag{6.33}$$

When baseband BPSK modulation is employed, the instantaneous BER of any data stream $r_{(k,j)}$ can be expressed as

$$P_b(\gamma) = Q\left(\sqrt{2\gamma}\right). \tag{6.34}$$

Furthermore, in the case of a single user having a single antenna, the SNR in case of MMSE preprocesing is given by

$$\gamma = \frac{\beta^2|\mathbf{H}\mathbf{P}|^2}{\sigma_{dl}^2}, \tag{6.35}$$

where $\mathbf{P}$ is determined by (6.16) and β is determined by (6.31). Likewise, when baseband BPSK modulation is employed, the instantaneous BER for this single user can be expressed as

$$P_b(\gamma) = Q\left(\sqrt{2\gamma}\right). \tag{6.36}$$

It may be readily shown that the achieveable BER is the same for both the MMSE and zero-forcing transmit preprocessing in the case of a single user having a single antenna.

6.4 MIMO-Aided SDMA Uplink Transmission

Let us now consider uplink (UL) transmissions, where $\mathbf{x}_{(ul,k)} = [x_{(ul,k1)}, x_{(ul,k2)}, \cdots, x_{(ul,kN_k)}]^T$ is the N_k-dimensional UL transmitted symbol vector of the kth user, as seen in Figure 6.2 Since TDD transmissions are considered in this chapter, we assume that the UL and DL CIRs are indentical. Hence the M-dimensional received UL signal $\mathbf{y}$ at BS is expressed as

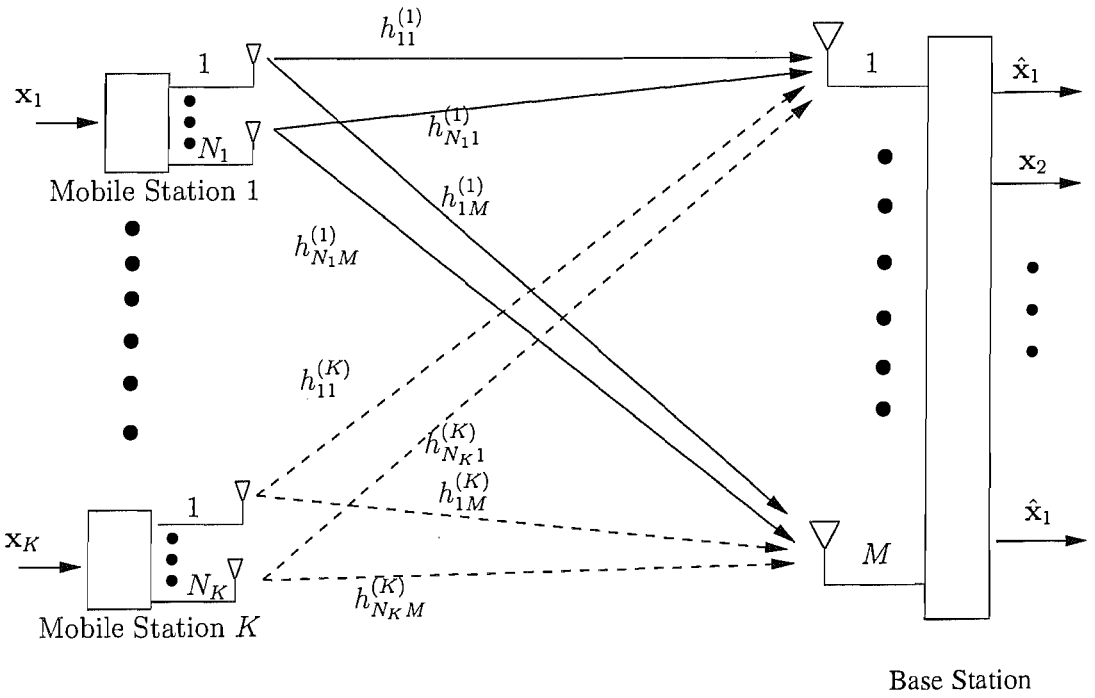


Figure 6.2: Schematic of the SDMA uplink transmitter

$$
\begin{aligned}
\mathbf{y} &= \sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{x}_{(ul,k)} + \mathbf{n}_{ul} \\
&= \begin{bmatrix} \mathbf{H}_1^T & \mathbf{H}_2^T & \cdots & \mathbf{H}_K^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_{(ul,1)} \\ \mathbf{x}_{(ul,2)} \\ \vdots \\ \mathbf{x}_{(ul,K)} \end{bmatrix} + \mathbf{n}_{ul} \\
&= \mathbf{H}^T \mathbf{x}_{ul} + \mathbf{n}_{ul},
\end{aligned} \tag{6.37}
$$

where $\mathbf{x}_{ul}$ is the $\mathcal{U}$-dimensional composite transmitted symbol vector of the K users, which is given by $\mathbf{x}_{ul} = [\mathbf{x}^T_{(ul,1)}, \mathbf{x}^T_{(ul,2)}, \cdots, \mathbf{x}^T_{(ul,K)}]^T$, while $\mathbf{n}_{ul}$ is an M-dimensional AWGN vector having zero mean and an autocorrelation matrix of $E[\mathbf{n}_{ul}\mathbf{n}^H_{ul}] = \sigma^2_{ul}\mathbf{I}_M$.

In the context of UL transmissions the $\mathcal{U}$-dimensional decision variable vector $\mathbf{z}$ at the BS can be expressed as

$$\mathbf{z} = \mathbf{W}^H\mathbf{y}, \tag{6.38}$$

where $\mathbf{W}$ is a $(M \times \mathcal{U})$-dimensional receiver weight matrix. When BPSK modulation is employed, hard-decisions are made as follows:

$$\hat{\mathbf{x}}_{ul} = sgn(Re\{\mathbf{z}\}), \tag{6.39}$$

where $sgn(\cdot)$ is the sign function. If the classic MMSE detector is invoked, the optimum weight matrix $\mathbf{W}^H_o$ is given by

$$\mathbf{W}_o = \mathbf{R}^{-1}_{\mathbf{y}}\mathbf{R}_{\mathbf{yx}}, \tag{6.40}$$

where $\mathbf{R}_{\mathbf{y}}$ is a $(M \times M)$-dimensional autocorrelation matrix of $\mathbf{y}$, which is given by

$$\begin{aligned} \mathbf{R}_{\mathbf{y}} &= E[\mathbf{y}\mathbf{y}^H] \\ &= E[(\mathbf{H}^T\mathbf{x}_{ul} + \mathbf{n}_{ul})(\mathbf{H}^T\mathbf{x}_{ul} + \mathbf{n}_{ul})^H] \\ &= \mathbf{H}^T\mathbf{H}^* + \sigma^2_{ul}\mathbf{I}_M, \end{aligned} \tag{6.41}$$

while $\mathbf{R}_{\mathbf{yx}}$ is the $(M \times \mathcal{U})$-dimensional cross-correlation matrix of $\mathbf{y}$ and $\mathbf{x}$, which is given by

$$\begin{aligned} \mathbf{R}_{\mathbf{yx}} &= E[\mathbf{y}\mathbf{x}^H] \\ &= E[(\mathbf{H}^T\mathbf{x}_{ul} + \mathbf{n}_{ul})\mathbf{x}^H_{ul}] \\ &= \mathbf{H}^T. \end{aligned} \tag{6.42}$$

Consequently, we have

$$\mathbf{W}_o = (\mathbf{H}^T\mathbf{H}^* + \sigma^2_{ul}\mathbf{I}_M)^{-1}\mathbf{H}^T. \tag{6.43}$$

As seen from (6.43), the knowledge of the UL CIR matrix is needed at the BS for the MMSE SDMA detector. Alternatively, the least mean square (LMS) or recursive least square (RLS) algorithms [1] can be used for detection, when the CIR matrix is not available.

6.5 Channel Prediction Based SDMA DL Transmitter Preprocessing

In order to perform BS transmit preprocessing for the SDMA downlink as described in Section 6.2, the CIR matrix hosting DL CIRs must be availble at the BS station. As a benefit of the TDD mode, the downlink CIRs can be estimated or predicted based on the CIRs of the uplink transmission [1]. However, the inherent delay of the estimated CIRs may seriously degrade the attainable performance [156]. Hence, the employment of prediction is preferred [153,154] for the SDMA DL. In this chapter, Kalman filtering assisted MIMO channel prediction is invoked [161]. To this end, (6.37) is rewritten for the time instant n as

$$\mathbf{y}(n) = [\mathbf{h}_1(n), \cdots, \mathbf{h}_{\mathcal{U}}(n)]\mathbf{x}_{ul}(n) + \mathbf{n}_{ul}(n), \tag{6.44}$$

where the M-dimensional vector $\mathbf{h}_j(n), 1 \leq j \leq \mathcal{U}$ is the jth column of $\mathbf{H}^T$, which represents the CIR vector between the jth MS transmitter antenna and all the BS receiver antennas. Then, we construct a $(M\mathcal{U})$-length channel vector

$$\mathbf{h}(n) = [\mathbf{h}_1^T(n), \mathbf{h}_2^T(n), \cdots, \mathbf{h}_{\mathcal{U}}^T(n)]^T, \tag{6.45}$$

According to (2.32) the ith, $1 \leq i \leq M$ element $h_{ji}(n)$ in $\mathbf{h}_j(n)$, which represents the channel CIR tap between the jth MS transmitter antenna and the ith receiver BS receiver antenna, can be expressed by an AutoRegressive (AR) model as [79]

$$h_{ji}(n) = \sum_{q=1}^{p} a_{(ji,q)} h_{ji}(n-q) + w_{ji}(n), \tag{6.46}$$

where p is the order of the channel's AR model, $\{a_{(ji,q)}\}$ represents the AR model coefficients and $w_{ji}(n)$ is an AWGN process having a zero mean and an autocorrelation coefficient of $E[w_{ji}(n)w_{ji}^*(n)] = \sigma^2_{w_{ji}(n)}$, where $\sigma^2_{w_{ji}(n)}$ can be obtained from (2.38). According to (6.46), the CIR vector between the jth MS transmitter antenna and all the BS receiver antennas $\mathbf{h}_j(n)$ can be

[1] As it was shown in Figure 15.3, p559 of [5], estimation, is carried out on the basis of previous CIR tap values for the current instant, while prediction determines their future values.

expressed as

$$\mathbf{h}_j(n) = \sum_{q=1}^{p} \mathbf{A}_{(j,q)} \mathbf{h}_j(n-q) + \mathbf{w}_j(n), \tag{6.47}$$

where $\{\mathbf{A}_{j,q}\}$ represents the $(M \times M)$-dimensional diagonal AR model coefficient matrices in the jth column of the CIR tap matrix $\mathbf{H}^T$, which are given by

$$\mathbf{A}_{j,q} = \begin{bmatrix} a_{j1,q} & 0 & \cdots & 0 \\ 0 & a_{j2,q} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{jM,q} \end{bmatrix}, \tag{6.48}$$

while $\mathbf{w}_j(n)$ is an M-dimensional AWGN vector, which is given by

$$\mathbf{w}_j(n) = [w_{j1}(n), \cdots, w_{jM}(n)]^T. \tag{6.49}$$

Upon substituting (6.47) into (6.45), the whole channel CIR vector $\mathbf{h}(n)$ can be expressed with the aid of AR model as

$$\mathbf{h}(n) = \sum_{q=1}^{p} \mathbf{A}_q \mathbf{h}(n-q) + \mathbf{w}(n), \tag{6.50}$$

where $\mathbf{A}_q$ is an $(M\mathcal{U} \times M\mathcal{U})$-dimensional diagonal matrix, which is expressed as

$$\mathbf{A}_q = \begin{bmatrix} \mathbf{A}_{1,q} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{2,q} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{\mathcal{U},q} \end{bmatrix}, \tag{6.51}$$

while $\mathbf{w}(n)$ is a $(M \times \mathcal{U})$-dimensional AWGN vector, which is given by

$$\mathbf{w}(n) = [\mathbf{w}_1(n)^T, \cdots, \mathbf{w}_{\mathcal{U}}(n)^T]^T. \tag{6.52}$$

If the p CIR vectors $\mathbf{h}(n-q)$ $(0 \leq q \leq (p-1))$ corresponding to p consecutive time instants are combined into a new vector $\tilde{\mathbf{h}}(n) = [\mathbf{h}(n)^T, \mathbf{h}(n-1)^T, \cdots, \mathbf{h}(n-p+1)^T]^T$, based on (6.50) we

arrive at

$$\tilde{\mathbf{h}}(n) = \tilde{\mathbf{F}}(n)\tilde{\mathbf{h}}(n-1) + \tilde{\mathbf{w}}(n), \tag{6.53}$$

where $\tilde{\mathbf{F}}(n)$ is a $[(M \times \mathcal{U} \times p) \times (M \times \mathcal{U} \times p)]$-dimensional transition matrix [1], describing the state transition from time instant $(n-1)$ to n, which is given by

$$\tilde{\mathbf{F}}_n = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_p \\ \mathbf{I}_{(M\times\mathcal{U})} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I}_{(M\times\mathcal{U})} & \mathbf{0} \end{bmatrix}, \tag{6.54}$$

while $\tilde{\mathbf{w}}(n)$ is a $((M \times \mathcal{U} \times p)$-dimensional noise vector, which is formulated as

$$\tilde{\mathbf{w}}(n) = [\mathbf{w}(n)^T, \mathbf{0}^T, \cdots, \mathbf{0}^T]^T. \tag{6.55}$$

Now the received UL signal vector of (6.44) can be rewritten as [161]

$$\mathbf{y}(n) = \mathbf{X}(n)\tilde{\mathbf{h}}(n) + \mathbf{n}_{ul}(n), \tag{6.56}$$

where $\mathbf{X}$ is a $(M \times (M \times \mathcal{U} \times p))$-dimensional matrix, which is given by

$$\mathbf{X} = \tilde{\mathbf{x}} \bigotimes \mathbf{I}_M, \tag{6.57}$$

with $\bigotimes$ representing the Kronecker product. Furthermore, in (6.57) $\tilde{\mathbf{x}}$ is a $(\mathcal{U} \times p)$-dimensional row vector, given by

$$\tilde{\mathbf{x}}(n) = [\mathbf{x}_{ul}^T, \mathbf{0}]. \tag{6.58}$$

Given the process equation and measurement equation of (6.53) and (6.56) [1], respectively, vector Kalman filtering assisted MIMO channel prediction of $\tilde{\mathbf{h}}(n+1|n)$ based on all the observations up to the time instant n, can be performed using the vector-based Kalman filtering assisted MIMO CIR-tap prediction procedure described in Table 6.1 [1,2].

$$
\begin{array}{l}
\tilde{\mathbf{h}}(n+1|n) = \tilde{\mathbf{F}}(n)\tilde{\mathbf{h}}(n|n-1) + \tilde{\mathbf{G}}(n)\tilde{\boldsymbol{\alpha}}(n); \\
\tilde{\boldsymbol{\alpha}}(n) = \mathbf{y}(n) - \mathbf{X}(n)\tilde{\mathbf{h}}(n|n-1); \\
\tilde{\mathbf{G}}(n) = \tilde{\mathbf{F}}(n)\tilde{\mathbf{K}}(n)\mathbf{X}^H(n)\tilde{\mathbf{R}}^{-1}(n); \\
\tilde{\mathbf{R}}(n) = \mathbf{X}(n)\tilde{\mathbf{K}}(n)\mathbf{X}^H(n) + \tilde{\mathbf{Q}}_{\mathbf{v}}(n); \\
\tilde{\mathbf{Q}}_{\mathbf{v}}(n) = E[\mathbf{n}_{ul}(n)\mathbf{n}_{ul}^H(n)] = \sigma_{ul}^2\mathbf{I}_M; \\
\tilde{\mathbf{K}}(n+1) = \tilde{\mathbf{F}}(n)\tilde{\mathbf{M}}(n)\tilde{\mathbf{F}}^H(n) + \tilde{\mathbf{Q}}_w(n+1); \\
\tilde{\mathbf{Q}}_w(n) = E[\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}(n)^H]; \\
\tilde{\mathbf{M}}(n) = \tilde{\mathbf{K}}(n) - \tilde{\mathbf{K}}(n)\mathbf{X}^H(n)(\mathbf{X}(n)\tilde{\mathbf{K}}(n)\mathbf{X}^H(n) + \tilde{\mathbf{Q}}_v(n))^{-1}\mathbf{X}(n)\tilde{\mathbf{K}}(n).
\end{array}
$$

Table 6.1: Vector Kalman Filtering Assisted MIMO Channel Prediction

6.6 TDD Standard

All of our above discussions were based on using a TDD transmission mode. Hence below we provide a brief overview of the existing third-generation TDD standard, namely the UMTS Terrestrial Radio Access (UTRA) TDD [123, 124]. The UTRA TDD time slot allocation is exemplified in Figure 6.3 [123,124], where ↑ denotes an uplink time slot and ↓ denotes a downlink time slot. Furthermore, Time Division-Synchronous Code Division Multiple Access (TD-SCDMA) also contributes a promising TDD transmission mode, [125, 163], which may be adopted as the third generation scheme in China. An example of the corresponding time slot allocation schemes is shown in Figure 6.4 [163].

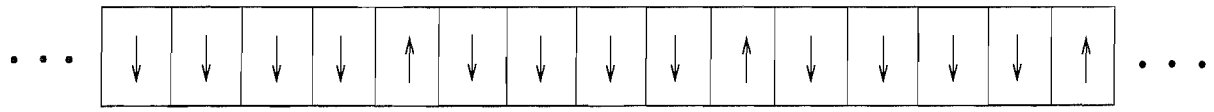


Figure 6.3: Example of UTRA TDD mode transmission.

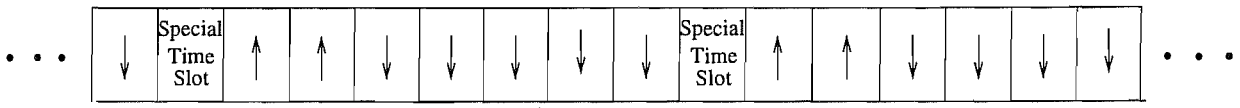


Figure 6.4: Example of TD-SCDMA mode transmission.

6.7 Performance Results

In this section, we provide simulation results for characterizing the performance of channel prediction aided downlink preprocessing. More specifically, a generalized TDD mode is assumed, where the

	Scheme 1	Scheme 2	Scheme 3	Scheme 4
Number of users	1	2	3	4
No. of receiver antennas at each user	1	1	1	1
Modulation scheme	BPSK	BPSK	BPSK	BPSK

Table 6.2: Parameters used for SDMA downlink (DL) preprocessing

time-span between two transmitted uplink symbols used for CIR tap prediction is assumed to be L and the range of prediction $\mathcal{L}$, where the predicted CIR taps are used is shown in Figure 6.5. Furthermore, T_s is the symbol duration, while the normalized maximum Doppler frequency is given by $f_{dm}T_sL$. Additionally, BPSK modulation is employed for both uplink and downlink transmissions. Finally, error-freely detected uplink symbols are assumed instead of invoking any specific uplink detector, implying that our results represent an idealistic upper bound.

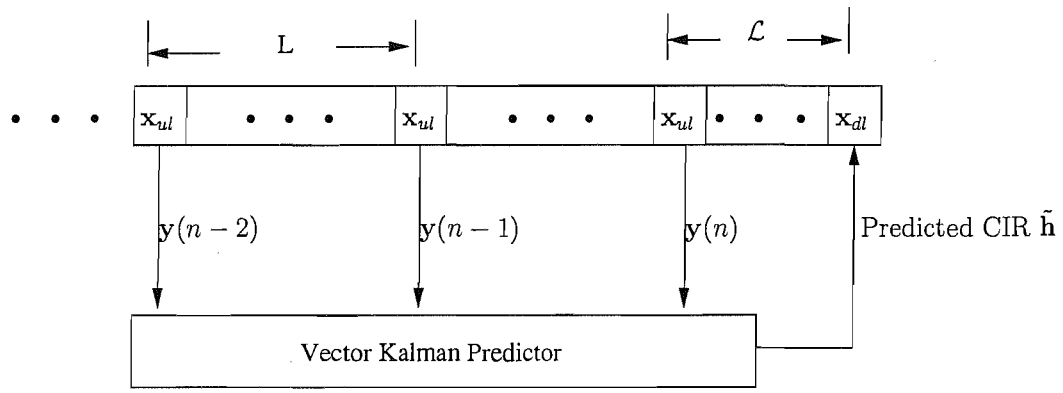


Figure 6.5: Schematic of the MIMO channel prediction scheme porposed for TDD systems.

6.7.1 Effect of the Number of Transmitter Antennas at the BS

In Figures 6.6 to 6.12 the achievable BER performance versus average SNR per symbol performance is plotted for Schemes 1-4 of Table 6.2, when the number of transmitter antennas at the BS ranges from $M = 1$ to $M = 10$, for both zero-forcing and MMSE BS transmit preprocessing. Observe from Figures 6.6 to 6.12 that as expected, the BER performance is improved upon increasing the number of BS transmit antennas. However, the extra transmit diversity gain attained every time, when increasing M by one gradually erodes, since transmit diversity gain does not increase linearly with M.

Additionally, the SNR required for maintaining BER=10^{-3} for Schemes 1-4 for both zero forcing and MMSE preprocessing was extracted from Figures 6.6 to 6.12 and summarized in Table 6.3

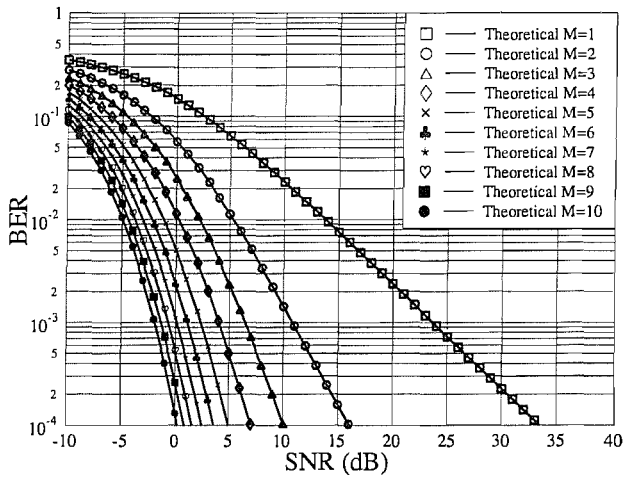


Figure 6.6: BER versus average SNR per symbol performance for **Scheme 1** of Table 6.2, when the number of transmitter antennas M at the BS ranges from $M = 1$ to $M = 10$ for **zero-forcing** BS transmit preprocessing according to (6.34).

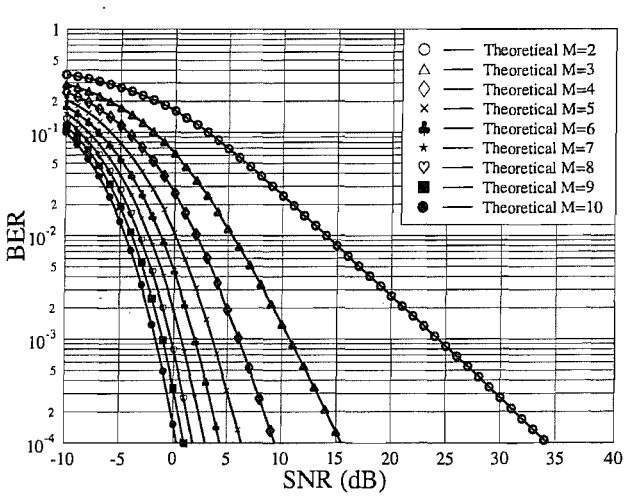


Figure 6.7: BER versus average SNR per symbol performance for **Scheme 2** of Table 6.2, when the number of transmitter antennas M at the BS ranges from $M = 2$ to $M = 10$ for **zero-forcing** BS transmit preprocessing according to (6.34).

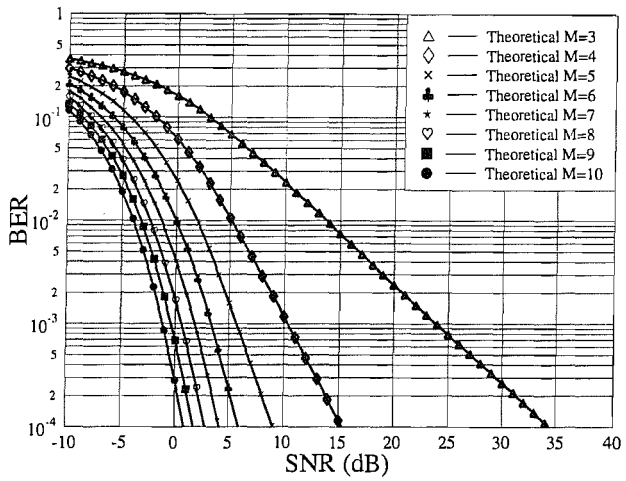


Figure 6.8: BER versus average SNR per symbol performance for **Scheme 3** of Table 6.2, when the number of transmitter antennas M at the BS ranges from $M = 3$ to $M = 10$ for **zero-forcing** BS transmit preprocessing according to (6.34).

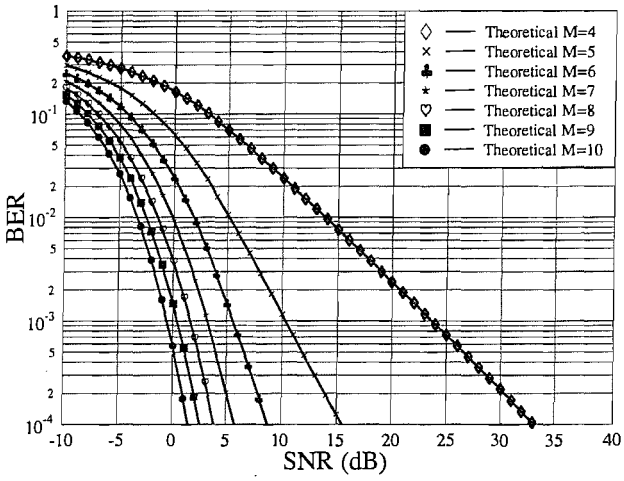


Figure 6.9: BER versus average SNR per symbol performance for **Scheme 4** of Table 6.2, when the number of transmitter antennas M at the BS ranges from $M = 4$ to $M = 10$ for **zero-forcing** BS transmit preprocessing according to (6.34).

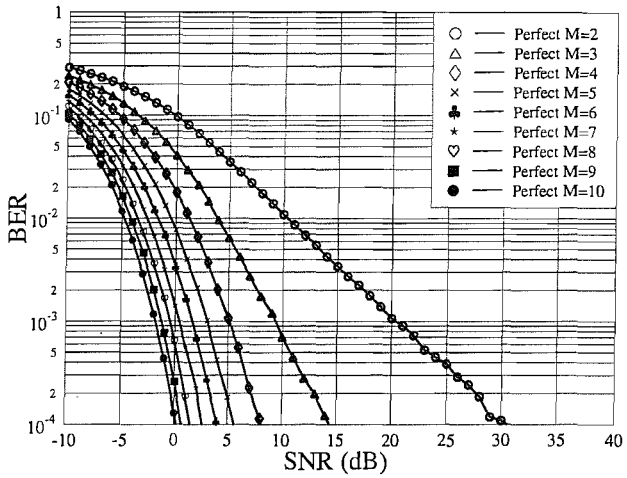


Figure 6.10: BER versus average SNR per symbol performance for the first user for **Scheme 2** of Table 6.2, when the number of transmitter antennas M at the BS ranges from $M = 2$ to $M = 10$ for **MMSE** BS transmit preprocessing using perfect CIR.

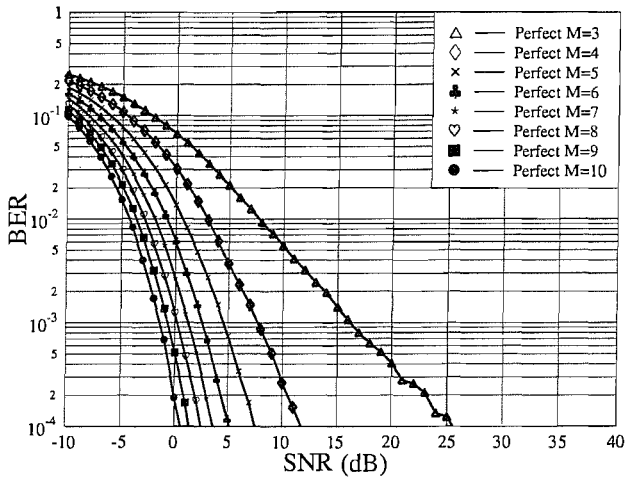


Figure 6.11: BER versus average SNR per symbol performance for the first user for **Scheme 3** of Table 6.2, when the number of transmitter antennas M at the BS ranges from $M = 3$ to $M = 10$ for **MMSE** BS transmit preprocessing using perfect CIR.

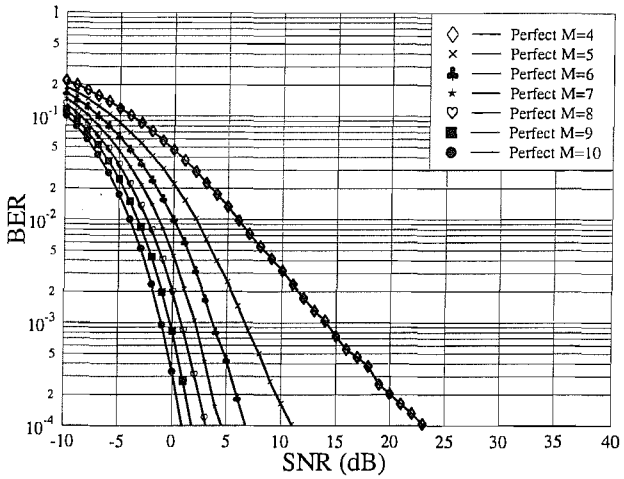
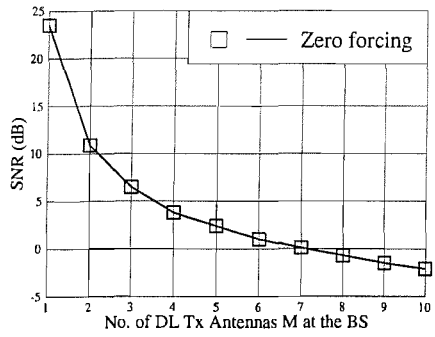


Figure 6.12: BER versus average SNR per symbol performance for the first antenna of first user for **Scheme 4** of Table 6.2, when the number of transmitter antennas M at the BS ranges from $M = 4$ to $M = 10$ for **MMSE** BS transmit preprocessing using perfect CIR.

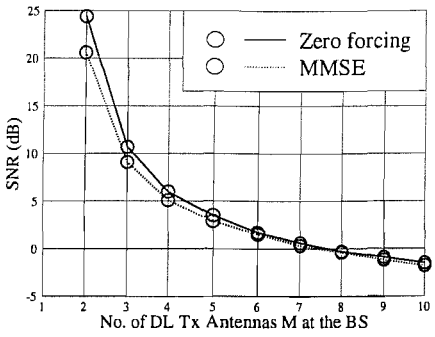
	Scheme 1 [dB]	Scheme 2 [dB]		Scheme 3 [dB]		Scheme 4 [dB]	
Throughput	1 bit	2 bits		3 bits		4 bits	
	Zero forcing	Zero forcing	MMSE	Zero forcing	MMSE	Zero forcing	MMSE
From	Fig.6.6	Fig.6.7	Fig.6.10	Fig.6.8	Fig.6.11	Fig.6.9	Fig.6.12
M=1	23.5						
M=2	10.9	24.4	20.6				
M=3	6.5	10.7	9.1	23.8	16.2		
M=4	3.8	6.0	5.1	10.3	7.6	23.8	14.1
M=5	2.4	3.5	2.9	5.7	4.6	10.3	6.6
M=6	1.0	1.7	1.5	3.2	2.5	5.6	3.7
M=7	0.1	0.6	0.3	1.6	1.2	3.2	2.1
M=8	-0.7	-0.3	-0.4	0.6	0.1	1.6	0.7
M=9	-1.5	-0.9	-1.2	-0.3	-0.6	0.4	-0.2
M=10	-2.1	-1.5	-1.8	-1.0	-1.3	-0.6	-1.0

Table 6.3: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.2

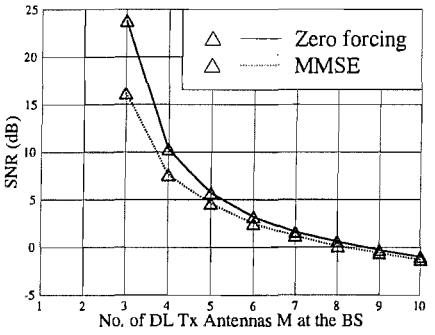
for Schemes 1-4 of Table 6.2, when maintaining BER=10^{-3}, which is also presented in Fig.6.13. As we can see from Fig.6.13, MMSE preprocessing required a lower SNR for attaining the same performance than that necessitated by zero forcing preprocessing. This is due to the fact that MMSE preprocessing takes into accout the effects of both the MUI and of the noise jointly, while zero forcing only considers those of the MUI.



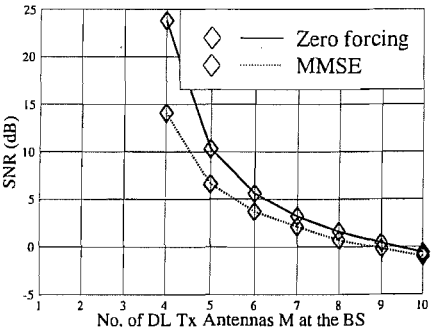
(a) Scheme 1



(b) Scheme 2



(c) Scheme 3



(d) Scheme 4

Figure 6.13: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.2

Number of users K	2
Number of antennas per user N_k	2
Order of Kalman filtering assisted predictor p	2
Range of prediction $\mathcal{L}$	L/2

Table 6.4: Parameters of the channel prediction based SDMA downlink preprocessing in TDD mode

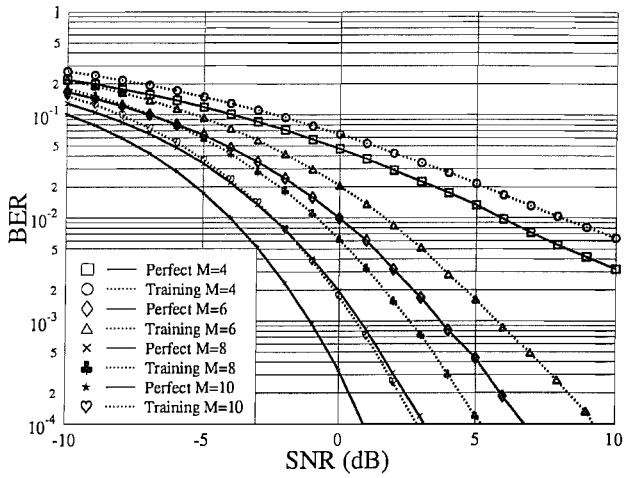


Figure 6.14: BER versus average SNR per symbol performance for downlink transmission for the first receive antenna of the first user, when the number of transmitter antennas at the BS is assumed to be 4, 6, 8, and 10, respectively for MMSE BS transmit preprocessing. Furthermore, the maximum normalized Doppler frequency is $f_{dm}T_sL = 0.001$ and the remaining parameters are the same as in Table 6.4.

In Figure 6.14, the attainable BER performance versus average SNR is plotted for the first receive antenna of the first user, when the number of antennas used for MMSE BS transmit preprocessing is assumed to be $M = 4$, 6, 8, and 10, respectively. Furthermore, the maximum normalized Doppler frequency is $f_{dm}T_sL = 0.001$ and the remaining parameters are the same as in Table 6.4. As we can see from Figure 6.14, once the number of transmitter antennas at the BS becomes more than the sum of the number of each MS's receiver antennas, the achievable BER performance is significantlly improved for both the perfect and predicted CIR-tap scenarios as a benefit of tranmitter diversity. Furthermore, observe in Figure 6.14 that the performance difference between $M = 4$ and 6, $M = 6$ and 8, as well as $M = 8$ and 10 recorded for both the perfect and predicted CIR-tap scenarios, becomes narrower. This is because the achieveable extra additional transmit diversity gain becomes lower upon every further increase of the number of antennas M at the BS. Moreover, the performance discrepancy seen in Figure 6.14 between the perfect and predicted CIR-tap scenarios becomes lower upon increasing the number of antennas at the BS since the performance loss due to prediction can be effectively compensated by increasing the transmit diversity gain and hence approaching a near AWGN performance. Additionally, the performance discrepancy between the scenarios using predicted CIR

taps in conjunction with $M = 10$ and perfect CIR taps with $M = 8$ is narrower than that between the predicted CIR scenario using $M = 8$ and the perfect CIR scenario employing $M = 6$. Similarly, observe in Figure 6.14 that the latter scenario has a lower performace discrepancy than that between the predicted CIR case employing $M = 6$ and the perfect CIR scenario using $M = 4$. Especially, the BER performance of the predicted CIR scenario using $M = 10$ was shown to be better in Figure 6.14 than that using perfect CIR-tap knowledge in conjunction with $M = 8$ in the lower range of SNRs. The reason for this is because the extra transmit gain attained compensates for the performance loss imposed by the channel's prediction error. The same trend can be observed in Figure 6.15 in the context of zero-forcing BS transmit preprocessing, where the theoretical BER performance of (6.34) is also plotted. Reassuringly, the theoretical results closely agree with those corresponding to the perfect CIR scenario.

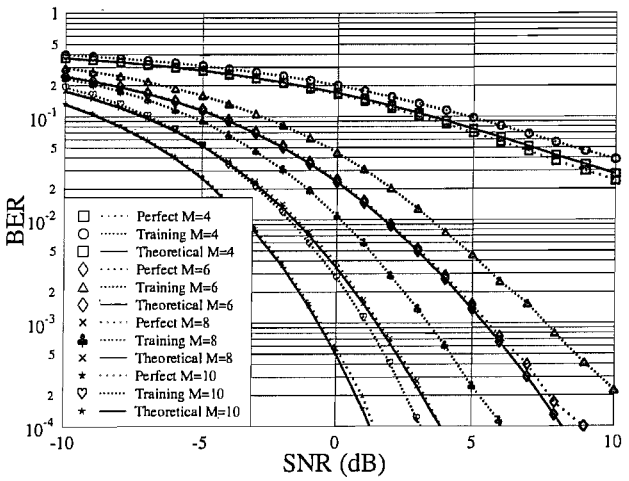


Figure 6.15: BER versus average SNR per symbol performance for downlink transmission for the first receive antenna of the first user when the number of transmitter antennas at BS is assumed to be 4, 6, 8, and 10, respectively for zero-forcing BS transmit preprocessing. Furthermore, the maximum normalized Doppler frequency is $f_{dm}T_sL = 0.001$ and the remaining parameters are the same as in Table 6.4.

6.7.2 Effect of the Number of MSs

In Figures 6.16 to 6.23 the achievable BER performance versus average SNR per symbol performance is plotted for Schemes 1-4 of Table 6.5, when the number of MSs K ranges from $K = 1$ to $K = 10$,

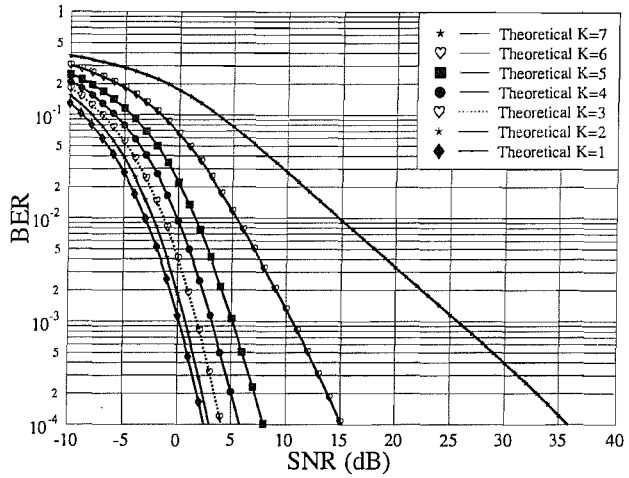


Figure 6.16: BER versus average SNR per symbol performance for **Scheme 1** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **zero forcing** BS transmit preprocessing according to (6.34).

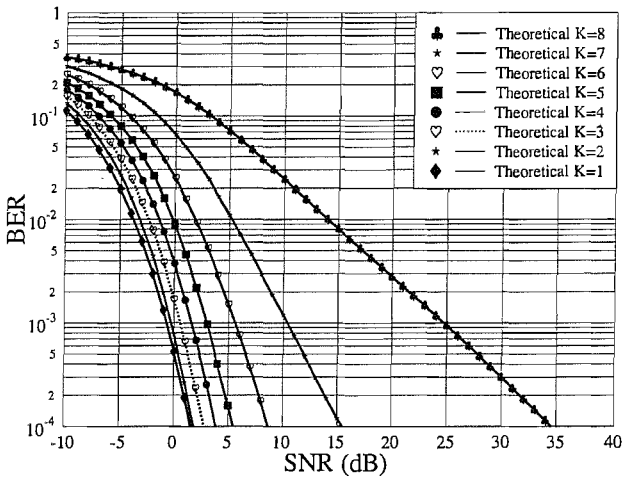


Figure 6.17: BER versus average SNR per symbol performance for **Scheme 2** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **zero forcing** BS transmit preprocessing according to (6.34).

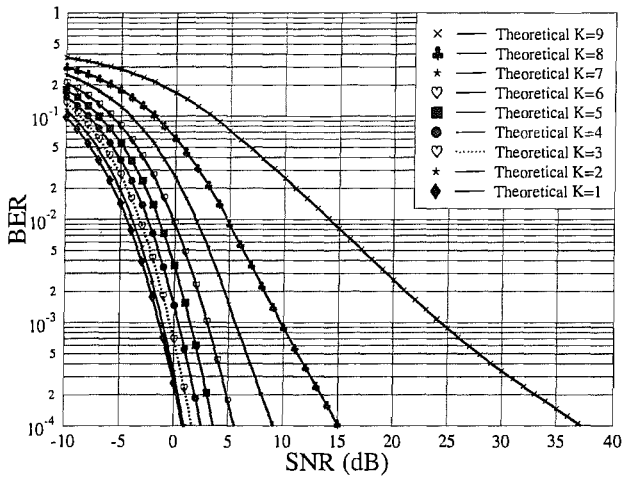


Figure 6.18: BER versus average SNR per symbol performance for **Scheme 3** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **zero forcing** BS transmit preprocessing according to (6.34).

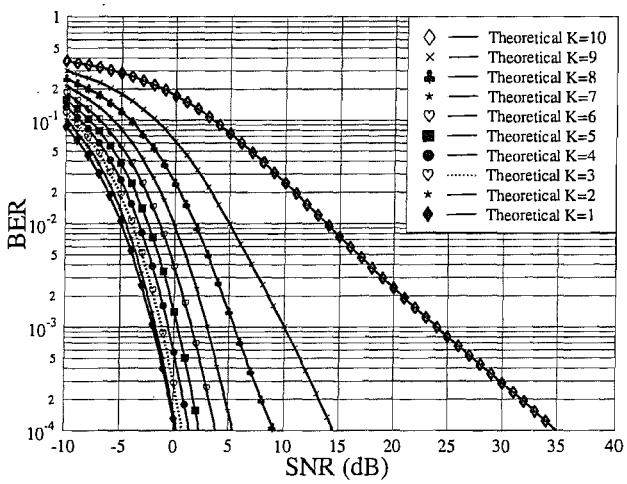


Figure 6.19: BER versus average SNR per symbol performance for **Scheme 4** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **zero forcing** BS transmit preprocessing according to (6.34).

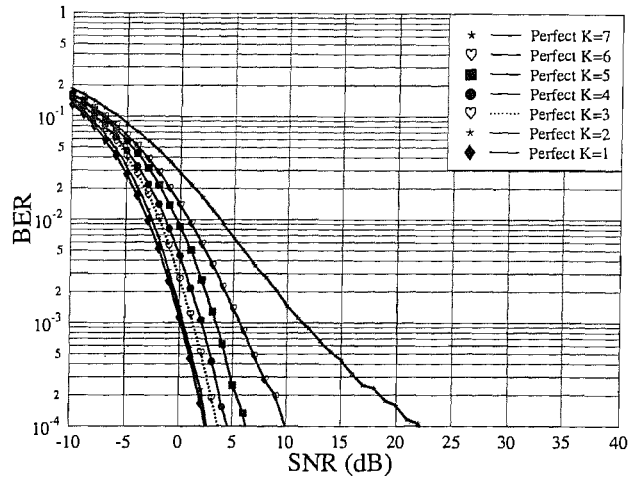


Figure 6.20: BER versus average SNR per symbol performance for **Scheme 1** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **MMSE** BS transmit preprocessing using perfect CIR.

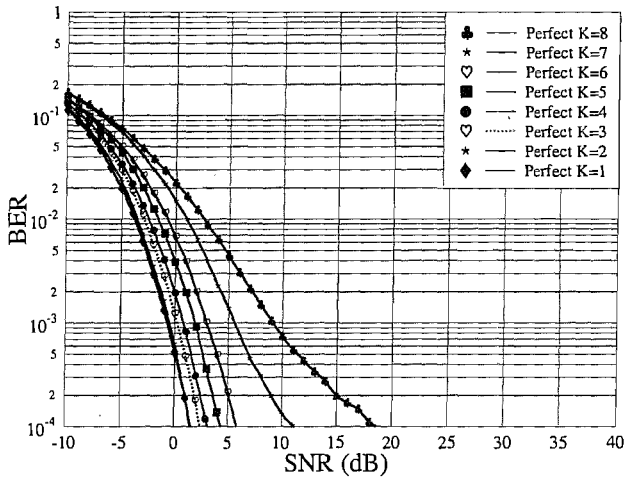


Figure 6.21: BER versus average SNR per symbol performance for **Scheme 2** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **MMSE** BS transmit preprocessing using perfect CIR.

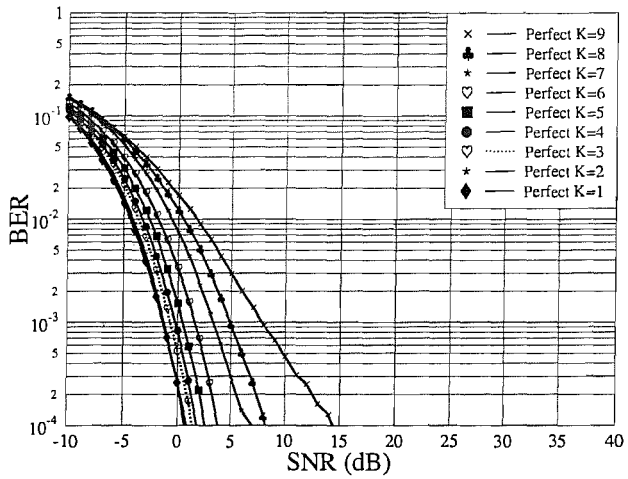


Figure 6.22: BER versus average SNR per symbol performance for **Scheme 3** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **MMSE** BS transmit preprocessing using perfect CIR.

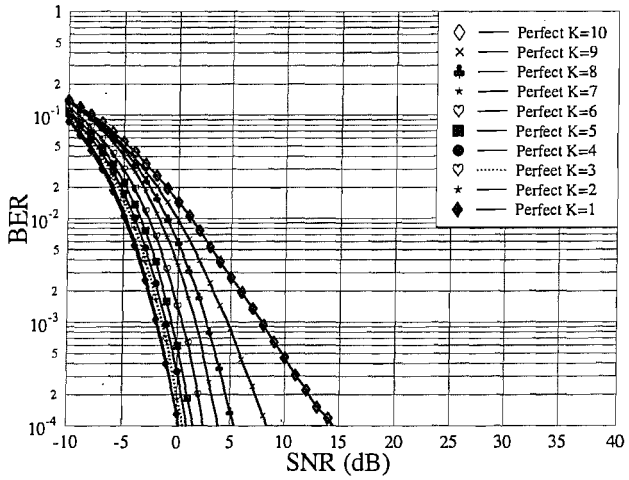


Figure 6.23: BER versus average SNR per symbol performance for **Scheme 4** of Table 6.5, when the number of MS K ranges from $K = 1$ to $K = 10$ for **MMSE** BS transmit preprocessing using perfect CIR.

	Scheme 1	Scheme 2	Scheme 3	Scheme 4
No. of antennas at each MS	1	1	1	1
No. of antennas at the BS	7	8	9	10
Modulation scheme	BPSK	BPSK	BPSK	BPSK

Table 6.5: Parameters II used for SDMA downlink (DL) preprocessing

		Scheme 1 [dB]		Scheme 2 [dB]		Scheme 3 [dB]		Scheme 4 [dB]	
		Zero forcing	MMSE	Zero forcing	MMSE	Zero forcing	MMSE	Zero forcing	MMSE
	Throughput	Fig.6.16	Fig.6.20	Fig.6.17	Fig.6.21	Fig.6.18	Fig.6.22	Fig.6.19	Fig.6.23
K=1	1 bits	0.1	0.1	-0.7	-0.7	-1.3	-1.3	-2.1	-2.1
K=2	2 bits	0.6	0.4	-0.3	-0.4	-1.0	-1.2	-1.5	-1.8
K=3	3 bits	1.8	1.2	0.6	0.3	-0.3	-0.7	-1.0	-1.3
K=4	4 bits	3.1	2.1	1.5	0.7	0.3	-0.3	-0.6	-1.0
K=5	5 bits	5.0	3.2	3.0	1.9	1.5	0.4	0.6	-0.4
K=6	6 bits	10.6	5.6	5.6	2.9	2.9	1.5	1.5	0.4
K=7	7 bits	25.7	11.5	10.3	5.6	5.6	3.2	2.9	1.8
K=8	8 bits			24.7	9.0	9.7	4.7	5.4	2.6
K=9	9 bits					24.4	7.8	10.0	4.7
K=10	10 bits							24.1	7.6

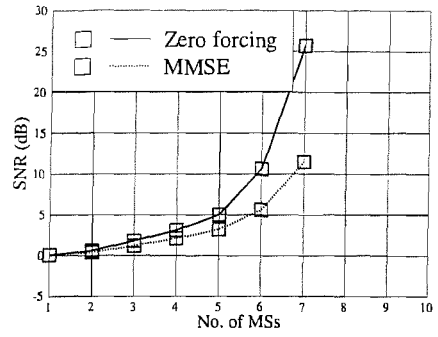
Table 6.6: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.5

for both zero-forcing and MMSE BS transmit preprocessing. Observe from Figures 6.16 to 6.23 that as expected, the BER performance is improved upon decreasing the number of MSs. However, the extra transmit diversity gain attained every time, when decreasing K by one gradually erodes, since the achievable transmit diversity gain does not increase linearly upon decreasing K.

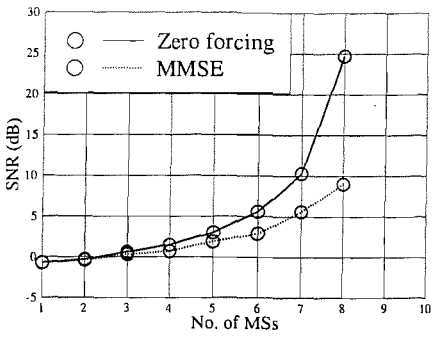
Additionally, the SNR required for maintaining BER=10^{-3} for Schemes 1-4 of Table 6.5 for both zero forcing and MMSE preprocessing was extracted from Figures 6.16 to 6.23 and summarized in Table 6.6 for Schemes 1-4 of Table 6.5, when maintaining BER=10^{-3}, which is also depicted in Fig.6.24. As we can see from Fig.6.24, MMSE preprocessing required a lower SNR than zero forcing preprocessing for attaining the same BER performance. The reason for this observation is the same as that discussed before in the context of Table 6.3.

6.7.3 Performance With or Without the Knowledge of AWGN Variance

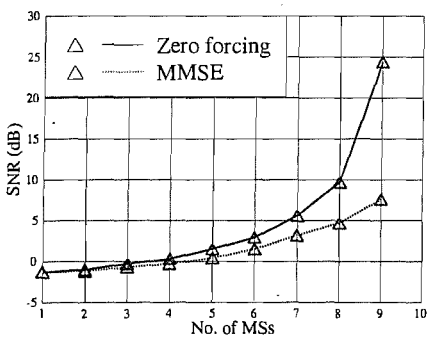
In Figure 6.25 the BER versus average SNR per symbol performance was recorded for downlink transmission in the context of the first receive antenna of the first user when the number of antennas at the BS is assumed to be M =4 and 6, respectively, for both MMSE and zero-forcing BS transmit



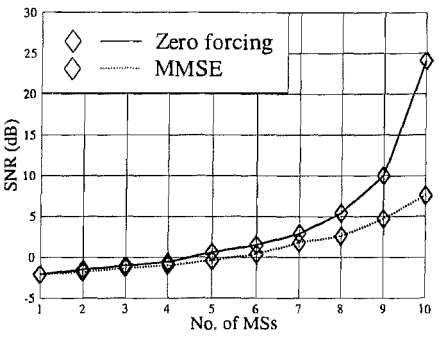
(a) Scheme 1



(b) Scheme 2



(c) Scheme 3



(d) Scheme 4

Figure 6.24: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.5

preprocessing. Furthermore, the maximum normalized Doppler frequency is $f_{dm}T_sL = 0.001$. We can see from Figure 6.25 that as expected, the BER performance of both MMSE and zero-forcing BS transmit preprocessing is improved upon increasing the SNR, regardless whether perfect or prediced CIR taps are used. Apart from the obvious effect of noise reduction at the detector, this BER performance improvement is also due to the associated more accurate prediction, when using predicted CIR taps. Moreover, observe in Figure 6.25 for a given number of antennas at the BS the achievable BER performance of MMSE preprocessing is better than that of zero-forcing preprocessing for both the perfect and predicted CIR scenarios. This is because the effects of both the MUI and AWGN are jointly taken into account by the MMSE preprocessing. Furthmore, the performance discrepancy between MMSE and zero-forcing preprocessing recorded in Figure 6.25 for both the perfect and predicted CIR taps scenarios becomes narrower upon increasing the number of transmitter antennas from

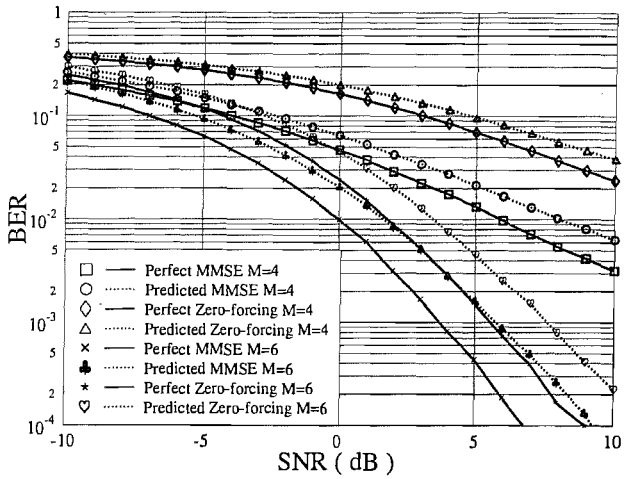


Figure 6.25: BER versus average SNR per symbol performance for downlink transmission for the first receive antenna of the first user when the number of transmitter antennas at BS is assumed to be 4 and 6, respectively, for both the MMSE and zero-forcing BS transmit preprocessing. Furthermore, the maximum normalized Doppler frequency is $f_{dm}T_sL = 0.001$. The remaining parameters are the same as in Table 6.4.

M =4 to 6. This is because zero-forcing preprocessing benefits more from the associated increased transmit diversity than MMSE preprocessing, since the transmitter diversity significantly mitigates not only the effect of the fading, but also those of the AWGN, as a benefit of noise averaging. Additionally, the discrepancy between the MMSE and zero-forcing based preprocessing techniques becomes narrower upon increasing the SNR for both the perfect and prediced CIR taps scenarios when considering the same number of antennas at BS. This is because MMSE preprocessing loses its benefits of mitigating the effects of the AWGN upon increasing the SNR. Moreover, as seen in Figure 6.25 the MMSE preprocessing using M =4 transmitter antennas outperforms the zero-forcing scheme for both the perfect and predicted CIR-tap scenarios, regardless of the SNR. Moreover, the MMSE preprocessor using M =4 transmitter antennas outperforms the zero-forcing scheme, even when M =6 tranmitter antennas are used in both the perfect and predicted CIR scenarios in the lower SNR range of Figure 6.25. This is because the MMSE preprocessing has the ability of mitigating the effects of the AWGN. Furthermore, for M =6 transmitter antennas the MMSE preprocessor using predicted CIR taps becomes capable of outperforming the zero-forcing preprocessor benefitting from perfect CIR tap in the lower range of SNR of Figure 6.25. The reason for this is because the detrimental

effects of CIR-tap prediction error are lower for MMSE preprocessing than the effects of the AWGN imposed on zero-forcing preprocessing using perfect CIR taps, especially for lower SNR range of Figure 6.25.

6.7.4 Effects of the Maximum Normalized Doppler Frequency

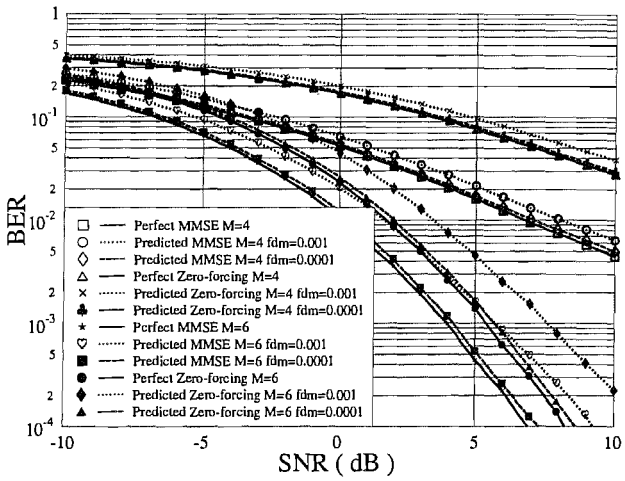


Figure 6.26: BER versus average SNR per symbol performance for downlink transmission for the first receive antenna of the first user, when the number of transmitter antennas at BS is assumed to be 4 and 6, respectively, for both the MMSE and zero-forcing BS transmit preprocessing. Furthermore the maximum normalized Doppler frequency is $f_{dm}T_sL = 0.001$ and $f_{dm}T_sL = 0.0001$, respectively. The remaining parameters are the same as in Table 6.4.

In Figure 6.26 the achievable BER performance is plotted against different values of the SNR for the first receive antenna of the first user, when the number of transmitter antennas at the BS is assumed to be M =4 and 6, respectively, for both MMSE and zero-forcing BS transmit preprocessing. The maximum normalized Doppler frequency was $f_{dm}T_sL = 0.001$ and $f_{dm}T_sL = 0.0001$, respectively. The remaining parameters are the same as in Table 6.4. We can see from Figure 6.26 that the BER performance is improved for both MMSE and zero-forcing preprocessing upon increasing the number of transmitter antennas from M =4 to 6 regardless of the maxmimum normalized Doppler frequency $f_{dm}T_sL$, which is a plausible benefit of the increased transmitter diversity gain. Additionally, for a given number of transmitter antennas, the attainable BER performance was improved for both the MMSE and zero-forcing preprocessing for a given number of predictor's order, when the

maximum normalized Doppler frequency was reduced from 0.001 to 0.0001. Furthermore, observe in Figure 6.26 that the BER performance recorded for predicted CIR taps was close to that of perfect CIR knowledge for both MMSE and zero-forcing preprocessing. Moreover, for M =6 transmitter antennas and $f_{dm}T_sL = 0.0001$ the BER performance of MMSE preprocessing using predicted CIR taps became better than that of zero-focring preprocessing using perfect CIR tap knowledge, regardless of the SNR. This is becasue the effect of CIR-tap prediction error on MMSE preprocessing using predicted CIR taps was always found to be lower compared to that of the AWGN imposed on zero-forcing preprocessing benefitting from perfect CIR tap knowledge in the case of slow fading.

6.8 Conclusion

Transmitter preprocessing has been investigated as a key technique of simplifying the MS's receiver. A crucial requirement for its success is the accurate and prompt knowledge of the CIR taps at the BS. The quantized and outdated CIR tap knowledge results in a performance degration. Hence, CIR-tap prediction becomes an essential technique in this situation. Furthermore, the CIR tap values extracted from the uplink transmissions can be used for predicting the downlink CIR taps in TDD systems. In this chapter, a TDD based SDMA system using a vector Kalman filtering assisted predictor was used at the BS for predicting the downlink CIR taps in order to invoke transmitter preprocessing. The MMSE criterion was adopted for designing the BS's transmit preprocessing matrix and two different forms of preprocessing, namely MMSE and zero-forcing were used corresponding to known or unknon backgroud AWGN variance, respectively. Our simulation results portrayed in Figures 6.14 to 6.26 demonstrated that with the aid of the Kalman filtering assisted channel predictor, BS transmitter preprocessing becomes capable of achieving an attractive performance. Furthermore, MMSE preprocessing was shown to outperform zero-forcing preprocessing as a benefit of its ability of mitigating the AWGN, as seen in Figures 6.25 and 6.26.

Chapter 7

Singular-Value-Decomposition Assisted Transmission and Detection in Multiuser MIMO Systems

7.1 Introduction

In MIMO-aided multiple users systems both the downlink (DL) and uplink (UL) transmissions experience multiple user interference (MUI), mutiple access interference (MAI) as well as inter-antenna interference (IAI). The optimum maximum likelihood (ML) receiver employed at the MS imposes a high computational complexity. In order to reduce the computational complexity of the MS, multiuser transmission techniques can be invoked at the BS [63, 151–153, 164–168]. Widely used linear preprocessing techniques, such as the MMSE and the zero-forcing arrangements were detailed in [151, 153, 165]. However, both the MMSE and zero-forcing techniques treat the composite channel matrix of all users jointly, rather than dealing with the individual user's channel. By contrast, in [152, 164], the effective channel of each user was determined by invoking the so-called block diagonalization technique at the BS, which removed the MUI, while in [168] a linear precoder was designed based on striking a compromise between the transmit power required and the SINR imposed. However, these schemes are only applicable for downlink transmission.

In fact, the optimum ML receiver may even be excessively complex for employment at the BS. By contrast, the linear MMSE or zero-forcing receiver is unable to take the individual users' specific channel into account. Furthermore, in [169] an MMSE-based criterion was used for designing the

preprocessing and postprocessing matrix, where cooperation of the MSs was required, while in [170] the so-called maximum ratio [171] uplink transmission scheme was investigated, where only the dominant right-hand-side (rhs) and left-hand-side (lhs) singular eigen vectors were adopted as the preprocessing and postprocessing eigen vectors, hence increasing the diversity gain at the cost of reducing the multiplexing gain.

It has been shown in [9, 10, 172–177] that when accurate and prompt CSI is available at both the transmitter and receiver, SVD-based adaptive modulation (AM) techniques applied in the context of MIMO systems are capable of achieving a high average spectral efficiency (ASE). Moreover, both SVD-assisted space time block coding (STBC) and V-BLAST have found numerous applications [178–180]. However, these proposals were based on point-to-point communcations. In the context of multiple users, SVD based MUD was discussed in [181, 182], when only the largest eigenvalue was invoked for uplink transmission, while in [183] multiple eigenvalue were invoked for downlink transmission, but only the inter-stream interference of the same user is cancelled with the aid of joint preprocessing and postprocessing.

In this chapter, both SVD-based SDMA MUDs designed for the uplink as well as multiple user downlink transmissions are investigated. When using combined SVD-based preprocessing and post-processing, both the multiple access interference (MAI) as well as the inter-stream interference of the same user can be completely removed in the case of uplink transmissions. Similarly, both the MUI as well as the inter-stream interference of a given user can be completely cancelled, when considering downlink transmissions. The proposed algorithm facilitates the employment of the adaptive modulation (AM) in the context of MIMO-aided multiple users and allows the extension of SVD-assisted STBC and V-BLAST to multiple user scenarios.

In contrast to the block diagonalization techniques of [152, 164], which only consider downlink transmissions, both the uplink and downlink have been considered by the algorithm proposed in this chapter. Furthermore, the individual users' channel was specifically taken into account, which is a substantial benefit in comparision to the transmit MMSE or zero-forcing technique. Another potential advantage is that only the knowledge of the singular vectors is required by the proposed algorithm, hence, subspace tracking or estimation algorithms may be used for directly tracking the preprocessing and postprocessing matrix, without regularly estimating the channel matrix and carrying out its SVD, which would result in a high complexity [130–132].

The chapter is structured as follows. In Section 7.2, SVD-based joint preprocessing and postprocessing designed for MIMO-aided SDMA MUD in the uplink is discussed. In Section 7.3 SVD-based joint preprocessing and postprocessing conceived for MIMO-assisted SDMA mupltiuser downlink

transmission is investigated, which is further discussed in a TDD context in Section 7.4. In Section 7.5 our simulation results are provided. Finally, our conclusions are offered in Section 7.6.

Note that, for the sake of simplifying our notation, in this chapter the variables without overbars are either related to the uplink or are common for both the uplink and downlink, while the variables having an overbar specifically denote the downlink.

7.2 SVD-Based Uplink Transmission and Detection

In this section we consider both uplink transmission and detection in a multiuser MIMO system, where the base-station (BS) supports multiple mobile-stations (MSs). Although the extension of these principles to other types of MIMO systems is straightforward, the multiuser MIMO system considered here is in fact a SDMA system, where both the BS and MSs may employ multiple antennas both for reception and transmission. In our study we assume that the BS is capable of acquiring the uplink channel impulse response (CIR) knowledge in the context of all the uplink users. By contrast, a MS is only capable of acquiring the uplink CIR knowledge of itself. Furthermore, we assume that there is no cooperation among the uplink users.

Since each of the MSs employs its corresponding uplink CIR knowledge, the SVD-based technique of [64] may hence be invoked, so that the transmission power can be optimally allocated to the transmit antennas of an uplink user in order to maximise the achievable performance. Specifically, in this section we investigate the uplink transmission power allocation under the criterion of maximal information rate per user [64] or of maximal signal-to-interference-plus-noise ratio (SINR) [88]. Again, there is no cooperation among the uplink users, hence the uplink users experience MAI, which may significantly degrade the achievable performance. Therefore, in this section the family of SVD-based techniques is also investigated, in order to mitigate the effects of MAI among the uplink users. It can be shown that the SVD-based MAI suppression satisfies the zero-forcing (ZF) condition, which is capable of entirely removing the residual uplink MAI.

7.2.1 Representation of the Uplink Signal

The schematic of the uplink multiuser MIMO system considered in this chapter is shown in Fig. 7.1, where the base-station (BS) employs M number of receive antennas and the kth ($k = 1, 2, \ldots, K$) mobile station (MS) employs N_k number of transmit antennas. In Fig. 7.1 $\boldsymbol{Q}_k$ ($k = 1, 2, \ldots, K$) represents the uplink transmitter's preprocessing matrix formulated for detecting the signal of the kth MS. The kth MS's data $\boldsymbol{x}_k$ is first preprocessed using $\boldsymbol{Q}_k$ before its transmission. In Fig. 7.1 $\boldsymbol{T}_k$

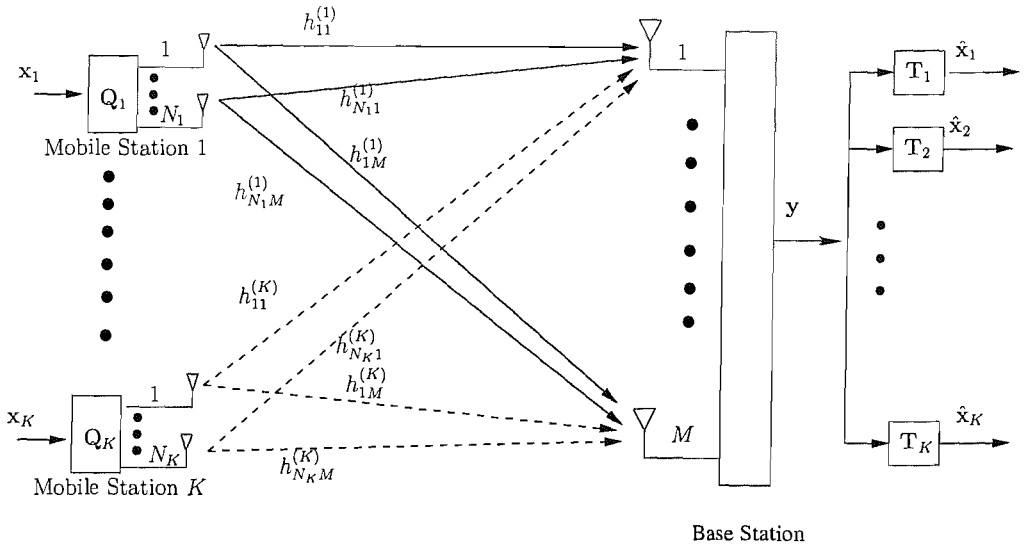


Figure 7.1: Schematic of an uplink multiuser MIMO system, where the BS employs M receive antennas, while the mobile stations (MSs) may employ different number of transmit antennas.

$(k = 1, 2, \ldots, K)$ represents the receiver's post-processing matrix formulated for detecting of the data transmitted by the kth MS. Let us now describe in detail both the transmission and detection schemes, which are based on the SVD.

Let the N_k data sysmbols to be transmitted by the kth MS to the BS be hosted by a vector expressed as [169]

$$\boldsymbol{x}_k = [x_{k1}, x_{k2}, \cdots, x_{kN_k}]^T, \; k = 1, 2, \ldots, K. \tag{7.1}$$

As shown in Fig. 7.1, $\boldsymbol{x}_k$ is preprocessed using the kth MS's transmitter preprocessing matrix $\boldsymbol{Q}_k$, yielding the output [169]

$$\boldsymbol{d}_k = \boldsymbol{Q}_k \boldsymbol{x}_k, \; k = 1, 2, \ldots, K. \tag{7.2}$$

Let the CIR matrix connecting the N_k transmit antennas of the kth MS with the M receive antennas

at the BS be expressed as

$$\boldsymbol{H}_k = \begin{bmatrix} h_{11}^{(k)} & h_{12}^{(k)} & \cdots & h_{1N_k}^{(k)} \\ h_{21}^{(k)} & h_{22}^{(k)} & \cdots & h_{2N_k}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1}^{(k)} & h_{M2}^{(k)} & \cdots & h_{MN_k}^{(k)} \end{bmatrix}, \ k = 1, 2, \ldots, K, \tag{7.3}$$

which is a $(M \times N_k)$-component matrix. Then, the received length-M observation vector $\boldsymbol{y}$ at the BS can be expressed as [169]

$$\begin{aligned} \boldsymbol{y} &= \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{d}_k + \boldsymbol{n} \\ &= \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{x}_k + \boldsymbol{n}, \end{aligned} \tag{7.4}$$

where $\boldsymbol{n}$ is a length-M noise observation vector, which is assumed to be Gaussian distributed with a zero mean and a covariance matrix given by $\sigma^2 \boldsymbol{I}_M$.

Let us define

$$\begin{aligned} \boldsymbol{H} &= [\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_K]^T, \\ \boldsymbol{Q} &= \text{diag}\{\boldsymbol{Q}_1, \boldsymbol{Q}_2, \cdots, \boldsymbol{Q}_K\}, \\ \boldsymbol{d} &= [\boldsymbol{d}_1^T, \boldsymbol{d}_2^T, \cdots, \boldsymbol{d}_K^T]^T, \\ \boldsymbol{x} &= [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_K^T]^T, \end{aligned} \tag{7.5}$$

where $\boldsymbol{H}$ is the $(\sum_{k=1}^{K} N_k \times M)$-component combined channel matrix of the uplink, $\boldsymbol{Q}$ is the $(\sum_{k=1}^{K} N_k \times \sum_{k=1}^{K} N_k)$-component overall preprocessing matrix, $\boldsymbol{d}$ is the length-$(\sum_{k=1}^{K} N_k)$ overall preprocessed data vector and $\boldsymbol{x}$ is the length-$(\sum_{k=1}^{K} N_k)$ overall transmitted data vector.

Then, (7.4) can also be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{H}\boldsymbol{d} + \boldsymbol{n} \\ &= \boldsymbol{H}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{n}. \end{aligned} \tag{7.6}$$

As shown in Fig. 7.1, at the BS's receiver the observation vector $\boldsymbol{y}$ is processed in order to generate the estimates of the transmitted data symbols. Specifically, the kth MS's transmitted data is recovered

by processing the observation vector y using a $(N_k \times M)$-component weight matrix $\boldsymbol{T}_k$, which can be expressed as

$$\hat{\boldsymbol{x}}_k = \boldsymbol{T}_k \boldsymbol{y}, \; k = 1, 2, \ldots, K. \tag{7.7}$$

Let us collect all the estimates of the K uplink users into a single vector $\hat{\boldsymbol{x}}$, which is defined as

$$\hat{\boldsymbol{x}} = \left[\hat{\boldsymbol{x}}_1^T, \hat{\boldsymbol{x}}_2^T, \cdots, \hat{\boldsymbol{x}}_K^T\right]^T. \tag{7.8}$$

Furthermore, let the overall $(\sum_{k=1}^{K} N_k \times M)$-component weight matrix $\boldsymbol{T}$ be formulated as

$$\boldsymbol{T} = \left[\boldsymbol{T}_1^T, \boldsymbol{T}_2^T, \ldots, \boldsymbol{T}_K^T\right]^T. \tag{7.9}$$

Then, it can be shown that (7.8) may be expressed as

$$\begin{aligned} \hat{\boldsymbol{x}} &= \boldsymbol{T}\boldsymbol{y} \\ &= \boldsymbol{T}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{T}\boldsymbol{n}, \end{aligned} \tag{7.10}$$

where the matrix operation of $\boldsymbol{T}$ is referred to as receiver post-processing [60] in contrast to the transmitter preprocessing operation of $\boldsymbol{Q}$ [60, 153].

For the traditional zero-forcing receiver [184], which does not use transmit preprocessing, the overall preprocessing matrix $\boldsymbol{Q}$ and the overall postprocessing matrix $\boldsymbol{T}$ are expressed, respectively, as

$$\boldsymbol{Q} = \boldsymbol{I}_{\sum_{k=1}^{K} N_k}, \tag{7.11}$$

$$\boldsymbol{T} = (\boldsymbol{H}^H \boldsymbol{H})^{-1} \boldsymbol{H}^H. \tag{7.12}$$

Upon substituting (7.11) and (7.12) into (7.10), we have

$$\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{n}', \tag{7.13}$$

where

$$n' = (H^H H)^{-1} H^H n. \tag{7.14}$$

The covariance matrix R of n' is given by

$$R = E[n'(n')^H] = \sigma^2 (H^H H)^{-1}. \tag{7.15}$$

Above we have derived the expression of the discrete-time signal received at the BS, as shown in (7.6), when the MS's transmitter employs transmitter preprocessing. After the receiver post-processing seen in Fig. 7.1, the decision variable vector representing the transmitted data of the K uplink users is given by (7.10). Below we consider both the uplink preprocessing as well as detection and derive the expressions for both the transmitter preprocessing matrix $\boldsymbol{Q}$ and for the receiver post-processing matrix $\boldsymbol{T}$ using the classic SVD principles. In our derivation we assume that we have $M \geq \sum_{k=1}^{K} N_k$, which physically means that the number of antennas at the BS is equal to or higher than the sum of antennas of all the K MSs.

7.2.2 Uplink Transmitter Preprocessing Based on SVD

The uplink transmitter preprocesing is employed for facilitating efficient transmission power allocation and to assist the BS receiver in reliably detecting the uplink signals. Our study shows that the system employing transmission power allocation is capable of significantly outperforming that dispensing with it. Let us assume that H_k of (7.3) satisfies rank $(H_k) = N_k$. Then, the SVD of H_k can be expressed as

$$\begin{aligned} H_k &= \boldsymbol{U}_k \begin{bmatrix} \boldsymbol{\Lambda}_k^{1/2} \\ \mathbf{0} \end{bmatrix} \boldsymbol{V}_k^H \\ &= [\boldsymbol{U}_{ks}\, \boldsymbol{U}_{kn}] \begin{bmatrix} \boldsymbol{\Lambda}_k^{1/2} \\ \mathbf{0} \end{bmatrix} \boldsymbol{V}_k^H \qquad (7.16) \\ &= \boldsymbol{U}_{ks} \boldsymbol{\Lambda}_k^{1/2} \boldsymbol{V}_k^H, \; k = 1, 2, \ldots, K, \qquad (7.17) \end{aligned}$$

where $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are $(M \times M)$ and $(N_k \times N_k)$-component unitary matrices, respectively, while $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_{N_k}\}$ contains the N_k non-zero eigenvalues of $H_k^H H_k$ or $H_k H_k^H$. Furthermore, in (7.16) the columns of $\boldsymbol{U}_k$ are constituted by the eigenvectors of $H_k H_k^H$, $\boldsymbol{U}_{ks}$ consists of the N_k

eigenvectors corresponding to the signal subspace of $\boldsymbol{H}_k\boldsymbol{H}_k^H$, while $\boldsymbol{U}_{kn}$ consists of the $(M - N_k)$ eigenvectors corresponding to the null subspace of $\boldsymbol{H}_k\boldsymbol{H}_k^H$. Similarly, the columns of $\boldsymbol{V}_k$ correspond to the eigenvectors of $\boldsymbol{H}_k^H\boldsymbol{H}_k$.

Upon substituting (7.17) into (7.4), the vector $\boldsymbol{y}$ of the received signal seen in Figure 7.1 can be expressed as

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{U}_{ks}\boldsymbol{\Lambda}_k^{1/2}\boldsymbol{V}_k^H\boldsymbol{Q}_k\boldsymbol{x}_k + \boldsymbol{n}, \tag{7.18}$$

where the channel matrix of the kth user $\boldsymbol{H}_k$ is replaced by its SVD. Let the transmitter preprocessing matrix $\boldsymbol{Q}_k$ of Figure 7.1 be formulated as

$$\boldsymbol{Q}_k = \boldsymbol{V}_k\boldsymbol{\beta}_k, \; k = 1, 2, \ldots, K, \tag{7.19}$$

where $\boldsymbol{\beta}_k = \text{diag}\{\beta_{k1}, \cdots, \beta_{kN_k}\}$ is a $(N_k \times N_k)$-component diagonal matrix, which is employed for implementing the transmission power-allocation, as we will discuss in detail in our forthcoming discourse in Subsection 7.2.4. After substituting (7.19) into (7.18) and exploiting the property $\boldsymbol{V}_k^H\boldsymbol{V}_k = \boldsymbol{I}_{N_k}$, the vector $\boldsymbol{y}$ of the received signal seen in Figure 7.1 can be simplified to

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{U}_{ks}\boldsymbol{\Lambda}_k^{1/2}\boldsymbol{\beta}_k\boldsymbol{x}_k + \boldsymbol{n}, \tag{7.20}$$

where the right-hand-side (rhs) singular vectors of the channel matrix $\boldsymbol{H}_k$ of the kth user has been cancelled out by the corresponding preprocessing matrix $\boldsymbol{Q}_k$ of Figure 7.1 at the Kth MS.

Equation (7.20) shows that the preprocessing operation represented by the preprocessing matrix, say $\boldsymbol{Q}_k$, of (7.19) decouples the N_k transmitted data symbols of the kth MS from the transmitter side. In other words, as shown in (7.20), the preprocessing of $\boldsymbol{x}_k$ projects the N_k symbols of $\boldsymbol{x}_k$ onto N_k different orthogonal subspaces. In this case a specific symbol $\boldsymbol{x}_k$ can be readily detected on the basis of the subspace it belongs to without encountering interference from the other symbols of MS k.

Let us define

$$\begin{aligned} \boldsymbol{U}_s &= [\boldsymbol{U}_{1s}, \boldsymbol{U}_{2s}, \cdots, \boldsymbol{U}_{Ks}], \\ \boldsymbol{\Lambda}^{1/2} &= \text{diag}\left\{\boldsymbol{\Lambda}_1^{1/2}, \boldsymbol{\Lambda}_2^{1/2}, \cdots, \boldsymbol{\Lambda}_K^{1/2}\right\}, \\ \boldsymbol{\beta} &= \text{diag}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_K\}, \end{aligned} \tag{7.21}$$

where the $(M \times \sum_{k=1}^{K} N_k)$-component matrix $\boldsymbol{U}_s$ consists of K signal spaces of $\boldsymbol{H}_k \boldsymbol{H}_k^H$, the $(\sum_{k=1}^{K} N_k \times \sum_{k=1}^{K} N_k)$-component matrix $\boldsymbol{\Lambda}^{1/2}$ is composed of the K number of diagonal matrices $\boldsymbol{\Lambda}_k^{1/2}$, while the $(\sum_{k=1}^{K} N_k \times \sum_{k=1}^{K} N_k)$-component diagonal matrix $\boldsymbol{\beta}$ hosts the power allocation coefficients for all the K MSs.

Then, the received signal vector $\boldsymbol{y}$ of Figure 7.1 can be expressed as

$$\boldsymbol{y} = \boldsymbol{U}_s \boldsymbol{\Lambda}^{1/2} \boldsymbol{\beta} \boldsymbol{x} + \boldsymbol{n}. \tag{7.22}$$

Note that although the columns of $\boldsymbol{U}_{ks}$ $(k = 1, 2, \ldots, K)$ are orthogonal, suggesting that there is no IAI, the columns of $\boldsymbol{U}_s$ in (7.22) corresponding to the different users are non-orthogonal. Therefore, there is MAI, which should be cancelled by the BS's receiver. Let us now consider the issues of uplink detection.

7.2.3 Uplink Detection

Upon left multiplying the overall transmit postprocessing matrix $\boldsymbol{T}$ with the received signal vector $\boldsymbol{y}$ of (7.22), the decision variable vector of (7.10) formulated for all the K users can be expressed as

$$\hat{\boldsymbol{x}} = \boldsymbol{T} \boldsymbol{U}_s \boldsymbol{\Lambda}^{1/2} \boldsymbol{\beta} \boldsymbol{x} + \boldsymbol{T} \boldsymbol{n}. \tag{7.23}$$

It can be shown that there are many alternatives for the design of the receiver post-processing matrix $\boldsymbol{T}$, as discussed in [184]. As an example, in this chapter we focus our attention on the zero-forcing (ZF) detection scheme, which is a linear detector and is capable of entirely removing the MAI.

The ZF solution encapsulated in $\boldsymbol{T}$ can be readily derived in the context of [184]

$$\boldsymbol{T} = \begin{cases} [\boldsymbol{U}_s]^+ = \left(\boldsymbol{U}_s^H \boldsymbol{U}_s\right)^{-1} \boldsymbol{U}_s^H, & \text{if } \sum_{k=1}^{K} N_k < M; \\ [\boldsymbol{U}_s]^{-1}, & \text{if } \sum_{k=1}^{K} N_k = M, \end{cases} \tag{7.24}$$

where $[\cdot]^+$ denotes the pseudo inverse of the matrix $\boldsymbol{U}_s$. Upon substituting (7.24) into (7.23), we arrive at

$$\hat{\boldsymbol{x}} = \boldsymbol{\Lambda}^{1/2} \boldsymbol{\beta} \boldsymbol{x} + \boldsymbol{n}'. \tag{7.25}$$

Explicitly, the MAI is entirely removed. In (7.25) the noise term n' is given by

$$n' = Tn, \tag{7.26}$$

which still represents a Gaussian noise vector with zero mean, but its covariance matrix is given by

$$E\left[n'(n')^H\right] = \sigma^2 \left(U_s^H U_s\right)^{-1}, \tag{7.27}$$

which indicates that the noise observations become correlated after receiver post-processing.

Since the signals transmitted from a given MS are decoupled by its transmitter preprocessing, it can be shown that the diagonal entries of $U_s^H U_s$ are constituted by K unity matrices having the sizes of $(N_k \times N_k)$ for $k = 1, 2, \ldots, K$, respectively. Therefore, a given MS does not impose correlation on its own noise samples. This property makes it possible for us to study the power-allocation for a specific MS without considering the correlation among the noise observation samples. Let us now consider the UL power-allocation in the next section.

7.2.4 Power-Allocation

In the context of uplink transmission, the originally allocated transmission power of the kth MS is given by $E\left[\|\boldsymbol{x}_k\|^2\right]$. Hence the power-allocation is carried out under the constraint of

$$E\left[\|\boldsymbol{\beta}_k \boldsymbol{x}_k\|^2\right] \leq E\left[\|\boldsymbol{x}_k\|^2\right], \tag{7.28}$$

which means that the transmission power after transmit preprocessing cannot exceed the originally allocated power. Let us assume that the transmitted symbols are normalized to satisfy $E\left[\|x_{ki}\|^2\right] = 1$. Then, the above constraint can be written as

$$\sum_{i=1}^{N_k} \beta_{ki}^2 = N_k. \tag{7.29}$$

In our the power-allocation study we assume that a MS employs the knowledge of the MIMO channels connecting its transmit antennas with the M BS receive antennas, i.e. that the kth MS has the knowledge of $\boldsymbol{H}_k$. Therefore, the kth MS has the knowledge of $\boldsymbol{\Lambda}_k$. However, the kth MS is unable to acquire any knowledge about the other MSs' CIRs. By contrast, the BS is capable of acquiring channel knowledge in the context of all the MSs, in order to carry out zero-forcing based detection.

In this section power-allocation is carried out in the uplink multiuser MIMO system either based on maximizing the information rate [64] or on maximizing the overall SNR [88]. Let us first consider the power-allocation scheme, which achieves the maximal information rate for an individual MS.

7.2.4.1 Maximum Information Rate Based Power-Allocation

Since in the considered system there is no cooperation among the MSs, a MS can only exploit the knowledge of it own CIR and its shared power can only be across to its own transmit antennas, while ignoring the existance of all the other MSs. In this case the 'water-filling' principle [64] may be employed for allocating the MS's total transmission power, in order to maximize the achievable information rate of the corresponding user. Specifically, as shown in (7.25), the decision variable vector of the kth MS can be expressed as

$$\hat{\boldsymbol{x}}_k = \boldsymbol{\Lambda}_k^{1/2} \boldsymbol{\beta}_k \boldsymbol{x}_k + \boldsymbol{n}'_k, \; k = 1, 2, \ldots, K. \tag{7.30}$$

When treating the components of $\boldsymbol{n}'_k$ as independent Gaussian random variables, it can be shown that, in order to maximize the achievable information rate, the power-allocation related matrix $\boldsymbol{\beta}_k$ should be chosen according to [64]

$$\begin{aligned} \boldsymbol{\beta}_k &= \max_{\sum_{i=1}^{N_k} \beta_{ki}^2 = N_k} \left\{ I\left(\hat{\boldsymbol{x}}_k, \boldsymbol{x}_k\right) = \log_2 \left[\det \left(\boldsymbol{I}_{N_k} + \boldsymbol{\Lambda}_k \boldsymbol{\beta}_k^2 \right) \right] \right\} \\ &= \max_{\sum_{i=1}^{N_k} \beta_{ki}^2 = N_k} \left\{ I\left(\hat{\boldsymbol{x}}_k, \boldsymbol{x}_k\right) = \sum_{i=1}^{N_k} \log_2 \left[\left(1 + \lambda_{ki} \beta_{ki}^2\right) \right] \right\}, \; k = 1, 2, \ldots, K, \end{aligned} \tag{7.31}$$

where $I\left(\hat{\boldsymbol{x}}_k, \boldsymbol{x}_k\right)$ denotes the channel capacity associated with $\boldsymbol{H}_k$, which quantifies the maximum number of error free information bits per channel use that can be transmitted over the channel $\boldsymbol{H}_k$ for the given noise covariance matrix, when assuming that the noise power is unity. Furthermore, since the channel matrix $\boldsymbol{H}_k$ is decomposed into N_k parallel SISO channels and $\boldsymbol{n}'_k$ is assumed to represent indepedent Gaussian random variables, the channel capacity associated with $\boldsymbol{H}_k$ can be expressed as the sum of N_k SISO channels' capacity.

Upon exploiting the 'water-filling' principle [64], the information rate of (7.31) can be maximized, if the coefficients $\{\beta_{ki}\}$ of the $(N_k \times N_k)$-component diagonal matrix of (7.19) hosting the power allocation coefficients are chosen as [64]

$$\beta_{ki}^2 = \left(v_k - \frac{1}{\lambda_{ki}} \right)^+, \; i = 1, 2, \ldots, N_k \tag{7.32}$$

where $(x)^+$ is defined as

$$(x)^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{7.33}$$

and v_k is chosen so that

$$\sum_{i=1}^{N_k} \beta_{ki}^2 = \sum_{i=1}^{N_k} \left(v_k - \frac{1}{\lambda_{ki}} \right)^+ \leq N_k \tag{7.34}$$

is satisfied.

Furthermore, when substituting (7.32) into (7.31), it can be shown that the maximum information rate of the kth MS is given by [64]

$$I_{\max}(\hat{\boldsymbol{x}}_k, \boldsymbol{x}_k) = \sum_{i=1}^{N_k} \left(\log_2 \left[v_k \lambda_{ki} \right] \right)^+, \; k = 1, 2, \ldots, K. \tag{7.35}$$

The above result has been obtained, when assuming that the noise power is unity. When the signal-to-noise ratio (SNR) is known, the maximum information rate normalized by the number of transmit antennas of MS k can be expressed as [147]

$$\mathcal{I}_{\max}(\hat{\boldsymbol{x}}_k, \boldsymbol{x}_k) = \frac{1}{N_k} \sum_{i=1}^{N_k} \log_2 \left[\left(1 + \frac{\lambda_{ki}}{\sigma^2} \left(v_k - \frac{\sigma^2}{\lambda_{ki}} \right)^+ \right) \right], \; k = 1, 2, \ldots, K \tag{7.36}$$

which denotes the maximum number of error free information bits per channel use for the channel $\boldsymbol{H}_k$ at a given SNR, provided that the noise samples observed at the BS's receiver are independent.

When the power-allocation regime of (7.32) and assuming noise samples are correlated, the maximal achievable rate must be modified as follows. Let us denote the the covariance matrix of the noise samples by

$$\boldsymbol{R}_k = E \left[\boldsymbol{n}_k' (\boldsymbol{n}_k')^H \right]. \tag{7.37}$$

The maximal achievable normalized rate can be expressed as [185]

$$\mathcal{I}_{\max}(\hat{\boldsymbol{x}}_k, \boldsymbol{x}_k) = \frac{1}{N_k} \log_2 \left[\det \left(\boldsymbol{I}_{N_k} + \boldsymbol{\Lambda}_k \boldsymbol{\beta}_k^2 \boldsymbol{R}_k^{-1} \right) \right], \; k = 1, 2, \ldots, K \tag{7.38}$$

where the diagonal entries of $\boldsymbol{\beta}_k^2$ are given by (7.36). From (7.38) we can see that the channel capacity

associated with $\boldsymbol{H}_k$ is not treated as the sum of N_k SISO channel's capacity, since the correlation among the noise components is known at the receiver and the receiver is capable of exploiting this correlation for improving the channel capacity of $\boldsymbol{H}_k$. The maximum attainable information rate of (7.38) is achieved, when the receiver is capable of exploiting the knowledge of $\boldsymbol{R}_k$ for the detection of the kth user, despite having no knowledge concerning the correlation of the noise experienced by the different users, since there is no cooperation among the MSs.

Note furthermore that when the conventional zero-forcing detection of (7.13) is considered, the achievable normalized capacity can be expressed as [186]

$$\mathcal{I}_{\max}(\hat{\boldsymbol{x}}, \boldsymbol{x}) = \frac{1}{\sum_{k=1}^{K} N_k} \sum_{i=1}^{\sum_{k=1}^{K} N_k} \log_2 \left[\left(1 + \frac{1}{[\boldsymbol{R}]_{(i,i)}} \right) \right], \tag{7.39}$$

where $\boldsymbol{R}$ is a noise auto-correlation matrix of the noise given by (7.15).

7.2.4.2 Power-Allocation Designed for Achieving the Maximum SNR

For this scenario, we approximate the entries of the vector $\boldsymbol{n}'_k$ as the independent identically distributed (i.i.d) Gaussian random variables having a common variance of $\sigma^2/2$ per dimension. Given the decision variable vector of (7.30), the SNR of the ith antenna's symbol of MS k can be expressed as

$$\gamma_{ki} = \frac{\lambda_{ki}\beta_{ki}^2}{\sigma^2}, \; i = 1, 2, \ldots, N_k; \; k = 1, 2, \ldots, K \tag{7.40}$$

and the sum of the SNRs of all the N_k antennas is given by $\gamma_k = \sum_{i=1}^{N_k} \gamma_{ki}$. However, it can be shown that maximizing γ_k does not result in a meaningful solution [88]. Hence we opt the following expression for minimizing [88]

$$\xi_k = \sum_{i=1}^{N_k} \frac{1}{\gamma_{ki}} = \sum_{i=1}^{N_k} \frac{\sigma^2}{\lambda_{ki}\beta_{ki}^2} \tag{7.41}$$

under the constraint of (7.29) for the total transmission power. In this case the Lagrangian constrained optimazation cost-function [88] can be formulated as

$$J = \sum_{i=1}^{N_k} \frac{1}{\gamma_{ki}} + \mu \left(\sum_{i=1}^{N_k} \beta_{ki}^2 - N_k \right) = \sum_{i=1}^{N_k} \frac{\sigma^2}{\lambda_{ki}\beta_{ki}^2} + \mu \left(\sum_{i=1}^{N_k} \beta_{ki}^2 - N_k \right). \tag{7.42}$$

Upon taking the derivative of J in (7.42) with respect to β_{ki} and setting the result to zero gives

$$\beta_{ki}^2 = \sqrt{\frac{\sigma^2}{\mu\lambda_{ki}}},\ i = 1, 2, \ldots, N_k, \tag{7.43}$$

where μ can be obtained from the power constraint of (7.29), which gives

$$\frac{1}{\sqrt{\mu}} = \left(\sum_{i=1}^{N_k} \sqrt{\frac{\sigma^2}{\lambda_{ki}}}\right)^{-1} N_k. \tag{7.44}$$

Finally, when substituting the above result into (7.43), we arrive at the power-allocation scheme optimized for maintaining the maximum achievable SNR in the form of

$$\begin{aligned}\beta_{ki}^2 &= N_k \left(\sum_{i=1}^{N_k} \sqrt{\frac{\sigma^2}{\lambda_{ki}}}\right)^{-1} \sqrt{\frac{\sigma^2}{\lambda_{ki}}} \\ &= N_k \left(\sum_{i=1}^{N_k} \frac{1}{\sqrt{\lambda_{ki}}}\right)^{-1} \frac{1}{\sqrt{\lambda_{ki}}},\ i = 1, 2, \ldots, N_k\end{aligned} \tag{7.45}$$

where β_{ki}^2 is the power allocated to the ith antenna of the kth MS, which is only related to the N_k eigenvalues of the channel matrix $\boldsymbol{H}_k$.

Upon comparing (7.45) to (7.32), we can see that the power-allocation scheme of (7.32) which is derived for the maximal information rate criterion assigns more transmission power to the transmit antennas having good channel conditions. By contrast, the power-allocation scheme of (7.45) optimized based on the maximum SNR principle assigns less transmission power to the transmit antennas benefitting from having good channel conditions, so that the SNR-related quantity of (7.41) can be minimized.

Upon substituting (7.45) into (7.40), the SNR experienced by the ith antenna's symbol of MS k is given by

$$\gamma_{ki} = \frac{N_k\sqrt{\lambda_{ki}}}{\sigma^2} \left(\sum_{i=1}^{N_k} \frac{1}{\sqrt{\lambda_{ki}}}\right)^{-1},\ i = 1, 2, \ldots, N_k;\ k = 1, 2, \ldots, K. \tag{7.46}$$

Furthermore, when the BPSK modulation is employed, the BER can be expressed as

$$P_b(\gamma_{ki}) = Q\left(\sqrt{2\gamma_{ki}}\right),\ i = 1, 2, \ldots, N_k;\ k = 1, 2, \ldots, K. \tag{7.47}$$

Note that there are other strategies [126, 187], which can also be invoked for optimizing the transmission power allocation. For example, the transmission power can be allocated, in order to achieve the minimum BER (MBER) [187]. It can also be allocated for minimizing the transmission power required maintaining a given data rate in the context of speficic modulation schemes. Furthermore, adaptive-rate modulation [6] can be invoked for practically achieving a transmission rate that is as high as possible.

Having considered uplink transmissions, let us now consider the downlink in the next section.

7.3 SVD-Based Downlink Transmission and Detection

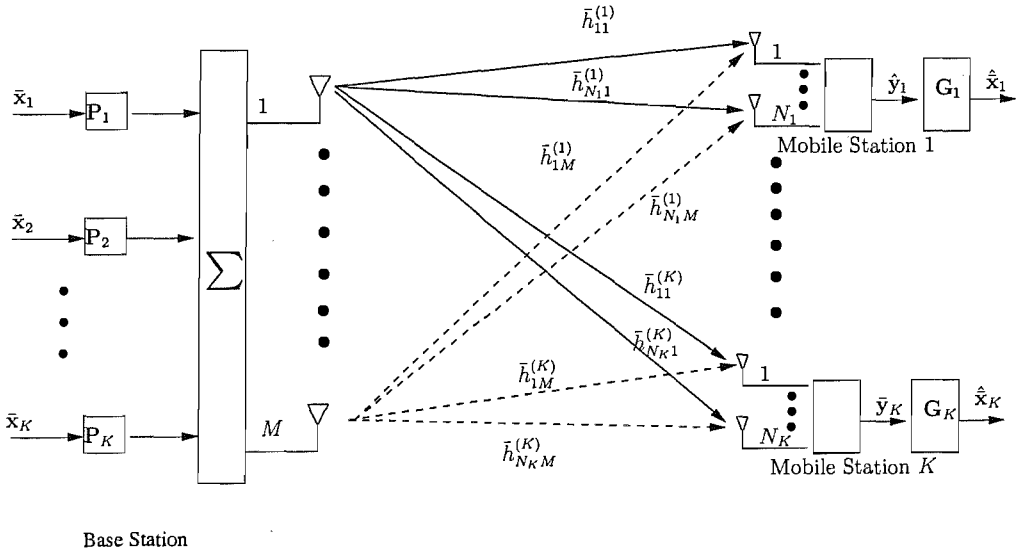


Figure 7.2: Schematic of SDMA downlink transmission using both preprocessing and postprocessing.

Similarly to the uplink, the downlink system considered has a single BS supporting K MSs as shown in Fig. 7.2. The BS is equipped with M transmit antennas, while the kth ($k = 1, 2, \ldots, K$) MS has N_k receive antennas. Furthermore, we assume that the channel between any pair of transmit and receive antennas is flat-fading. Note again that in this chapter the variables marked with overbars are specifically reserved for the downlink.

Let the N_k-component DL symbol vector $\bar{\boldsymbol{x}}_k$ to be transmitted to the kth MS be expressed as

$$\bar{\boldsymbol{x}}_k = [\bar{x}_{k1}, \bar{x}_{k2}, \ldots, \bar{x}_{kN_k}]^T. \tag{7.48}$$

As shown in Figure 7.2, $\bar{x}_k$ is preprocessed before its transmission by premutiplying it with an $(M \times N_k)$-component preprocessing matrix P_k, yielding

$$\bar{d}_k = \boldsymbol{P}_k \bar{\boldsymbol{x}}_k, \; k = 1, 2, \ldots, K. \tag{7.49}$$

After transmitter preprocessing, the M-component signal broadcast by the BS to the K MSs can be expressed as

$$\begin{aligned} \bar{\boldsymbol{d}} &= \sum_{k=1}^{K} \bar{\boldsymbol{d}}_k = \sum_{k=1}^{K} \boldsymbol{P}_k \bar{\boldsymbol{x}}_k \\ &= \boldsymbol{P} \bar{\boldsymbol{x}}, \end{aligned} \tag{7.50}$$

where $\boldsymbol{P}$ is a $\left(M \times \sum_{k=1}^{K} N_k\right)$-component matrix given by

$$\boldsymbol{P} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_K] \tag{7.51}$$

and $\bar{\boldsymbol{x}}$ is a $\left(\sum_{k=1}^{K} N_k\right)$-component vector containing the transmitted data, which is given by

$$\bar{\boldsymbol{x}} = \left[\bar{\boldsymbol{x}}_1^T, \bar{\boldsymbol{x}}_2^T, \ldots, \bar{\boldsymbol{x}}_K^T\right]^T. \tag{7.52}$$

When the signal of (7.50) is transmitted over the downlink MIMO channel, as shown in Figure 7.2, the received N_k-component vector $\bar{\boldsymbol{y}}_k$ of the kth MS can be expressed as

$$\begin{aligned} \bar{\boldsymbol{y}}_k &= \bar{\boldsymbol{H}}_k \bar{\boldsymbol{d}} + \bar{\boldsymbol{n}}_k \\ &= \bar{\boldsymbol{H}}_k \boldsymbol{P} \bar{\boldsymbol{x}} + \bar{\boldsymbol{n}}_k \\ &= \sum_{i=1}^{K} \bar{\boldsymbol{H}}_k \boldsymbol{P}_i \bar{\boldsymbol{x}}_i + \bar{\boldsymbol{n}}_k \\ &= \bar{\boldsymbol{H}}_k \boldsymbol{P}_k \bar{\boldsymbol{x}}_k + \sum_{i=1, i \neq k}^{K} \bar{\boldsymbol{H}}_k \boldsymbol{P}_i \bar{\boldsymbol{x}}_i + \bar{\boldsymbol{n}}_k, \; k = 1, 2, \ldots, K \end{aligned} \tag{7.53}$$

where $\bar{\boldsymbol{n}}_k$ is an N_k-length AWGN vector having zero mean and a covariance matrix of $E[\bar{\boldsymbol{n}}_k \bar{\boldsymbol{n}}_k^H] = \sigma^2 \boldsymbol{I}_{N_k}$, while $\bar{\boldsymbol{H}}_k$ is a $(N_k \times M)$-component channel transfer matrix connecting the M antennas of

the BS with the kth MS, which can be expressed as

$$\bar{\boldsymbol{H}}_k = \begin{bmatrix} \bar{h}_{11}^{(k)} & \bar{h}_{12}^{(k)} & \cdots & \bar{h}_{1M}^{(k)} \\ \bar{h}_{21}^{(k)} & \bar{h}_{22}^{(k)} & \cdots & \bar{h}_{2M}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{h}_{N_k1}^{(k)} & \bar{h}_{N_k2}^{(k)} & \cdots & \bar{h}_{N_kM}^{(k)} \end{bmatrix}, \tag{7.54}$$

where $\bar{h}_{ij}^{(k)}$ represents the CIR coefficients between the jth BS transmit antenna and the ith receive antenna of the kth MS. As we can see from (7.53), the downlink MSs experience multiuser inferference (MUI).

Let us assume that the rows of $\bar{\boldsymbol{H}}_k$ $(k = 1, 2, \ldots, K)$ have full rank, i.e. we have rank $\left(\bar{\boldsymbol{H}}_k\right) = N_k$, and that $M \geq \sum_{k=1}^{K} N_k$. Then, upon carrying out the SVD of $\bar{\boldsymbol{H}}_k$, we arrive at

$$\begin{aligned} \bar{\boldsymbol{H}}_k &= \bar{\boldsymbol{U}}_k \left[\bar{\boldsymbol{\Lambda}}_k^{1/2}, \boldsymbol{0}\right] \bar{\boldsymbol{V}}_k^H \\ &= \bar{\boldsymbol{U}}_k \left[\bar{\boldsymbol{\Lambda}}_k^{1/2}, \boldsymbol{0}\right] \begin{bmatrix} \bar{\boldsymbol{V}}_{ks}^H \\ \bar{\boldsymbol{V}}_{kn}^H \end{bmatrix} \\ &= \bar{\boldsymbol{U}}_k \bar{\boldsymbol{\Lambda}}_k^{1/2} \bar{\boldsymbol{V}}_{ks}^H, \end{aligned} \tag{7.55}$$

where $\bar{\boldsymbol{U}}_k$ and $\bar{\boldsymbol{V}}_k$ are $(N_k \times N_k)$-component and $(M \times M)$-component unitary matrices, respectively, $\bar{\boldsymbol{\Lambda}}_k$ is a $(N_k \times N_k)$-component diagonal matrix containing the eigenvalues of $\bar{\boldsymbol{H}}_k \bar{\boldsymbol{H}}_k^H$, i.e. we have $\bar{\boldsymbol{\Lambda}}_k = \text{diag}\{\bar{\lambda}_{k1}, \bar{\lambda}_{k2}, \cdots, \bar{\lambda}_{kN_k}\}$. Furthermore, in (7.55) $\bar{\boldsymbol{V}}_{ks}$ is a $(M \times N_k)$-component matrix, which is constituted by the eigenvectors corresponding to the non-zero eigenvalues of $\bar{\boldsymbol{H}}_k^H \bar{\boldsymbol{H}}_k$. By contrast, $\bar{\boldsymbol{V}}_{kn}$ is a $(M \times (M - N_k))$-component matrix, which is constituted by the eigenvectors corresponding to the zero eigenvalues of $\bar{\boldsymbol{H}}_k^H \bar{\boldsymbol{H}}_k$. Similarly, $\bar{\boldsymbol{U}}_k$ consists of the eigenvectors of $\bar{\boldsymbol{H}}_k \bar{\boldsymbol{H}}_k^H$.

Upon substituting (7.55) into the second equation in (7.53), the received signal $\bar{\boldsymbol{y}}_k$ of the kth MS seen in Figure 7.2 may be expressed as

$$\bar{\boldsymbol{y}}_k = \bar{\boldsymbol{U}}_k \bar{\boldsymbol{\Lambda}}_k^{1/2} \bar{\boldsymbol{V}}_{sk}^H \boldsymbol{P} \bar{\boldsymbol{x}} + \bar{\boldsymbol{n}}_k, \; k = 1, 2, \ldots, K. \tag{7.56}$$

Let us now collect all the K received signal vectors $\{\bar{\boldsymbol{y}}_k\}$ of (7.53) into a vector $\bar{\boldsymbol{y}}$, which is expressed

as

$$\bar{\boldsymbol{y}} = \left[\bar{\boldsymbol{y}}_1^T, \bar{\boldsymbol{y}}_2^T, \ldots, \bar{\boldsymbol{y}}_K^T\right]^T. \tag{7.57}$$

Then, according to (7.56) it can be shown that the overall received signal vector $\bar{\boldsymbol{y}}$ of all K MSs can be expressed as

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{U}}\bar{\boldsymbol{\Lambda}}^{1/2}\bar{\boldsymbol{V}}_s^H \boldsymbol{P}\bar{\boldsymbol{x}} + \bar{\boldsymbol{n}}, \tag{7.58}$$

where we introduced the following definitions,

$$\begin{aligned}
\bar{\boldsymbol{U}} &= \text{diag}\left\{\bar{\boldsymbol{U}}_1, \bar{\boldsymbol{U}}_2, \cdots, \bar{\boldsymbol{U}}_K\right\}, \\
\bar{\boldsymbol{\Lambda}} &= \text{diag}\left\{\bar{\boldsymbol{\Lambda}}_1, \bar{\boldsymbol{\Lambda}}_2, \ldots, \bar{\boldsymbol{\Lambda}}_K\right\}, \\
\bar{\boldsymbol{V}}_s &= \left[\bar{\boldsymbol{V}}_{1s}, \bar{\boldsymbol{V}}_{2s}, \cdots, \bar{\boldsymbol{V}}_{Ks}\right], \\
\bar{\boldsymbol{n}} &= \left[\bar{\boldsymbol{n}}_1^T, \bar{\boldsymbol{n}}_2^T, \cdots, \bar{\boldsymbol{n}}_K^T\right]^T.
\end{aligned} \tag{7.59}$$

In (7.59) $\bar{\boldsymbol{U}}$ and $\bar{\boldsymbol{\Lambda}}$ are $\left(\sum_{k=1}^K N_k \times \sum_{k=1}^K N_k\right)$-component matrices, $\bar{\boldsymbol{V}}_s$ is a $\left(M \times \sum_{k=1}^K N_k\right)$-component matrix and $\bar{\boldsymbol{n}}$ is an AWGN vector, which is Gaussian distributed with zero-mean and a covariance matrix of $\sigma^2 \boldsymbol{I}_{\sum_{k=1}^K N_k}$.

The preprocessing matrix $\boldsymbol{P}$ is designed, so that the downlink MUI can be efficiently suppressed. As shown in (7.58), the MUI can be fully removed, when the preprocessing matrix $\boldsymbol{P}$ is chosen to satisfy

$$\bar{\boldsymbol{V}}_s^H \boldsymbol{P} = \bar{\boldsymbol{\beta}}, \tag{7.60}$$

where the power-allocation regime of $\bar{\boldsymbol{\beta}} = \text{diag}\left\{\bar{\beta}_1, \bar{\beta}_2, \cdots, \bar{\beta}_{\sum_{k=1}^K N_k}\right\} = \text{diag}\left\{\bar{\beta}_{11}, \cdots, \bar{\beta}_{1N_1}; \cdots; \bar{\beta}_{K1}, \cdots, \bar{\beta}_{KN_K}\right\}$ represents our transmission power constraint, which will be considered in detail in our forthcoming discourse of Subsection 7.3.1.

In order to satisfy (7.60), $\boldsymbol{P}$ can be set to be

$$\begin{aligned}
\boldsymbol{P} &= \left[\bar{\boldsymbol{V}}_s^H\right]^+ \bar{\boldsymbol{\beta}} \\
&= \bar{\boldsymbol{V}}_s \left[\bar{\boldsymbol{V}}_s^H \bar{\boldsymbol{V}}_s\right]^{-1} \bar{\boldsymbol{\beta}} \\
&= \tilde{\boldsymbol{P}}\bar{\boldsymbol{\beta}},
\end{aligned} \tag{7.61}$$

where $\left[V_s^H\right]^+$ denotes the pseudo inverse of the matrix V_s^H and $\tilde{P} = \left[V_s^H\right]^+ = \bar{V}_s \left[\bar{V}_s^H \bar{V}_s\right]^{-1}$.

When substituting the overall preprocessing matrix of (7.60) into (7.58), the overall received signal vector $\bar{y}$ of all K MSs can be simplified to

$$\bar{y} = \bar{U}\bar{\Lambda}^{1/2}\bar{\beta}\bar{x} + \bar{n}. \tag{7.62}$$

To be more specific, the N_k-length observation vector of the kth MS can be expressed as

$$\bar{y}_k = \bar{U}_k\bar{\Lambda}_k^{1/2}\bar{\beta}_k\bar{x}_k + \bar{n}_k,\ k = 1, 2, \ldots, K, \tag{7.63}$$

where we have $\bar{\beta}_k = \text{diag}\left\{\bar{\beta}_{k1}, \bar{\beta}_{k2}, \cdots, \bar{\beta}_{kN_k}\right\}$. Explicitly, the kth user endures no interference conflicted by the other users. However, there may exist interference among the symbols transmitted by the BS to the kth MS. This inter-element interference can be suppressed with the aid of the SVD-based matrices $\left\{\bar{U}_k\right\}$. Consequently, after the postprocessing the received signal vectors $\{\bar{y}_k\}$ by $\left\{G_k = \bar{U}_k^H\right\}$ according to Figure 7.2, the decision variables can be individually as

$$\hat{\bar{x}}_k = \bar{\Lambda}_k^{1/2}\bar{\beta}_k\bar{x}_k + \bar{U}_k^H\bar{n}_k,\ k = 1, 2, \ldots, K \tag{7.64}$$

or jointly as

$$\begin{aligned}\hat{\bar{x}} &= [\hat{\bar{x}}_1^T, \cdots, \hat{\bar{x}}_K^T]^T \\ &= \bar{\Lambda}^{1/2}\bar{\beta}\bar{x} + \bar{U}^H\bar{n}.\end{aligned} \tag{7.65}$$

Let us now consider the design of power allocation for the downlink.

7.3.1 Power-Allocation

In the context of downlink transmissions, the total transmission power $E\left[\|\boldsymbol{P}\bar{\boldsymbol{x}}\|^2\right]$ of all the K MSs after transmit preprocessing must not exceed the original total transmission power $E\left[\|\boldsymbol{x}\|^2\right]$ of all the K MSs before transmit preprocessing, hence the power-allocation designed for downlink transmissions is implemented under the constraint of [153]

$$E\left[\|\boldsymbol{P}\bar{\boldsymbol{x}}\|^2\right] \leq E\left[\|\boldsymbol{x}\|^2\right]. \tag{7.66}$$

Upon substituting (7.61) into (7.66), we arrive at [153]

$$\text{trace}(\tilde{\boldsymbol{P}}^H \tilde{\boldsymbol{P}} \bar{\boldsymbol{\beta}}^2) \leq \sum_{k=1}^{K} N_k. \tag{7.67}$$

Here we have assumed that $E[\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H] = I_{\sum_{k=1}^{K} N_k}$.

Explicitly, $\tilde{\boldsymbol{P}}^H \tilde{\boldsymbol{P}}$ is a positive definite full-rank matrix. Let $\{\vartheta_{ii}\}$ $(1 \leq i \leq \sum_{k=1}^{K} N_k)$ be the diagonal elements of matrix $\tilde{\boldsymbol{P}}^H \tilde{\boldsymbol{P}}$, where we have $\{\vartheta_{ii}\} > 0$. Consequently, the constraint of (7.67) can be written as

$$\sum_{i=1}^{\sum_{k=1}^{K} N_k} \vartheta_{ii} \bar{\beta}_{ii}^2 \leq \sum_{k=1}^{K} N_k, \tag{7.68}$$

and furthermore, we have

$$\begin{aligned} \boldsymbol{P} &= \tilde{\boldsymbol{P}} \bar{\boldsymbol{\beta}} \\ &= [\tilde{\boldsymbol{P}}_1 \bar{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{P}}_K \bar{\boldsymbol{\beta}}_K] \\ &= [\boldsymbol{P}_1, \cdots, \boldsymbol{P}_K], \end{aligned} \tag{7.69}$$

where $\boldsymbol{P}_k = \tilde{\boldsymbol{P}}_k \bar{\boldsymbol{\beta}}_k$.

Alternatively, we can also impose a power constraint on each user as

$$E\left[\|\boldsymbol{P}_k \bar{\boldsymbol{x}}_k\|^2\right] \leq E\left[\|\bar{\boldsymbol{x}}_k\|^2\right]. \tag{7.70}$$

Hence we have

$$\text{trace}(\tilde{\boldsymbol{P}}_k^H \tilde{\boldsymbol{P}}_k \bar{\boldsymbol{\beta}}_k^2) \leq N_k. \tag{7.71}$$

Let $\{\vartheta_{ii}^{(k)}\}$ $(1 \leq i \leq N_k)$ be the diagonal elements of matrix $\tilde{\boldsymbol{P}}_k^H \tilde{\boldsymbol{P}}_k$, where we have $\{\vartheta_{ii}^{(k)}\} > 0$. Consequently, the constraint of (7.70) can be written as

$$\sum_{i=1}^{N_k} \vartheta_{ii}^{(k)} \bar{\beta}_{ki}^2 \leq N_k. \tag{7.72}$$

In this section four different types of power-allocation schemes are considered. The first one maximizes the overall information rate of the K downlink users. We refer to this power-allocation scheme

as that designed for achieving joint maximum information rate. The second one aims for maximizing the information rate for each of the downlink users. We term this power-allocation scheme as the one designed for maintaining the individual maximum information rate. Moreover, similarly to the uplink, the downlink power may be optimized for maximizing the average SNR for the individual users, which is hence referred to as the maximum SNR (MSNR) assisted power-allocation regime. Finally, a widely used power-allocation scheme, which is referred to as equal power-allocation scheme in this chapter, since it simply assigns the same power to each antenna of each MS, is also investigated in the context of the proposed transmission regime.

7.3.1.1 Joint Maximum Information Rate Assisted Power-Allocation

When following the approach of Section 7.2.4.1 and using the 'water-filling' principle [64], it can be shown that the overall information rate of the downlink can be maximized, if $\{\bar{\beta}_{ii}\}$ are chosen as [64]

$$\bar{\beta}_{ii}^2 = \left(v - \frac{1}{\bar{\lambda}_{ii}}\right)^+, \; i = 1, 2, \ldots, \sum_{k=1}^{K} N_k, \tag{7.73}$$

where v is chosen so that the power allocation constraint of (7.68) is satisfied [64], i.e. we have

$$\sum_{i=1}^{\sum_{k=1}^{K} N_k} \vartheta_{ii} \bar{\beta}_{ii}^2 \leq \sum_{k=1}^{K} N_k. \tag{7.74}$$

The proof of this is given in Appendix B. Furthermore, it can be shown that the maximum overall information rate of the downlink can be computed as [64]

$$I_{\max}(\hat{\bar{\boldsymbol{x}}}, \bar{\boldsymbol{x}}) = \sum_{i=1}^{\sum_{k=1}^{k} N_k} \left(\log_2 \left[v\bar{\lambda}_i\right]\right)^+ \tag{7.75}$$

which represents the maximum number of error free information bits per channel use for downlink transmission, when jointly considering the power allocation of all the K MSs, and when the noise power is unity.

Additionally, given a specific SNR, the maximum overall information rate normalized by the total number of receive antennas $\sum_{k=1}^{K} N_k$ can be expressed as [147]

$$\mathcal{I}_{\max}(\hat{\bar{\boldsymbol{x}}}, \bar{\boldsymbol{x}}) = \frac{1}{\sum_{k=1}^{K} N_k} \sum_{i=1}^{\sum_{k=1}^{K} N_k} \log_2 \left[\left(1 + \frac{\bar{\lambda}_i}{\sigma^2}\left(v - \frac{\sigma^2}{\bar{\lambda}_i}\right)^+\right)\right], \tag{7.76}$$

which represents the normalized maximum number of error free information bits per channel use for a given SNR, when jointly considering the power allocation among all the K MSs, and when ZF-assisted transmit preprocessing based on SVD is applied, assuming that the noise samples observed at the MSs are independent.

7.3.1.2 Power-Allocation Designed for Maximum Individual Information Rate

When the power-allocation regime based on the maximum individual information rate is considered, the power adjustment is carried out for a specific user without exploiting any information concerning the other users. Specifically, for the kth downlink MS the power is specifically allocated so that the kth user's information rate is maximized, while satisfying the power constraint of

$$\sum_{i=1}^{N_k} \vartheta_{ii}^{(k)} \bar{\beta}_{ki}^2 \leq N_k. \tag{7.77}$$

When we compare (7.64) to (7.30), we find that they have the same structure. Hence, in this case the power $\bar{\beta}_{ki}^2$ allocated for the kth downlink MS can be represented as [64]

$$\bar{\beta}_{ki}^2 = \left(v_k - \frac{1}{\bar{\lambda}_{ki}}\right)^+, \; i = 1, 2, \ldots, N_k; \; k = 1, 2, \ldots, K, \tag{7.78}$$

where v_k is chosen so that the the transmission power constraint of (7.77) is satisfied [64], i.e. we have

$$\sum_{i=1}^{N_k} \vartheta_{ii}^{(k)} \bar{\beta}_{ki}^2 = \sum_{i=1}^{N_k} \vartheta_{ii}^{(k)} \left(v_k - \frac{1}{\bar{\lambda}_{ki}}\right)^+ \leq N_k, \; k = 1, 2, \ldots, K. \tag{7.79}$$

The proof of this can be found in Appendix B.

Furthermore, it can be shown that the maximum information rate of the kth MS can be expressed as [64]

$$I_{\max}(\hat{\tilde{\boldsymbol{x}}}_k, \bar{\boldsymbol{x}}_k) = \sum_{i=1}^{N_k} \left(\log_2 \left[v_k \bar{\lambda}_{ki}\right]\right)^+, \; k = 1, 2, \ldots, K \tag{7.80}$$

which represents the maximum number of error free information bits per channel use for the kth MS, when noise power is unity. Alternatively, when the SNR is known, the maximum information rate of

the kth MS normalized by the number of receive antennas N_k can be expressed as [147]

$$\mathcal{I}_{\max}(\hat{\bar{\boldsymbol{x}}}_k, \bar{\boldsymbol{x}}_k) = \frac{1}{N_k} \sum_{i=1}^{N_k} \log_2 \left[\left(1 + \frac{\bar{\lambda}_{ki}}{\sigma^2} \left(\upsilon_k - \frac{\sigma^2}{\bar{\lambda}_{ki}} \right)^+ \right) \right], \tag{7.81}$$

which represents the average maximum number of error free information bits per channel use for the kth MS for a given SNR, again assuming that the noise samples observed at the kth MS are independent.

7.3.1.3 Power-Allocation Designed for Maximum SNR

Similarly to the uplink scenario of (7.41), when the power-allocation is optimized for maintaining the MSNR, the transmitter minimizes

$$\xi = \sum_{i=1}^{\sum_{k=1}^{K} N_k} \frac{1}{\gamma_i} = \sum_{i=1}^{\sum_{k=1}^{K} N_k} \frac{\sigma^2}{\bar{\lambda}_{ii} \bar{\beta}_{ii}^2} \tag{7.82}$$

under the constraint of (7.68) imposed on the transmission power.

Upon following an approach used for the uplink in Section 7.2.4.2, it can be shown that in the case of MSNR-optimization the power $\bar{\beta}_{ii}^2$ allocated to the ith data stream is given by [88]

$$\begin{aligned} \bar{\beta}_{ii}^2 &= \sum_{k=1}^{K} N_k \left(\sum_{j=1}^{\sum_{k=1}^{K} N_k} \sqrt{\frac{\sigma^2 \vartheta_{jj}}{\bar{\lambda}_{jj}}} \right)^{-1} \sqrt{\frac{\sigma^2}{\bar{\lambda}_{ii} \vartheta_{ii}}} \\ &= \sum_{k=1}^{K} N_k \left(\sum_{j=1}^{\sum_{k=1}^{K} N_k} \sqrt{\frac{\vartheta_{jj}}{\bar{\lambda}_{jj}}} \right)^{-1} \frac{1}{\sqrt{\bar{\lambda}_{ii} \vartheta_{ii}}}, \; i = 1, 2, \ldots, \sum_{k=1}^{K} N_k. \end{aligned} \tag{7.83}$$

Furthermore, similarly to the uplink scenario of (7.40), the resultant SNR of the ith data stream can be expressed as

$$\gamma_i = \frac{\sum_{k=1}^{K} N_k}{\sigma^2} \left(\sum_{j=1}^{\sum_{k=1}^{K} N_k} \sqrt{\frac{\vartheta_{jj}}{\bar{\lambda}_{jj}}} \right)^{-1} \sqrt{\frac{\bar{\lambda}_{ii}}{\vartheta_{ii}}}, \; i = 1, 2, \ldots, \sum_{k=1}^{K} N_k. \tag{7.84}$$

Furthermore, when BPSK modulation is employed, the BER can be computed as

$$P_b(\gamma_i) = Q\left(\sqrt{2\gamma_i}\right), \ i = 1, 2, \ldots, \sum_{k=1}^{K} N_k. \tag{7.85}$$

7.3.1.4 Equal Power Allocation

In this case, the coefficients $\bar{\beta}_i$ are set according to $\bar{\beta}_1 = \cdots = \bar{\beta}_{\sum_{k=1}^{K} N_K} = \bar{\beta}$, where $\bar{\beta}$ is a constant, facilitating the normalization of all K MS's power [151, 153], hence (7.67) can be expressed as [153]

$$\bar{\beta}^2 \text{trace}(\tilde{\boldsymbol{P}}^H \tilde{\boldsymbol{P}}) \leq \sum_{k=1}^{K} N_k. \tag{7.86}$$

It can be readily shown with the aid of (7.61) that in this case we have [151, 153]

$$\bar{\beta} = \sqrt{\frac{\sum_{k=1}^{K} N_k}{\text{trace}\left(\left[\bar{\boldsymbol{V}}_s^H \bar{\boldsymbol{V}}_s\right]^{-1}\right)}}. \tag{7.87}$$

In this case, the resulant SNR of (7.40) for the ith symbol of MS k can be expressed as

$$\gamma_i = \frac{\bar{\lambda}_{ii}}{\sigma^2} \frac{\sum_{k=1}^{K} N_k}{\text{trace}\left(\left[\bar{\boldsymbol{V}}_s^H \bar{\boldsymbol{V}}_s\right]^{-1}\right)}. \tag{7.88}$$

Specifically, we refer to this power-allocation scheme as the equal-power based scheme.

Furthermore, when BPSK modulation is employed, the resultant BER can be computed as

$$P_b(\gamma_i) = Q\left(\sqrt{2\gamma_i}\right), \ i = 1, 2, \ldots, \sum_{k=1}^{K} N_k. \tag{7.89}$$

Similarly to the uplink discussions of Subsection 7.2.4, there are also other strategies for downlink power allocation [126, 168, 187].

7.4 Downlink Transmitter Preprocessing and Detection in TDD MIMO Systems

The downlink transmitter preprocessing and detection techniques discussed in Section 7.3 are suitable for both FDD and TDD systems. When TDD-assisted MIMO systems are considered, both the downlink transmitter preprocessing and detection schemes can be further simplified by exploiting the channel's reciprocity manifesting itself in terms of a similar UL and DL CHTF, since both are transmitted at the same frequency, although naturally, in practice the effects of the different UL and DL interferences also have to be taken in account. Let us consider the benefits of UL/DL reciprocity in a little more detail.

In TDD-assisted MIMO systems, instead of transmitting the data as seen in (7.50), we arrange for the downlink to transmit

$$\bar{\boldsymbol{d}} = \sum_{k=1}^{K} \boldsymbol{P}_k \bar{\boldsymbol{x}}_k^* = \boldsymbol{P}\bar{\boldsymbol{x}}^* \tag{7.90}$$

where $\bar{\boldsymbol{x}}_k^*$ is the conjugate of $\bar{\boldsymbol{x}}_k$ as in [130].

Since in TDD-assisted MIMO systems we have $\bar{\boldsymbol{H}}_k = \boldsymbol{H}_k^T$, where $\boldsymbol{H}_k$ is the CIR matrix of the uplink channel, the N_k-component received vector $\boldsymbol{r}_k$ by the kth MS can be expressed as

$$\begin{aligned} \boldsymbol{r}_k &= \boldsymbol{H}_k^T \bar{\boldsymbol{d}} + \bar{\boldsymbol{n}}_k \\ &= \boldsymbol{H}_k^T \sum_{l=1}^{K} \boldsymbol{P}_l \bar{\boldsymbol{x}}_l^* + \bar{\boldsymbol{n}}_k \\ &= \boldsymbol{H}_k^T \boldsymbol{P}_k \bar{\boldsymbol{x}}_k^* + \sum_{l \neq k}^{K} \boldsymbol{H}_k^T \boldsymbol{P}_l \bar{\boldsymbol{x}}_l^* + \bar{\boldsymbol{n}}_k. \end{aligned} \tag{7.91}$$

Upon applying the conjugation operation to $\boldsymbol{r}_k$ [130], which allows us to recover the DL's transmitted data, we arrive at

$$\begin{aligned} \boldsymbol{r}_k^* &= \boldsymbol{H}_k^H \boldsymbol{P}^* \bar{\boldsymbol{x}} + \bar{\boldsymbol{n}}_k^* \\ &= \boldsymbol{H}_k^H \boldsymbol{P}_k^* \bar{\boldsymbol{x}}_k + \sum_{l \neq k}^{K} \boldsymbol{H}_k^H \boldsymbol{P}_l^* \bar{\boldsymbol{x}}_l + \bar{\boldsymbol{n}}_k^*. \end{aligned} \tag{7.92}$$

According to (7.17), the CIR matrix $\boldsymbol{H}_k$ of the uplink channel associated with the kth MS can be

represented with the aid of the SVD as

$$\boldsymbol{H}_k = \boldsymbol{U}_{ks}\boldsymbol{\Lambda}_k^{1/2}\boldsymbol{V}_k^H, \; k = 1, 2, \ldots, K. \tag{7.93}$$

Upon substituting (7.93) into (7.92), we arrive at

$$\boldsymbol{r}_k^* = \boldsymbol{V}_k\boldsymbol{\Lambda}_k^{1/2}\boldsymbol{U}_{ks}^H\boldsymbol{P}^*\bar{\boldsymbol{x}}_k + \bar{\boldsymbol{n}}_k^*. \tag{7.94}$$

Let $\boldsymbol{r}^*$ be the $\left(\sum_{k=1}^{K} N_k\right)$-component vector defined as

$$\boldsymbol{r}^* = [\boldsymbol{r}_1^H, \boldsymbol{r}_2^H, \cdots, \boldsymbol{r}_K^H]^T, \tag{7.95}$$

Then, it can be shown with the aid of (7.94) that $\boldsymbol{r}^*$ can be represented as

$$\boldsymbol{r}^* = \boldsymbol{V}_s\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}_s^H\boldsymbol{P}^*\bar{\boldsymbol{x}} + \bar{\boldsymbol{n}}^*, \tag{7.96}$$

where the arguments of $\boldsymbol{\Lambda}$ and $\boldsymbol{U}_s$ have been defined in the context of (7.21), while $\boldsymbol{V}_s$ and $\bar{\boldsymbol{n}}$ are given by

$$\begin{aligned} \boldsymbol{V}_s &= \text{diag}\,\{\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_K\} \\ \bar{\boldsymbol{n}} &= \left[\bar{\boldsymbol{n}}_1^T, \bar{\boldsymbol{n}}_2^T, \cdots, \bar{\boldsymbol{n}}_K^T\right]^T. \end{aligned} \tag{7.97}$$

Note that in the above equations the orthogonal matrices $\boldsymbol{U}_{ks}$ and $\boldsymbol{V}_k$ for $k = 1, 2, \ldots, K$ have been obtained for uplink transmission and detection, respectively. Therefore, in TDD-assisted MIMO systems the preprocessing and post-processing techniques designed for both the uplink and downlink can be implemented based on the same common SVD at each of the MSs for finding $\boldsymbol{V}_k$ for $k = 1, \ldots, K$, while K common SVDs are employed at the BS for deriving $\boldsymbol{U}_{ks}$ for $k = 1, \ldots, K$. By contrast, when the the FDCHTFs of the uplink and downlink cannot be deemed similar, which is often the case in FDD-assisted wireless systems, each MS requires two separate SVDs, one for uplink transmission and one for downlink detection. Similarly, the BS has to carry out $2K$ SVDs, K of them for uplink detection and the other K for downlink preprocessing.

Based on (7.96) it can be readily shown that in order to completely remove the multiuser interfer-

ence, the BS's transmitter preprocessing matrix P can be chosen to satisfy

$$U_s^H P^* = \bar{\beta}. \tag{7.98}$$

Upon solving this equation, we arrive at

$$P = U_s^* \left[U_s^T U_s^*\right]^{-1} \bar{\beta}. \tag{7.99}$$

Finally, the decision variables can be derived for the K MSs by substituting (7.99) into (7.96) and then multiplying both sides of (7.96) by V_s^H, yielding,

$$\begin{aligned} y &= V_s^H r^* \\ &= \Lambda^{1/2} \bar{\beta} \bar{x} + V_s^H \bar{n}^*, \end{aligned} \tag{7.100}$$

where $\bar{\beta}$ depends on the specific power-allocation scheme applied, as discussed in Section 7.3.

7.5 Simulations

In this section, simulation results are provided for characterizing the achievable performance of the proposed algorithm in conjunction with a specific power allocation scheme. Speficically, three different scenarios are considered in Tables 7.1, 7.2 and 7.3, where Table 7.1 outlines the parameters used for various fully loaded systems, where the MSs benefit from having $2, 4, 5$ or 10 antennas. By contrast, in Table 7.2 the total number of antennas employed was $M = 20$ and at the MS it was $N_k = 4$, while the number of MSs was $K = 2, 3, 4$ or 5. Finally, in Table 7.3 the number of MSs was fixed to $K = 3$ and the number of antennas at each MS was set to $N_k = 4$, while the number of antennas at the BS was $M = 12, 16, 20$ or 24. Furthermore, the symbols "$M - K - N_k$" in the legends of the simulation results seen in Figures 7.3 to 7.17 indicate that M antennas are employed at the BS for supporting K MSs and each of the MSs had N_k antennas.

As a benchmarker, the achievable system capacity versus average SNR per symbol recorded for a single-user MIMO system is evaluated from (7.36) and plotted in Figure 7.3, when uncorrlated noise contaminates the received signal of the four schemes. The number of antennas at the BS is fixed to $M = 20$ and the number of antennas at MS varies from $N_k = 1$ to 4. We can see from Figure 7.3 that the achievable system capacity increases significantly upon increasing the number of receive antennas at the MS. This is because the more antennas the system has, the higher the capacity that it is capable

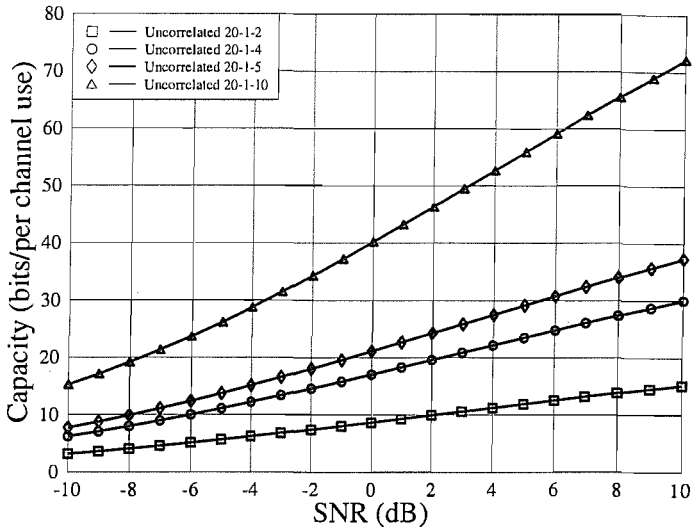


Figure 7.3: System capacity versus average SNR per symbol evaluated from (7.36) for a single user MIMO system when encountering the uncorrelated noise for four schemes where number of antennas at the BS is fixed to $M = 20$ and the number of antennas at MS varies from $N_k = 1$ to 4.

	Scheme 1	Scheme 2	Scheme 3	Scheme 4
Number of MSs (K)	10	5	4	2
Number of antennas at each MS (N_k)	2	4	5	10
Number of antennas at the BS (M)	20	20	20	20
Modulation scheme	BPSK	BPSK	BPSK	BPSK

Table 7.1: Parameters I for SDMA transmission based on SVD for various fully loaded systems, where the MSs benefit from having $2, 4, 5$ or 10 antennas.

of achieving [64]. However, as we can see in Figure 7.3, the attainable capacity does not increase proportionally with the number of transmitter antennas [64].

In Figure 7.4, the attainable average capacity versus average SNR per symbol recorded for uplink transmission is plotted, when invoking the maximum information rate based power allocation policy of [64, 185] for the four diffent transmission schemes, as shown in Table 7.1. We can see from Figure 7.4 that for a specific scheme of Table 7.1, the capacity achieved by assuming the presence of the uncorrelated noise of (7.36) is significantly higher than that associated with encountering the correlated noise of (7.38). This is due to the noise enhancement imposed by the postprocessing. Furthermore, the highest capacity is achieved by Schemes 1, 2, 3 and 4, when the presence of uncorrelated

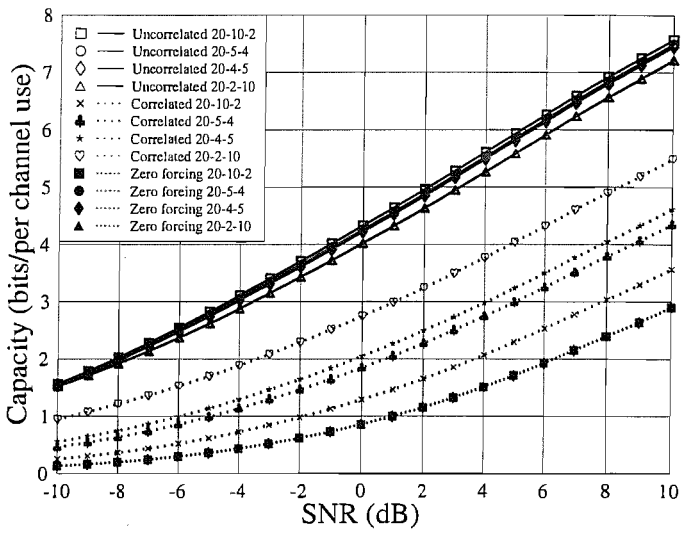


Figure 7.4: The average capacity versus average SNR per symbol evaluted from (7.36), (7.38) and (7.39) for uplink transmission, when invoking the maximum information rate based power allocation policy for the four diffent transmission schemes outlined in Table 7.1.

	Scheme 1	Scheme 2	Scheme 3	Scheme 4
Number of MSs (K)	2	3	4	5
Number of antenna at each MS (N_k)	4	4	4	4
Number of antennas at the BS (M)	20	20	20	20
Modulation scheme	BPSK	BPSK	BPSK	BPSK

Table 7.2: Parameters II for SDMA transmission based on SVD for different number of MSs, i.e. system loads.

noise is assumed. This can be explained that the system capacity does not increase proportionally with the number of transmitter antennas as evidenced in Figure 7.3. Moreover, when considering the correlated noise scenario of Figure 7.4, the more users the system supports, the lower the capacity it achieves. This is because for a fully loaded system a higher number of users results in lower number of antennas at each MS, resulting in a higher MUI for each individual user. Furthermore, the classic zero forcing receiver treats all other $(\sum_{k=1}^{K} N_k - 1)$ transmit antennas' signal as interference, which results in the lowest performance.

In Figure 7.5, the achievable average capacity versus average SNR per symbol performance is plotted for uplink transmission, when invoking the maximum information rate based power allocation

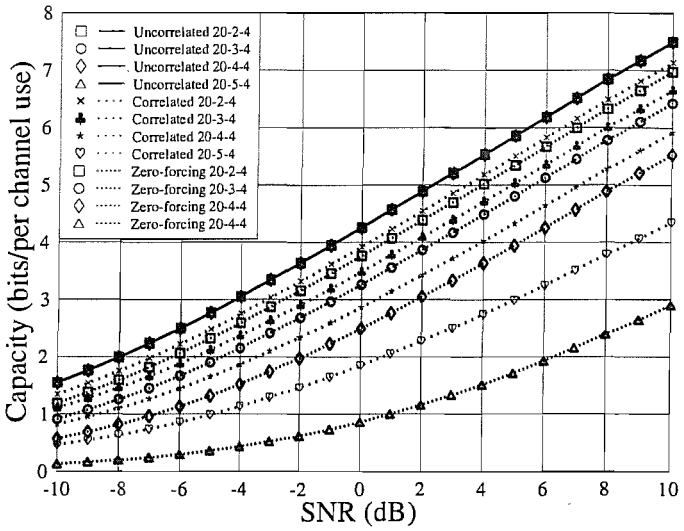


Figure 7.5: The average capacity versus average SNR per symbol evaluted from (7.36), (7.38) and (7.39) for uplink transmission, when invoking maximum information rate based power allocation policy for the four different schemes, as shown in Table 7.2.

policy of [64, 185] for the four different schemes outlined in Table 7.2. We can see from Figure 7.5 that the achievable capacity remains constant for the four schemes, when the presence of uncorrelated noise is assumed, because the system capacity improvement attained is proportional to the number of users supported in this scenario. However, the achievable capacity recorded in the presence of correlated noise decreases upon increasing the number of users, which results in an increased MUI. Furthermore, for uncorrelated noise, the system capacity increases proportionally upon increasing the number of users, since the average capacity is the same for any of the schemes in Table 7.2. However, for correlated noise, the situation is more complicated and will be further detailed in the context of Figure 7.6.

In Figure 7.6, the achievable system capacity versus the average SNR per symbol is plotted for uplink transmission, when invoking the maximum information rate based power allocation policy for correlated noise and using the four different schemes of Table 7.2. We can see from Figure 7.6 that the highest system capacity may be achieved by different schemes for different SNRs. In fact, there is a tradeoff between the number of users supported and the SNR. The more users are served, the more serious the MUI. However the effects of the MUI are more readily mitigated upon increasing the SNR,

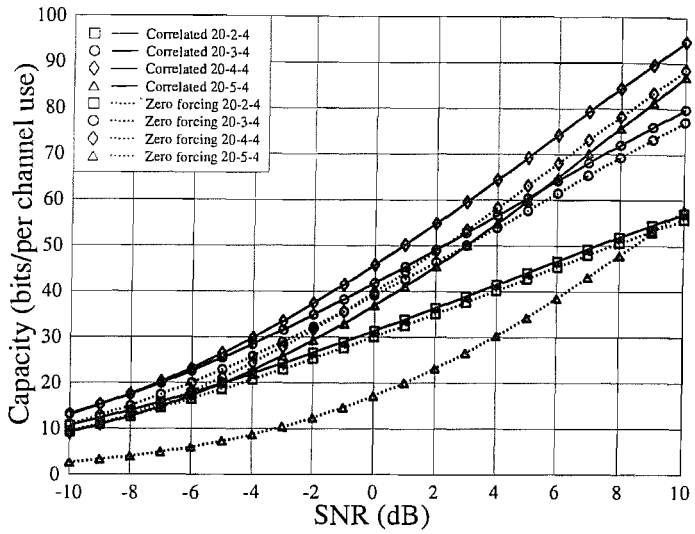


Figure 7.6: System capacity versus average SNR per symbol evaluted from (7.38) and (7.39) for uplink transmission when invoking maximum information rate based power allocation policy of [64, 185] for only correlated noise for the four different schemes of Table 7.2.

	Scheme 1	Scheme 2	Scheme 3	Scheme 4
Number of users	3	3	3	3
Number of antenna at each user	4	4	4	4
Number of antennas at the BS	12	16	20	24
Modulation scheme	BPSK	BPSK	BPSK	BPSK

Table 7.3: Parameters III for SDMA transmission based on SVD for a fixed number of 3 users and 4 antennas at each MS, while varying the number of BS antennas.

hence, the more users, the higher the capacity at high SNRs. Observe in Figure 7.6 that as expected, the highest capacity should be achieved in the 5-user scenario of Table 7.2, when the SNR approaches infinity, because no residual MUI is experienced. Furthermore, the traditioanl zero-forcing receiver achieves the lowest capacity, as discussed above.

In Figure 7.7, the achievable average capacity versus average SNR per symbol is plotted for uplink transmission, when invoking the maximum information rate based power allocation policy of [64,185] for the four different schemes of Table 7.3. We can see from Figure 7.7 that the attainable capacity corresponding to the uncorrelated noise scenario for a specific scheme is always better than that of the correlated noise scenario due to the noise enhancement experienced in the uncorrelated noise

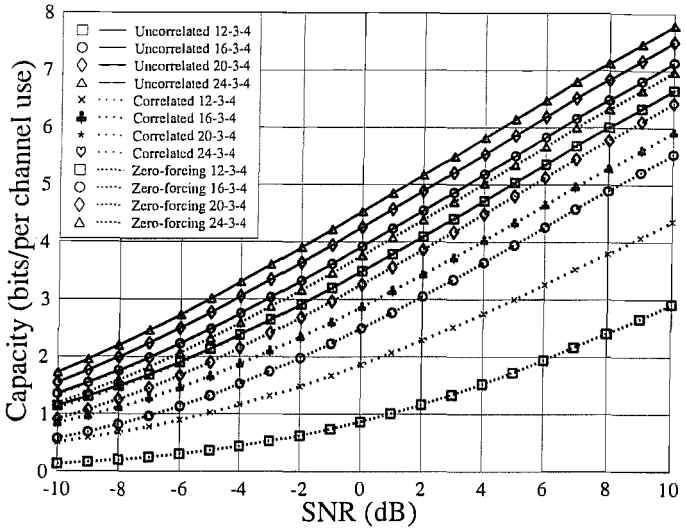


Figure 7.7: Average capacity versus average SNR per symbol evaluted from (7.38) and (7.39) for uplink transmission when invoking maximum information rate based power allocation policy for the four different schemes of Table 7.3.

scenario. Furthermore, the achievable capaticy increases upon increasing the number of antennas at the BS for both the uncorrelated and correlated noise scenarios due to associated increased receiver diversity gain. Moreover, the achievable capacity corresponding to $M = 20$ and 24 antennas at the BS in the presence of correlated noise is almost the same as that corresponding to $M = 12$ and 16 antennas at the BS recorded for uncorrelated noise. The reason for this is because the associated receive diversity gain compensates the performance loss due to noise enhancement. Unsurpersingly, the classic zero-forcing receiver still achieves the lowest capacity.

In Figure 7.8 the achievable average BER versus average SNR per symbol is plotted for uplink transmission, when invoking the maximum SNR based power allocation policy for the four different schemes of Table 7.1. We can see from Figure 7.8 that the BER performance recorded for uncorrelated noise is always better than that corresponding to the correlated noise scenario, which again is due to the more substantial noise enhancement experienced in the presence of correlated noise. Furthermore, the BER performance degrades upon increasing the number of transmit antennas at the MS, when uncorrelated noise is encountered. The reason for this is because having more transmit antennas at MS may result in a wider eigenvalue dynamic range. Consequently, low eigenvalues may be

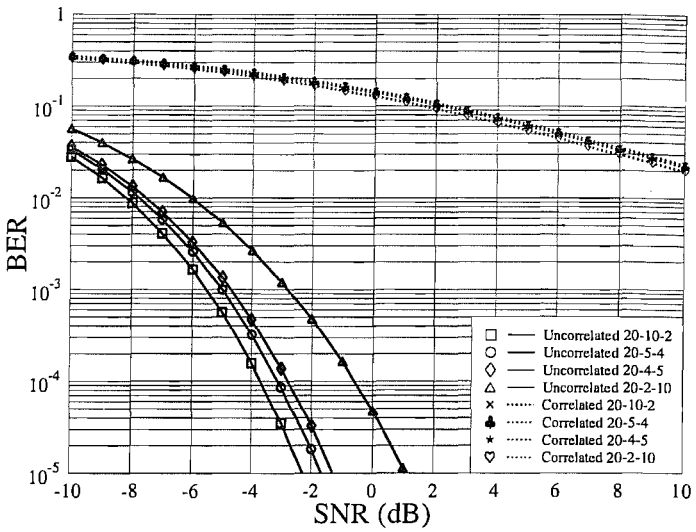


Figure 7.8: Average BER versus average SNR per symbol evaluted from (7.47) for uplink transmission when invoking the maximum SNR based power allocation policy for the four different schemes of Table 7.1.

obtained, which dominate the BER performance. However, in the presence of correlated noise the BER performance is similar for the schemes considered in Table 7.1. The reason for this is that when we have less antennas at the MS, we support more users, which results in more serious MUI.

In Figure 7.9, displays the achievable average BER versus average SNR per symbol performance for uplink transmissions, when invoking the maximum SNR based power allocation policy for the four different schemes of Table 7.2. We can see from Figure 7.9 that the BER performance remains similar upon increasing the number of users, when uncorrelated noise is assumed, since (7.46) shows that the SNR of a specific user is not related to the parameters of other users. However, in the uncorrelated noise scenario, the BER performance degrades upon increasing the number of users due to the more serious MUI inflicted.

In Figure 7.10, the achievable average BER versus average SNR per symbol is characterized in the context of uplink transmissions, when invoking the maximum SNR based power allocation policy for the four different schemes of Table 7.3. We can see from Figure 7.9 that the BER performance improves upon increasing the number of antennas at the BS for both the uncorrelated and correlated noise scenarios due to the associated receiver diversity gain. Furthermore, for a given scheme of

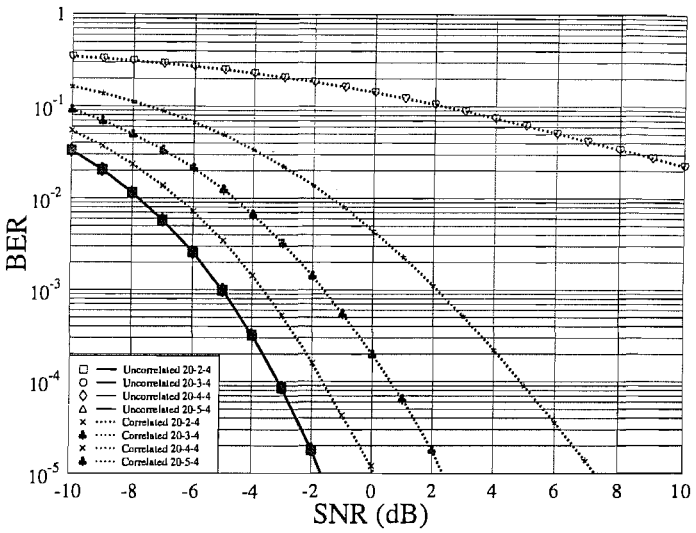


Figure 7.9: Average BER versus average SNR per symbol evaluted from (7.47) for uplink transmission when invoking the maximum SNR based power allocation policy for the four different schemes of Table 7.2.

Table 7.3 the BER performance recorded in the presence of correlated noise is worse than that of the uncorrelated noise scenario owing to the more substantial noise enhancement. Moreover, the achievable BER performance corresponding to $M = 20$ and 24 antennas at the BS and recorded for correlated noise is close to that corresponding to $M = 12$ and 16 antennas experienced in the presence of uncorrelated noise, since the increased receive diversity gain compensates for the performance loss imposed by the noise enhancement.

In Figure 7.11, the attainable average capacity versus average SNR per symbol recorded for downlink transmissions is plotted when invoking joint and individual maximum information rate based power allocation policy of [64], respectively, for the four diffent transmission schemes of Table 7.1. Surprisingly, we observe in Figure 7.11 that the average capacity achieved by the individual maximum information rate policy of [64] is slightly higher than that recorded for the joint maximum information rate policy [64]. The reason for this is because the power constraint of (7.74) is more stringent owing to the effect of ϑ_{ii} than (7.79). Furthermore the average capacity decreases upon increasing the number of users, since supporting more users results in a higher MUI.

Figure 7.12 portrays the achievable average capacity versus the average SNR per symbol perfor-

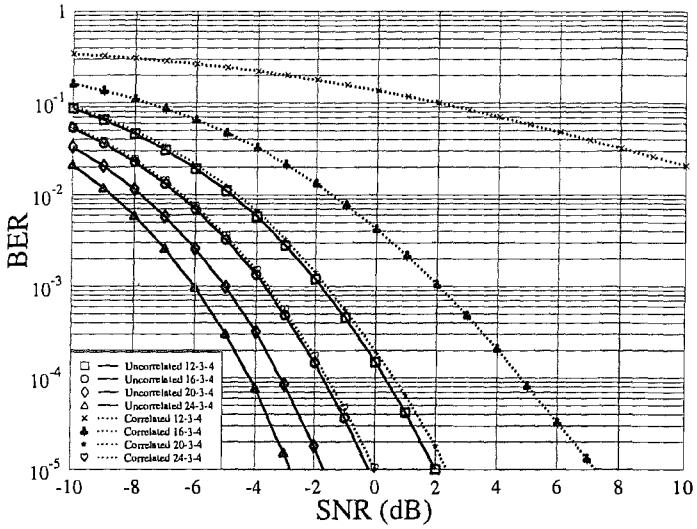


Figure 7.10: Average BER versus average SNR per symbol evaluted from (7.47) for uplink transmission when invoking the maximum SNR based power allocation policy for the four different schemes of Table 7.3.

mance is plotted for uplink transmissions, when invoking both the joint and the individual maximum information rate based power allocation policy of [64], respectively, for the four diffent transmission schemes of Table 7.2. We can see from Figure 7.12 that the attainable capacity is similar for both the joint and the individual policy for all four schemes of Table 7.2, when uncorrelated noise is assumed. Furthermore, the more users are supported, the more serious the MUI and hence the lower the average capacity.

Figure 7.13 shows the attainable system capacity versus the average SNR per symbol performance for downlink transmissions, when invoking both the joint and the individual maximum information rate based power allocation policy of [64], respectively, for the four different transmission schemes of Table 7.1. We can see from Figure 7.13 that the highest system capacity is achieved by different specific schemes of Table 7.1 for different SNR values. The power constraints of (7.74) and (7.79) have a significant influence on the system capacity.

Figure 7.14 characterizes the achievable average capacity versus the average SNR per symbol performance for downlink transmissions when invoking both the joint and the individual maximum information rate based power allocation policy of [64], respectively, for the four diffent transmission

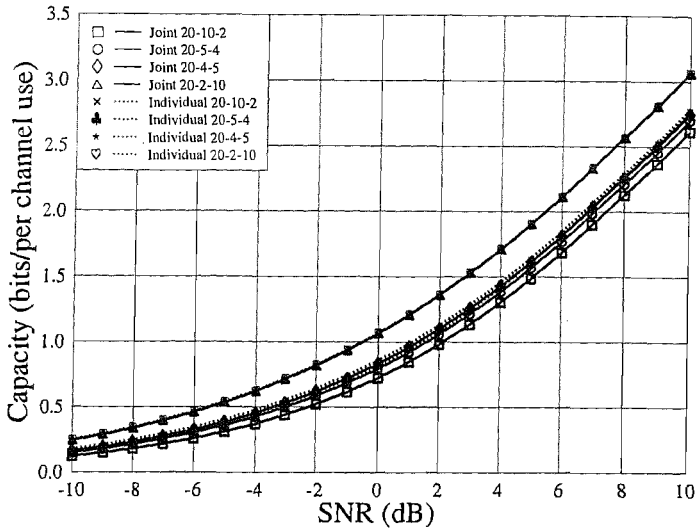


Figure 7.11: Average capacity versus average SNR per symbol evaluted from (7.75) and (7.81) for downlink transmission when invoking joint and individual maximum information rate based power allocation policy, respectively, for four schemes of Table 7.1.

schemes of Table 7.3. We can see from Figure 7.14 that the capacity increases upon increasing the number of antennas at the BS for both the joint and the individual policy of [64], as a benefit of the increased receiver diversity gain. Furthermore, for a given scheme, the capacity remains similar for both the joint and individual policy.

In Figure 7.15 we recorded the achievable average BER versus the average SNR per symbol performance for downlink transmission, when invoking the maximum SNR and the equally normalized power allocation of policy [88, 153], respectively, for the four different schemes of Table 7.1. We can see from Figure 7.15 that the BER performance corresponding to the maximum SNR based power allocation policy of [88] is always better than that of the equal-power policy of [153]. This is because the equal-power normalized policy assigns the same power to each data stream. Hence, some data streams having low SNR values may dominate and hence degrade the average BER performance. However, in order to maximize the expression in (7.82), the maximum SNR based power allocation policy may assign more power to weaker data stream in order to prevent this BER performance degradation. Moreover, for the equal-power policy, the BER remains similar upon increasing the number of users up to full load. The reason for this observation is similar to that in Figure 7.8, when we have

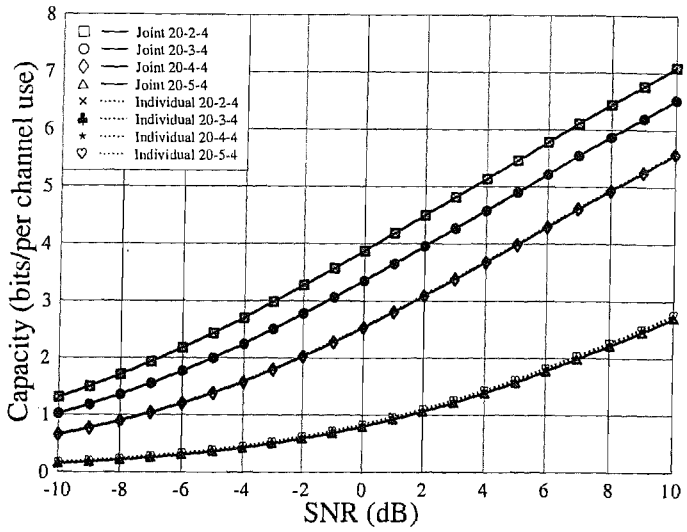


Figure 7.12: Average capacity versus average SNR per symbol evaluted from (7.75) and (7.81) for downlink transmission when invoking joint and individual maximum information rate based power allocation policy, respectively, for the four diffent transmission schemes of Table 7.2.

less antennas at the MS, we support more users, which results in more serious MUI.

In Figure 7.16 the achievable average BER versus average SNR per symbol is plotted for downlink transmissions, when employing maximum SNR and the equal-power policy of [88, 153], respectively, for the four different schemes of Table 7.2. We can see from Figure 7.16 that the BER performance of the maximum SNR policy of [88] is always better than that of the equal-power policy [153]. The reason for this observation is the same as that argued in the context of Figure 7.15. Furthermore, the BER performance improves upon decreasing the number of supported users in the context of both policies, since an extra transmit diversity gain is attained.

In Figure 7.17 the achievable average BER versus average SNR per symbol performance is shown for downlink transmissions, when using the maximum SNR and the equal-power policy of [88, 153], respectively, for the four different schemes of Table 7.3. As observed in Figure 7.17, the BER performance improves upon increasing the number of antennas at the MS for both the maximum SNR and the equal-power policy due to the increased transmit diversity gain experienced. Furthermore, for a given scheme of Table 7.3 the BER performance corresponding to the equal-power policy of [153] is always worse than that of the maximum SNR policy [88] . The reason for this is similar to those

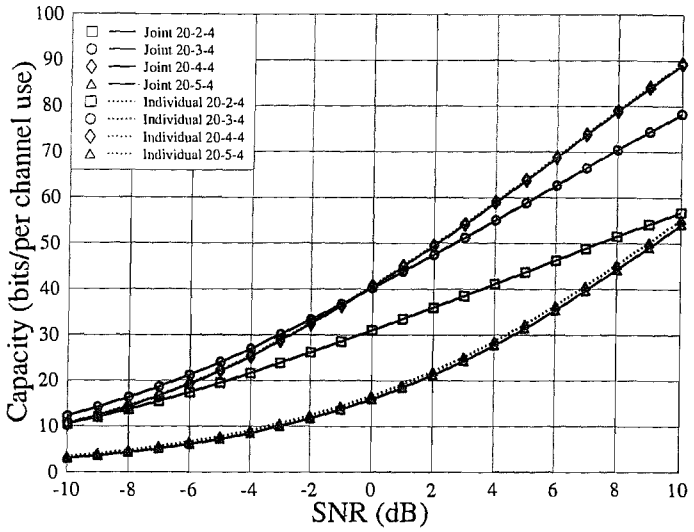


Figure 7.13: System capacity versus average SNR per symbol valuted from (7.75) and (7.81) for downlink transmission when invoking joint and individual maximum information rate based power allocation policy, respectively, for the four diffent transmission schemes of Table 7.2.

outlined in our arguments provided in the context of Figure 7.15. There may exist some lower eigenvalues of the channel matrix, which result in lower effective parallel channel gains. However, the equal-power policy simply allocates the same power to each transmitted symbol and ignores this effect. Hence there may be some low SNR data streams dominating the BER performance in the context of the equal-power policy. Moreover, the BER performance corresponding to a higher number of antennas at the BS for the equal-power policy may become better than that corresponding to a lower number of antennas at the BS combined with the maximum SNR policy. This is because the increased transmit diversity gain may be able to compensate the potential performance loss due to the lower-power data stream of the equal-power policy.

7.6 Conclusions

In this chapter, SVD based SDMA MUD algorithms were proposed for both uplink and downlink transmissions. Based on the proposed algorithm, both the maximum information rate and the maximum SNR based power allocation policies of [64, 88, 185] were considered for uplink transmisson

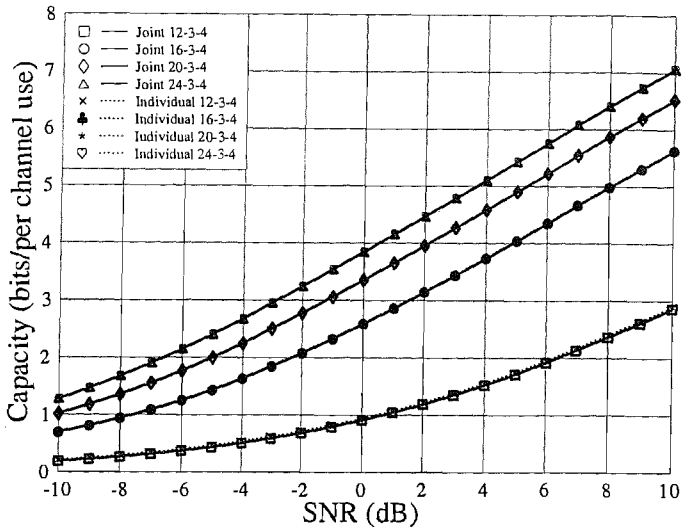


Figure 7.14: Average capacity versus average SNR per symbol evaluted from (7.75) and (7.81) for downlink transmission when invoking joint and individual maximum information rate based power allocation policy, respectively, for the four diffent transmission schemes of Table 7.3.

and the attainable capacity and BER performance was compared for both the uncorrelated and correlated noise scenarios. The simulation results of Figures 7.3 to 7.10 suggest that a performance loss is experienced due the experienced noise enhancement in the correlated noise scenarios. For downlink transmission, the joint and the individual maximum information rate based power allocation policies of [64] were considered first. The simulation results of Figures 7.11 to 7.14 suggest that both of them achieve a similar performance, while and the power constraints of (7.74) and (7.79) have a significant influence on the attainable system performance. Furthermore, the maximum SNR and equal-power allocation policies of [88, 151] were compared for downlink transmissions as well. The simulation results of Figures 7.15 to 7.17 suggest that the BER performance corresponding to the maximum SNR policy is better than that of the equal-power policy, since the equal-power policy ingores the fact that the lower eigenvalues of the channel matrix result in a lower effective channel gain and simply allocate an equal power to each transmitted symbol. Hence in the context of the equal-power policy some of the lower-SNR data streams will degrade the overall BER performance, while the maximum-SNR policy is capable of allocating more power to the specific data streams having lower eigenvalues in order to achieve an improved balance among all data streams. The TDD mode may exploit that the

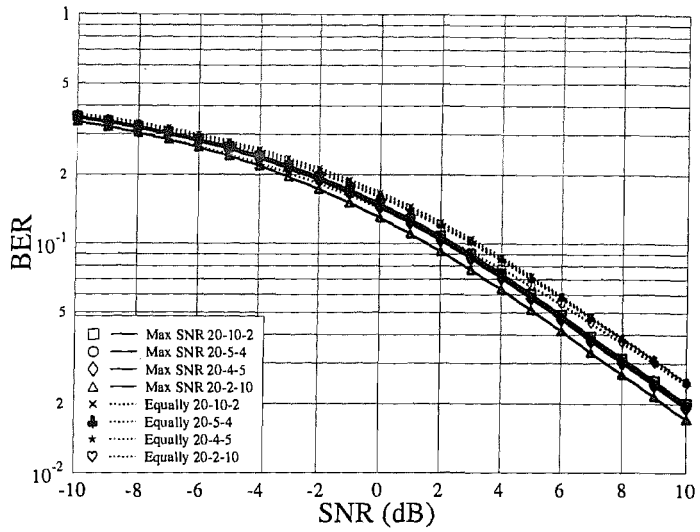


Figure 7.15: Average BER versus average SNR per symbol evaluated from (7.85) and (7.89) for downlink transmission when invoking maximum SNR and equally normalized policy, respectively, for the four different schemes of Table 7.1.

UL and DL channels are similar and invoke this knowledge for downlink transmit preprocessing, as discussed in Section 7.4.

The main advantage of the proposed algorithm is that the characteristics of individual users' channels are taken into account instead of treating all the users' channels jointly, as in the traditional MMSE or zero-forcing MUD technique. Secondly, both uplink and downlink transmissions were considered based on the same MUD scheme instead of considering only downlink transmission in the context of the block diagonalization algorithm of [152, 164].

Morever, based on the SVD-aided based SDMA MUD algorithm, SVD-assisted STBC and BLAST type system [178–180] may also be readily created for multiuser scenarios. Furthermore, sophisticated adaptive modulation schemes may be employed in the context of multiple users for both uplink and downlink transmissions by adjusting the related parameters, such as the transmit power, the transmission rate etc in order to maximize the throughput or minimize the transmission power and so on [9, 177]. A typical application of this scheme can be found in multimedia communications, where different modulation schemes can be chosen to satisfy the different quality-of-service (QoS) requirements imposed by different media [188, 189].

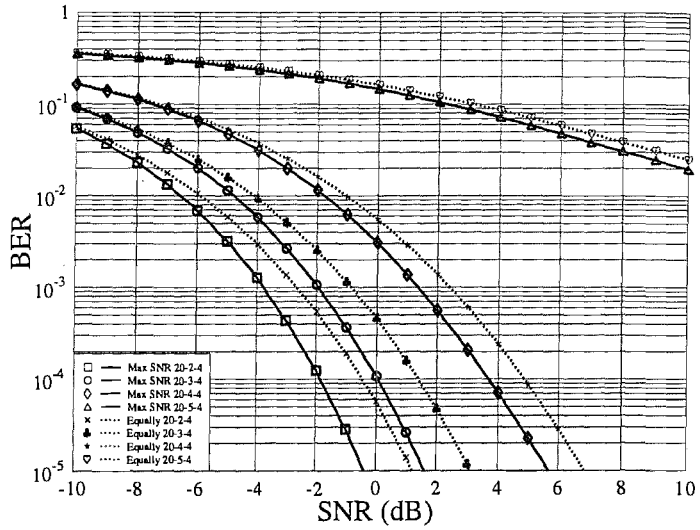


Figure 7.16: Average BER versus average SNR per symbol evaluated from (7.85) and (7.89) for downlink transmission when invoking maximum SNR and equally normalized policy, respectively, for the four different schemes of Table 7.2.

Moreover, since only the knowledge of the singular eigen vectors is required at the MS or the BS to implement the proposed algorithm, subspace tracking techniques may be invoked for directly determining the future SVD components, instead of estimating the channel matrix and repeatedly employing SVD, which results in a high complexity [130–132].

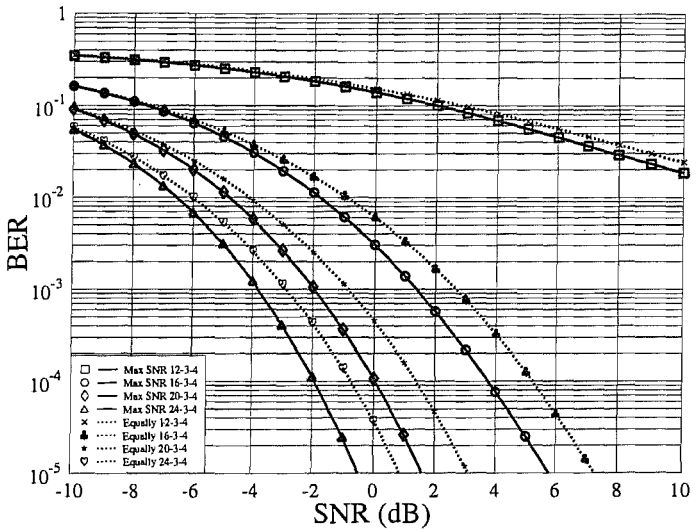


Figure 7.17: Average BER versus average SNR per symbol evaluated from (7.85) and (7.89) for downlink transmission when invoking maximum SNR and equally normalized policy, respectively, for the four different schemes of Table 7.3.

Chapter 8

Preliminary Results for Future Research

8.1 Introduction

In the previous chapters, the non-dispersive fading channel was linearly predicted which was readily justifiable, since the channal can be accurately modelled by a tapped delay line [5, 32]. However, the presence of strong nonlinear correlation among the channel coefficients has also been reported [46,47,49–51]. As a design alternative, neural network based nonlinear algorithms may be invoked for carrying out the task of channel prediction. Neural networks have also been proposed for numerous nonlinear modelling applications, for example in the context of audio signal processing [190–192], signal processing for tele-communication [193, 194] and so on. Recurrent neural networks (RNNs) were reported to benefit from having a feedback from their outputs [195]. The resultant scheme may be considered as an infinte impulse response filter [196]. This property renders them suitable for nonlinear speech processing [197–204], or for classification-based channel equalization [205, 206]. In this chapter, the application of RNNs for narrowband channel prediction is investigated.

The rest of this chapter has the following structure. The RNN-based channel predictor is introduced in Section 8.2, while the RNN's activation function and training algorithms are discussed in Section 8.3 and Section 8.4, respectively. The application of RNN-based channel predicors in non-Gaussian channels is discussed in Section 8.5. Our simulation results are provided in Section 8.6. Finally, our conclusions are offered in Section 8.7.

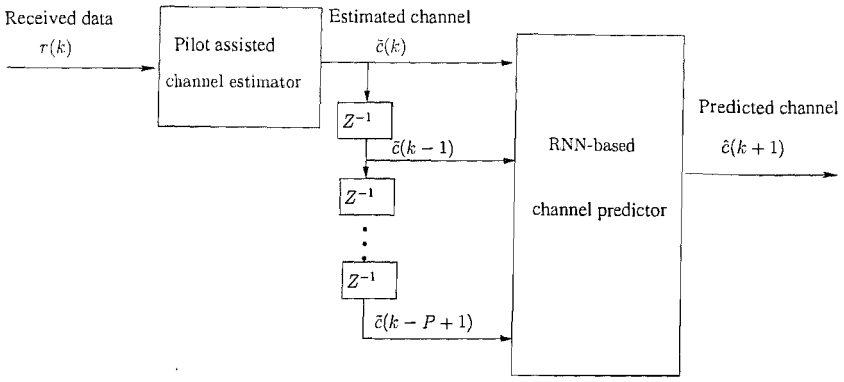


Figure 8.1: Schematic of an RNN-based channel predictor

8.2 Narrowband Channel Prediction Using FCRNNs

As shown in (3.1) of Chapter 3 for a narrowband fading channel, the sampled received signal $r(k)$ is given by

$$r(k) = c(k)b(k) + n(k), \tag{8.1}$$

where $c(k)$ is obtained by sampling the complex-valued fading channel $c(t)$ at the time instant of $t = kT_b$ and T_b is the data symbol duration, $b(k)$ is the kth transmitted symbol value, while $n(k)$ is a complex-valued discrete AWGN process having a variance of $N_0/2$ per dimension.

As we can see in Figure 8.1, if the $\{b(k)\}$ are known pilot symbols, the channel estimate $\tilde{c}(k)$ can be expressed as

$$\begin{aligned} \tilde{c}(k) &= \frac{r(k)}{b(k)} \\ &= c(k) + \frac{n(k)}{b(k)}. \end{aligned} \tag{8.2}$$

As pointed out in [46, 47, 50, 51], $\{\tilde{c}(k)\}$ has to be sampled and the sampling rate is determined based on the average mutual information of samples. In this chapter $\{\tilde{c}(k)\}$ is directly used for channel prediction and the delayed output of the decision-directed channel estimator (DDCE) constitues a $(p \times 1)$-dimensional vector $\tilde{\mathbf{c}}(k)$, which is expressed as

$$\tilde{\mathbf{c}}(k) = [\tilde{c}(k), \cdots, \tilde{c}(k-P+1)]. \tag{8.3}$$

Then $\tilde{\mathbf{c}}(k)$ of (8.3) is fed into the RNN-based channel predictor in order to generate the predicted

channel $\hat{c}(k+1)$.

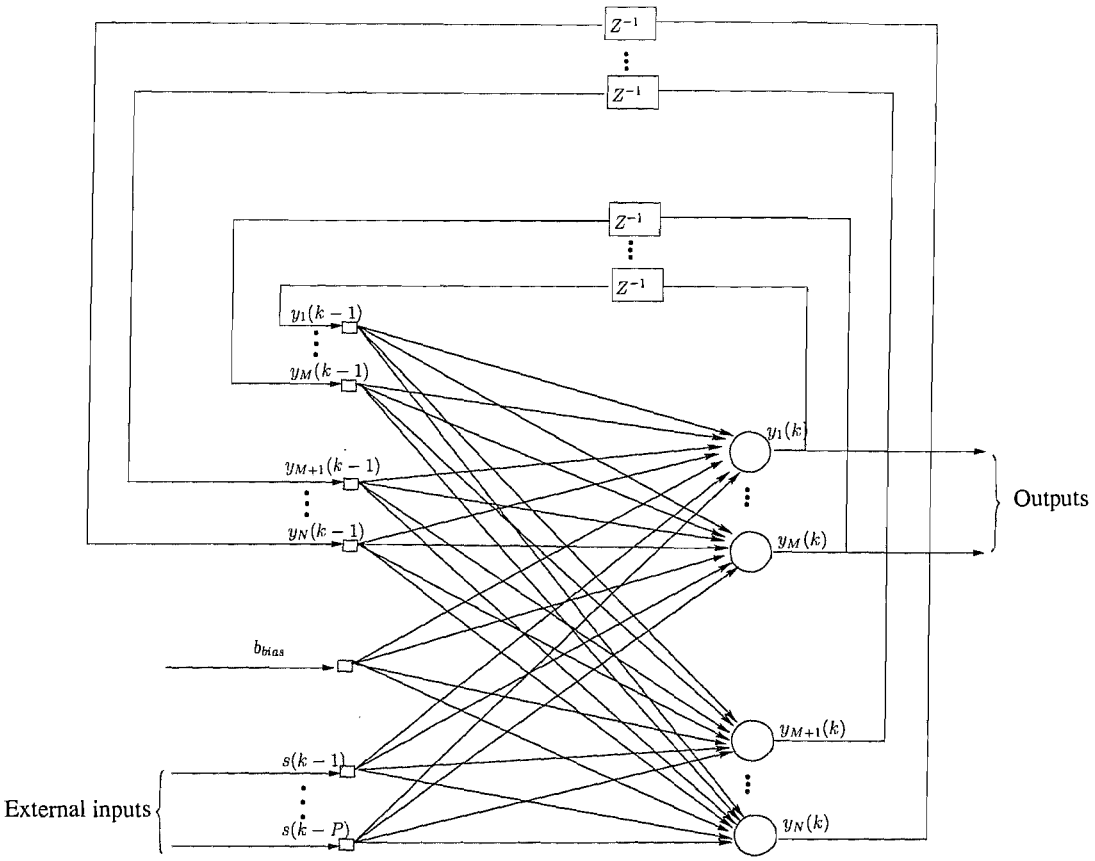


Figure 8.2: A fully connected recurrent neural network

The structure of the RNN-based channel predictor is shown in Figure 8.2. The family of fully connected recurrent neural networks (FCRNNs) constitutes a specific subclass of recurrent neural networks, where every neuron of the output layer is fed back into the input layer and every neuron of the input layer is connected to every neuron of the output layer in the network [195, 207]. Since typically complex-valued channels are encountered in wireless communications, complex-valued neural networks have been developed [208]. They tend to impose a lower computational complexity in comparison to their real-valued conterparts in the context of wireless communication applications [208]. Hence in this chapter the family of complex-valued FCRNNs is discussed and the set of real-valued FCRNNs can be considered as a special case.

Briefly, Figure 8.2 shows an FCRNN, which consists of N activation neurons having P external inputs as well as a fixed bias input b_{bias}. The network has two distinct layers consisting of the

external input-feedback layer and a layer of processing elements. Let the $(N \times 1)$-dimensional vector $\mathbf{y}(k) = [y_l(k), \cdots, y_N(k)]^T$ denote the complex-valued output vector of all the neurons at time index k, where $y_n(k)$ is the complex-valued output of the nth neuron, $n = 1, \cdots, N$, and the $(P \times 1)$-dimensional vector $\mathbf{s}(k) = [s(k-1), \cdots, s(k-P)]^T$ denotes the complex-valued external input vector at time index k. Furthermore, let the bias be $b_{bias} = (1+j)$, where $j = \sqrt{-1}$. Then the $(P+1+N) \times 1$-dimensional combined input vector $\boldsymbol{\rho}(k)$ of the network seen in Figure 8.2 represents the concatenation of the vectors $\mathbf{s}(k)$, b_{bias} and $\mathbf{y}(k-1)$, which is given by

$$\begin{aligned} \boldsymbol{\rho}(k) &= \begin{bmatrix} \mathbf{s}(k) \\ b_{bias} \\ \mathbf{y}(k-1) \end{bmatrix} \\ &= [s(k-1), \cdots, s(k-P), 1+j, y_1(k-1), \cdots, y_N(k-1)]^T \\ &= \boldsymbol{\rho}^{(r)}(k) + j\boldsymbol{\rho}^{(i)}(k), \end{aligned} \tag{8.4}$$

where the superscripts $(\cdot)^{(r)}$ and $(\cdot)^{(i)}$ denote the real and imaginary parts of the argument, respectively. Assuming that the outputs of the first M neurons are the outputs of the FCRNN, the $(M \times 1)$-dimensional output vector of the FCRNN $\mathbf{y}_o(k)$ can be expressed as

$$\mathbf{y}_o(k) = [y_1(k), \cdots, y_M(k)]^T, \; 1 \leq M \leq N. \tag{8.5}$$

Let $w_{n,l}$ denote the complex-valued weight of the FCRNN, which connects the nth neuron and lth input, where we have $1 \leq n \leq N$ and $1 \leq l \leq (P+1+N)$. Then the input of the nth node at time index k is given by

$$\begin{aligned} n_{et_n}(k) &= \sum_{l=1}^{P+1+N} w_{n,l}(k)\rho_l(k) \\ &= n_{et_n}^{(r)}(k) + jn_{et_n}^{(i)}(k), \end{aligned} \tag{8.6}$$

where $n_{et_n}^{(r)}(k)$ and $n_{et_n}^{(i)}(k)$ are the real and imaginary parts of $n_{et_n}(k)$, respectively, which can be

expressed as [209]

$$\begin{aligned} n_{et_n}^{(r)}(k) &= \sum_{l=1}^{P+1+N} [w_{n,l}^{(r)}(k)\rho_l^{(r)}(k) - w_{n,l}^{(i)}(k)\rho_l^{(i)}(k)] \\ &= \sum_{l=1}^{P} [w_{n,l}^{(r)}(k)s^{(r)}(k-l) - w_{n,l}^{(i)}(k)s^{(i)}(k-l)) + (w_{n,P+1}^{(r)}(k) - w_{n,P+1}^{(i)}(k)] + \\ &\quad \sum_{q=1}^{N} [w_{n,P+1+q}^{(r)}(k)y_q^{(r)}(k-1) - w_{n,P+1+q}^{(i)}(k)y_q^{(i)}(k-1)] \end{aligned} \tag{8.7}$$

$$\begin{aligned} n_{et_n}^{(i)}(k) &= \sum_{l=1}^{P+1+N} [w_{n,l}^{(r)}(k)\rho_l^{(i)}(k) + w_{n,l}^{(i)}(k)\rho_l^{(r)}(k)] \\ &= \sum_{l=1}^{P} [w_{n,l}^{(r)}(k)s^{(i)}(k-l) + w_{n,l}^{(i)}(k)s^{(r)}(k-l)) + (w_{n,P+1}^{(r)}(k) + w_{n,P+1}^{(i)}(k)] + \\ &\quad \sum_{q=1}^{N} [w_{n,P+1+q}^{(r)}(k)y_q^{(i)}(k-1) + w_{n,P+1+q}^{(i)}(k)y_q^{(i)}(k-1)]. \end{aligned} \tag{8.8}$$

The output of the nth activation neuron can be expressed as [209]

$$\begin{aligned} y_n(k) &= \Phi(n_{et_n}(k)) \\ &= \Phi^{(r)}(n_{et_n}(k)) + j\Phi^{(i)}(n_{et_n}(k)) \\ &= y_n^{(r)}(k) + jy_n^{(i)}(k),\ n = 1, \cdots, N, \end{aligned} \tag{8.9}$$

where Φ is a complex-valued nonlinear activation function, while $y_n^{(r)}(k)$ and $y_n^{(i)}(k)$ are the real and imaginary parts of $y_n(k)$, respectively.

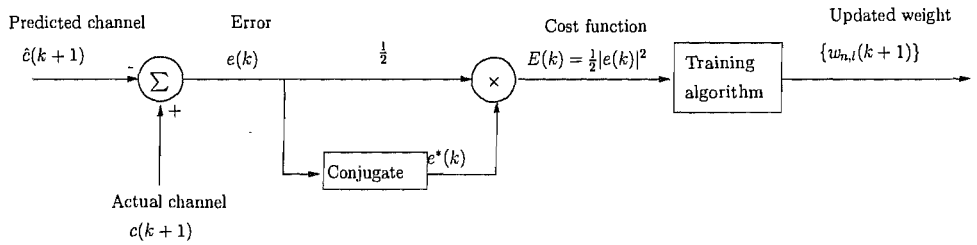


Figure 8.3: Schematic of the weight-update arrangement

Let $d_m(k), 1 \leq m \leq M$ be the correponding desired output of the mth neuron. Then the $(M \times 1)$-

dimensional desired output vector of the FCRNN $\mathbf{d}_o(k)$ can be expressed as

$$\mathbf{d}_o(k) = [d_1(k), \cdots, d_M(k)]^T, \ 1 \leq M \leq N. \tag{8.10}$$

In the context of narrowband channel prediction, we have

$$[s(k-1), \cdots, s(k-p)] = [\tilde{c}(k), \cdots, \tilde{c}(k-P+1)], \tag{8.11}$$

$$y_o(k) = y_1(k) = \hat{c}(k+1), \tag{8.12}$$

$$d_o(k) = d_1(k) = c(k+1). \tag{8.13}$$

Consequently, the error $e(k)$ seen in Figure 8.3 represents the discrepancy between the actual fading channel profile $c(k+1)$ as well as the predicted fading channel profile $\hat{c}(k+1)$ and is given by

$$e(k) = c(k+1) - \hat{c}(k+1). \tag{8.14}$$

Then the error $e(k)$ of (8.14) is fed to a training algorithm in order to generate the updated weights, as shown in Figure 8.3, until a sufficiently accurate prediction is achieved.

8.3 Activation Function

In this chapter, two different activation functions are investigated, namely the split activation and the full activation function [209]. The terminology "split activation function" is used for indicating that the complex-valued input is divided into two real-valued components which are treated seperately to derive the two real-valued outputs that are combined to generate a complex-valued output [210, 211]. By contrast, when using a "full activation function", the complex-valued input is fed into an activation function, which directly generates a complex-valued output [209]. Specifically, in conjunction with the split activation function, (8.9) is expressed as

$$\begin{aligned} y_n(k) &= f(n_{etn}^{(r)}(k)) + jf(n_{etn}^{(i)}(k)) \\ &= \frac{1}{1+e^{-\beta n_{etn}^{(r)}(k)}} + j\frac{1}{1+e^{-\beta n_{etn}^{(i)}(k)}}, \end{aligned} \tag{8.15}$$

where $f(x)$ is a function having real-valued input and real-valued output. Here we opted for using the so-called logistic sigmoid function, given by

$$f(x) = \frac{1}{1+e^{-x}}, \tag{8.16}$$

where x is real-valued.

For a full activation function, the complex-valued activation function was directly chosen to be the logistic sigmoid function. Then (8.9) is expressed as

$$y_n(k) = \frac{1}{1+e^{-(n_{etn}^{(r)}(k)+jn_{etn}^{(i)}(k))}}. \tag{8.17}$$

In the context of narrowband channel prediction, when the split activation function of (8.15) is used, the range of the real-valued logistic sigmoid function of (8.16) spans the interval of $(0,1)$, hence the desired outputs of the neural network have to be adjusted so that they fall within this range. The appropriately adjusted desired output of the neural network can be generated in conjunction with the split activation function $c_s(k+1)$ by appropriately shifting and scaling the original desired output with the aid of the following equation

$$\begin{aligned} c_s(k+1) &= \frac{c(k+1)+\alpha_s}{\iota_s} \\ &= \frac{(c^{(r)}(k+1)+\alpha_s^{(r)})+j(c^{(i)}(k+!)+\alpha_s^{(i)})}{\iota_s} \\ &= c_s^{(r)}(k+1)+jc_s^{(i)}(k+1) \quad \iota_s \neq 0, \end{aligned} \tag{8.18}$$

where α_s is the complex-valued shift coefficient and ι_s is the real-valued scaling coefficient. Furthermore, $\alpha_s^{(r)}$ and $\alpha_s^{(i)}$ are the real and imaginary parts of α_s in (8.18), respectively. Specifically, in our simulations outlined in this chapter, we had $\alpha_s = 5+j5$ and $\iota_s = 10$, which resulted in the minimum MSE, when they were varied within a reasonable range. Accordingly, the output $y_o(k)$ of the FCRNN is not the prediction of the actual channel profile $c(k+1)$, but the adjusted actual channel profile $c_s(k+1)$, which is expressed as

$$y_o(k) = \hat{c}_s(k+1). \tag{8.19}$$

Consequently, the error and the cost function can be expressed as

$$e(k) = c_s(k+1) - \hat{c}_s(k+1), \tag{8.20}$$

$$E(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}|c_s(k+1) - \hat{c}_s(k+1)|^2. \tag{8.21}$$

When the full activation function of (8.17) is used, the desired output of the neural network also has to be adjusted, so that it falls within the range the complex-valued of logistic sigmoid function of (8.17). In this case, the desired output of the neural network adjusted for invoking the full activation function is given by

$$\begin{aligned} c_f(k+1) &= \frac{c(k+1) + \alpha_f}{\iota_f} \\ &= \frac{(c^{(r)}(k+1) + \alpha_f^{(r)}) + j(c^{(i)}(k+1) + \alpha_f^{(i)})}{\iota_f} \\ &= c_f^{(r)}(k+1) + jc_f^{(i)}(k+1) \quad \iota_f \neq 0, \end{aligned} \tag{8.22}$$

where α_f is a complex-valued shift coefficient and ι_f is a real-valued scaling coefficient, respectively, while $\alpha_f^{(r)}$ and $\alpha_f^{(i)}$ are the real and imaginary parts of α_f in (8.22), respectively. In our simulations highlighed in Section 8.6 of this chapter, we used the values of $\alpha_f = 50 + j50$ and $\iota_f = 80$, which resulted in the minimum MSE.

Similarly, the output $y_o(k)$ of FCRNN is not the prediction of the actual channel profile $c(k+1)$, but the adjusted actual channel profile $c_f(k+1)$, which is expressed as

$$y_o(k) = \hat{c}_f(k+1). \tag{8.23}$$

By contrast, when the full activation function is used, the error and the cost function, respectively, can be expressed as

$$e(k) = c_f(k+1) - \hat{c}_f(k+1), \tag{8.24}$$

$$E(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}|c_f(k+1) - \hat{c}_f(k+1)|^2. \tag{8.25}$$

8.4 Training Algorithms

In this chapter, three different training algorithms are invoked, namely the real time recurrent learning (RTRL) scheme [209–215], the global extended Kalman filter (GEKF) [216] and the decoupled extended Kalman filter (DEKF) based training [216]. In this section we assume that the reader is familiar with the these principles, which have been outlined in Appendix A.

8.5 Non-Gaussian Channel Prediction

So far channel prediction has been carried out under the assumption that the noise $n(k)$ can modelled as AWGN. However, in practical communications systems we often encounter impulsive noise, which is non-Gaussian [217–219]. In this section, the application of RNNs in the prediction of non-Gaussian channels is investigated. For brevity and mathematical tractability, we only consider split activation, but this technique can be extended to the employment of the full activation function.

A popular impulsive noise model is the α-stable random distribution [218, 219]. The symmetric α-stable (SαS) Probability Density Function (PDF) is defined by its characteristic function which is the Fourier transform of the PDF, as given by [218, 219]

$$P(t) = e^{-\gamma|t|^{\alpha}}, \tag{8.26}$$

where the parameters γ ($\gamma > 0$) and α ($\alpha \in (0, 2]$) define the specific SαS distribution. When we have $\alpha = 2$, this scenario corresponds to the zero-mean normal distribution having a variance of 2γ, while it turns into the Cauchy distribution [218], when we have $\alpha = 1$. The SαS noise samples can be generated from the following random variable [218, 219]

$$p_{\alpha}(t) = \gamma^{\frac{1}{\alpha}} \frac{\sin(\alpha t)}{(\cos t)^{\frac{1}{\alpha}}} \left(\frac{\cos((1-\alpha)t)}{A} \right)^{\frac{1-\alpha}{\alpha}}, \alpha \neq 1, \tag{8.27}$$

where t is uniformly distributed in $(-\pi/2, \pi/2)$ and A obeys the standard exponential distribution.

The SαS noise has an infinite variance [218], which makes the use of the standard SNR meaningless. In this section, the so-called geometric signal-to-noise ratio (GSNR) defined in [220] is used in order to indicate the power ratio of the information-bearing signal and the SαS noise [218], which is expressed as

$$GSNR = \frac{1}{2C_g} \left(\sqrt{E} S_0 \right)^2, \tag{8.28}$$

where $C_g \simeq 1.78$ is the exponential of the Euler's constant [218], and $S_0 = (C_g\gamma)^{\frac{1}{\alpha}}/C_g$ denotes the geometric power. In general the noise power increases, as α decreases and when α approaches zero, the noise power tends to infinity.

8.6 Performance Results

8.6.1 Channel Prediction Performance for a Gaussian Channel

In this chapter, the mean square error (MSE) metric is used for quantifying the attainable performance of the various fading channel predictors. More specifically, for the split activation function of (8.15), the resulant MSE is given by

$$\begin{aligned}
\text{MSE} &= \frac{1}{K}\sum_{k=1}^{K}|c(k+1)-(\hat{c}_s(k+1)\iota_s-\alpha_s)|^2 \\
&= \frac{1}{K}\sum_{k=1}^{K}|(c_s(k+1)\iota_s-\alpha_s)-(\hat{c}_s(k+1)\iota_s-\alpha_s)|^2 \\
&= \frac{1}{K}\sum_{k=1}^{K}|c_s(k+1)\iota_s-\hat{c}_s(k+1)\iota_s|^2 \\
&= \frac{1}{K}\iota_s\sum_{k=1}^{K}|c_s(k+1)-\hat{c}_s(k+1)|^2 \\
&= \frac{2}{K}\iota_s^2\sum_{k=1}^{K}E(k). \qquad (8.29)
\end{aligned}$$

By contrast, for the full activation function of (8.17) the MSE expression is formulated as

$$\begin{aligned}
\text{MSE} &= \frac{1}{K}\sum_{k=1}^{K}|c(k+1)-(\hat{c}_f(k+1)\iota_f-\alpha_f)|^2 \\
&= \frac{2}{K}\iota_f^2\sum_{k=1}^{K}E(k), \qquad (8.30)
\end{aligned}$$

where K is the total number of the channel samples, which were predicted after the training has been completed. Furthermore, in our simulations, BPSK modulation was employed and $E(|c(k)|^2) = 1$ was assumed, as in Chapter 3. In Figure 8.4, the achieveable channel prediction MSE performance versus the number of external inputs is quantified for both the SCRTRL and FCRTRL schemes of (8.29) and (8.30) as a function of the different number of activations $N = 1, 2, 3$ and 4, when using

Learning rate η	0.01
Number of data for training	10000
Number of data for calculating MSE after training	10000
Maximum doppler frequency f_{dm}	120 Hz
Data rate f_b	4ksymbols/s
Sampling rate f_s	4kHz
SNR	10dB

Table 8.1: Parameters for calculating the MSE versus the number of external inputs

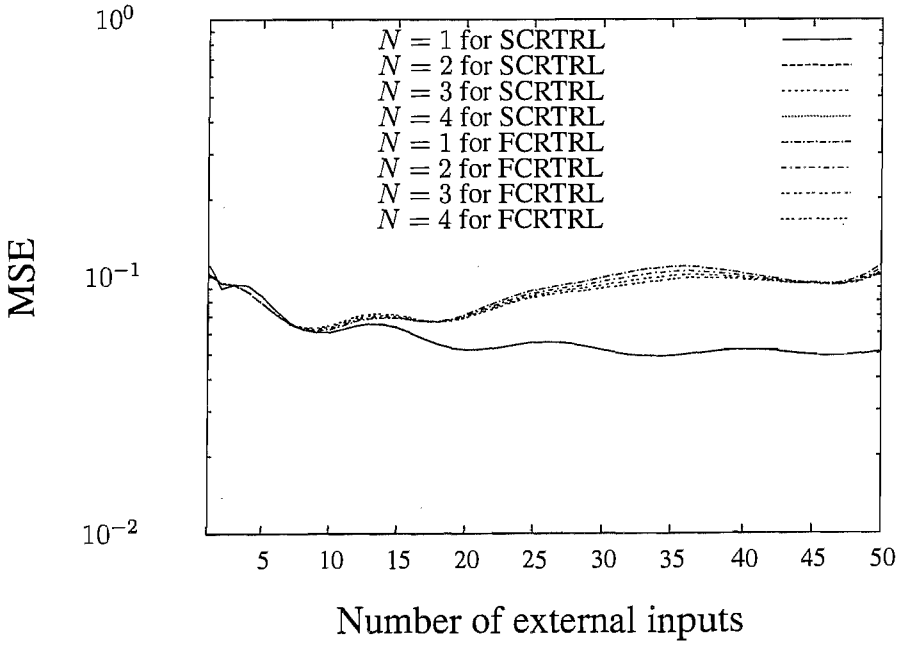


Figure 8.4: MSE performance versus the number of external inputs for both the SCRTRL and FCRTRL in conjunction with different number of activations $N = 1, 2, 3$ and 4 when the power intensity of the channel was set to 1.0. The remaining parameters are given in Table.8.1.

the parameters of Table.8.1. Observe from Figure 8.4 that for the SCRTRL scheme of (8.29) the MSE performaces of the different number of activations $N = 1, 2, 3$ and 4 are similar, regardless of the number of external inputs P. This indicates that the number of activation neurons does not appreciably affect the attainable performance and those findings are consistent with those disseminated in [198, 205]. Moreover, when P changes from 1 to 2, the achievable MSE performace decreases nearly linearly and when we have $P > 2$, the MSE performances fluctuate upon increasing the number of external inputs. The minimum MSE was achieved at $P = 34$. On the other hand, as seen in Figure 8.4 for the FCRTRL arrangement of (8.30), the MSE fluctuates upon increasing the number of external inputs for the different number of activations $N = 1, 2, 3$ and 4 and the minimum MSE recorded for the different number of activations $N = 1, 2, 3$ and 4 was obtained, when the number of external inputs was $P = 9$. It is worth mentioning at this stage that the channel predictor's complexity increases as a function of the number of activations. Moreover, when the number of external inputs was $P \leq 9$, the MSE performances recorded for the different number of activations $N = 1, 2, 3$ and 4 remained near-constant. By contrast, when the number of external inputs obeyed $9 < P \leq 18$, there were slight MSE differences among them, and the higher the number of activations, the better the MSE performace. However, observe in Figure 8.4 that when the number of external inputs obeyed $P > 18$, the lower the number of activations, the better the MSE performance and the difference among the various parameter settings became apparent, except when we have $43 < P < 47$, resulting again in similar MSE performances for the different number of activations $N = 1, 2, 3$ and 4. Furthermore, it is seen in Figure 8.4 for $P \leq 6$, that a better MSE performane was achieved by the SCRTRL or FCRTRL schemes of (8.29) and (8.30) and the MSE difference remained small between them. By contrast, observe in Figure 8.4 that for $6 < P < 10$, the associated MSE performances of the SCRTRL and FCRTRL arrangments of (8.29) and (8.30) are almost identical. However, as seen in Figure 8.4 when we have $P \geq 10$, a consistently better MSE performance was attained by the SCRTRL scheme of (8.29). Finally, the number of external inputs was set to $P = 9$ for the FCRTRL scheme of (8.30) in order to achieve the minimum MSE in our forthcoming investigations. Furthermore, the number of external inputs was chosen to be $P = 9$ also for the SCRTRL scheme of (8.29) for the sake of convient comparision to our forthcoming simulation experiments.

In Figure 8.5, the attainable MSE performance versus learning rate η of both the SCRTRL and FCRTRL schemes of (8.29) and (8.30) is characterized using the parameters of Table 8.2. We note from Figure 8.5 that for the SCRTRL scheme of (8.29) the MSE performaces recorded for different number of activations $N = 1, 2, 3$ and 4 are again, quite similar to each other, when the learning rate of (A.15) and (A.43) obeys $\eta < 0.05$. Furthermore, the MSE decreases near-linearly upon

Number of activation N	1
Number of external inputs P	9
Number of data for training	10000
Number of data for calculating MSE	10000
Maximum doppler frequency f_{dm}	120 Hz
Data rate f_b	4ksymbols/s
Sampling rate f_s	4kHz
SNR	10dB

Table 8.2: Parameters used for calculating the MSE versus learning rate

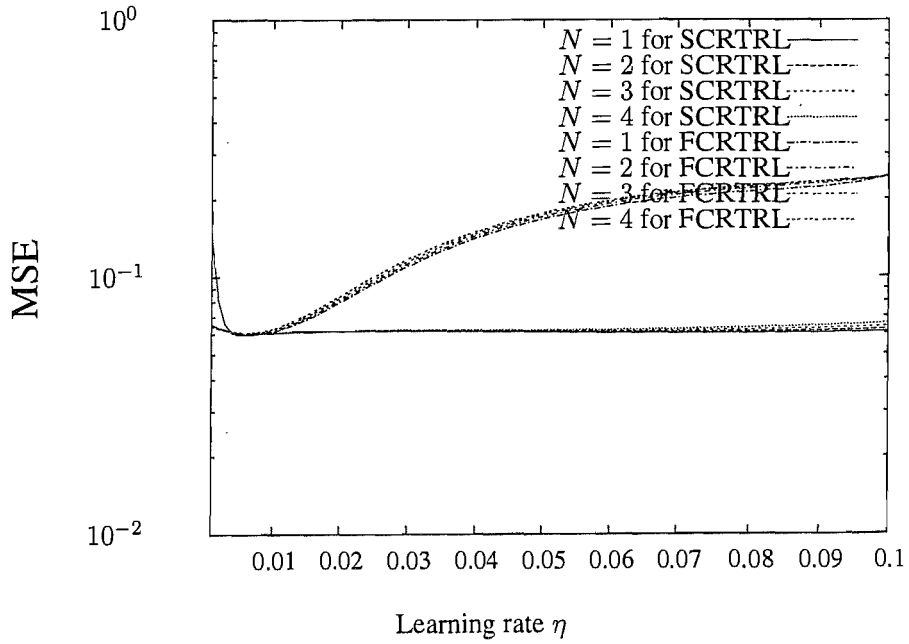


Figure 8.5: MSE performance versus the learning rate η for both the SCRTRL and FCRTRL schemes of (8.29) and (8.30). The remaining parameters were summarized in Table 8.2.

Learning rate η	0.01
Number of activation N	1
Number of external inputs P	9
Number of data for calculating MSE	10000
Maximum doppler frequency f_{dm}	120 Hz
Data rate f_b	4ksymbols/s
Sampling rate f_s	4kHz
SNR	10dB

Table 8.3: Parameters used, when comparing the MSE performances of SCRTRL and FCRTRL

increasing the learning rate η in the range of $\eta < 0.004$ and smoothly after that point, until it reaches its minimum at point of $\eta = 0.006$. Upon increasing η further, it remains near-constant, regardless of the number of activations. On the other hand, the MSE of FCRTRL recorded for different number of activations $N = 1, 2, 3$ and 4 decreases smoothly upon increasing the learning rate η, until it reaches its minimum at point at $\eta = 0.006$ or $\eta = 0.007$. Beyond that point, the MSE recorded for different number of activations $N = 1, 2, 3$ and 4 increases smoothly. Furthermore, the MSE performances found for different number of activations $N = 1, 2, 3$ and 4 were similar, when the learning rate obeyed $\eta < 0.005$ and diverged slightly beyond that point. More specifically, the lower the number of activations, the better the MSE performance, except for $\eta > 0.098$, where the MSE performaces converged to a similar value. Finally, when the learning rate obeyed $\eta < 0.003$, the MSE performace of FCRTRL was significantly better than those of SCRTRL. For values in excess of $\eta = 0.01$, the MSE performance of SCRTRL was significantly better than that of FCRTRL. Furthermore, because the higher the number of activation, the higher the complexity, the number of activations was set to $N = 1$ for both SCRTRL and FCRTRL in our forthcoming simulations.

In Figure 8.6 the MSE performance of the SCRTRL scheme of (8.29) is compared to that of the FCRTRL arrangement of (8.30), when using the parameters of Table 8.3. To elaborate a little further, we can see in Figure 8.6 that at the begining of the training the MSE performance of the SCRTRL scheme of (8.29) is significantly better than that of the FCRTRL arrangement of (8.30). The reason for this observation is that the scaling coefficient of the FCRTRL in (8.30) is $\iota_f = 80$, which is seven times higher than that of the SCRTRL in (8.29) during the course of training. The MSE of the FCRTRL scheme of (8.30) was about 50 times higher than that of the SCRTRL arrangement of (8.29) even for same cost function. By changing learning rate to $\eta = 0.005, 0.006, 0.007, 0.008$ and 0.009, similar performances have been observed for the two schemes. Finally, a learning rate of $\eta = 0.01$ was chosen for both the SCRTRL and FCRTRL arrangements in the following simulations, because

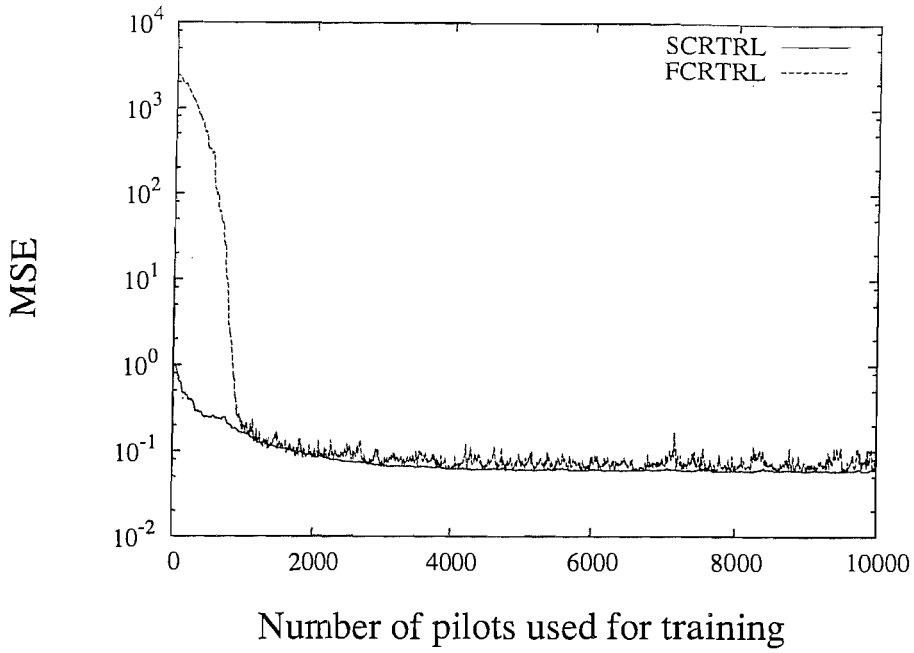


Figure 8.6: MSE performance versus the number of pilots used for training of both SCRTRL and FCRTRL. The remaining parameters were summarized in Table 8.3.

the larger the learning rate, the shorter the training time.

In Figure 8.7, the cost function value of (8.21) versus the number of pilots used for training of the SCRTRL scheme of (8.29) is plotted using the parameters of Table 8.3. We infer from Figure 8.7 that the cost function of the SCRTRL scheme of (8.29) formulated in (8.21) exhibits a gracefully decreasing trend upon increasing the number of training samples.

By contrast, in Figure 8.8 the cost function value of (8.25) versus the number of pilots used for training of the FCRTRL scheme of (8.30) is plotted using the parameters of Table 8.3. Observe in Figure 8.8 that the associated cost function values of the FCRTRL schems of (8.30) apprear to be about two orders of magnitude lower than those of the SCRTRL scheme of (8.29) seen in Figure 8.7.

In Figures 8.9 to 8.14, the complex-valued channel profile prediction performance of both the SCRTRL and FCRTRL schemes of (8.29) and (8.30) is compared in terms of their real part and imaginary part, as well as in terms of the magnitude of the channel envelope using the parameters of Table8.4. In this specific initialization scenario, we had an MSE of 0.060245349 for the SCRTRL scheme of (8.29) and MSE of 0.0752349 for the FCRTRL of (8.30). From Figure 8.9 to Figure 8.14, we can see that for some data points both the SCRTRL and the FCRTRL shemes of (8.29) and (8.30) are capable of accurately predicting the real part and imaginary part of the channel profile as well as

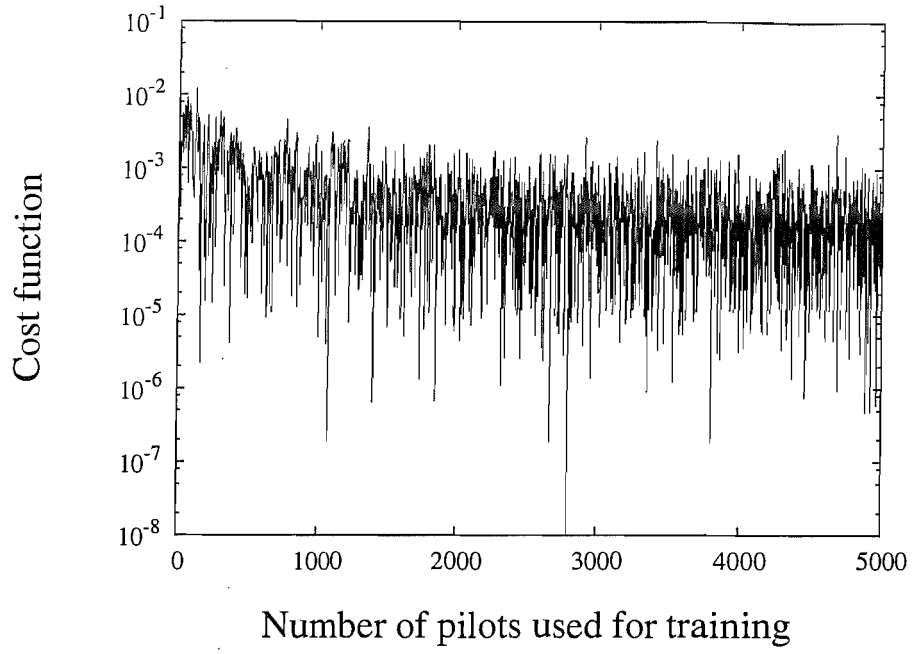


Figure 8.7: The cost function value of (8.21) versus the number of pilots used for training of the SCRTRL scheme of (8.29). The remaining parameters were summarized in Table.8.3.

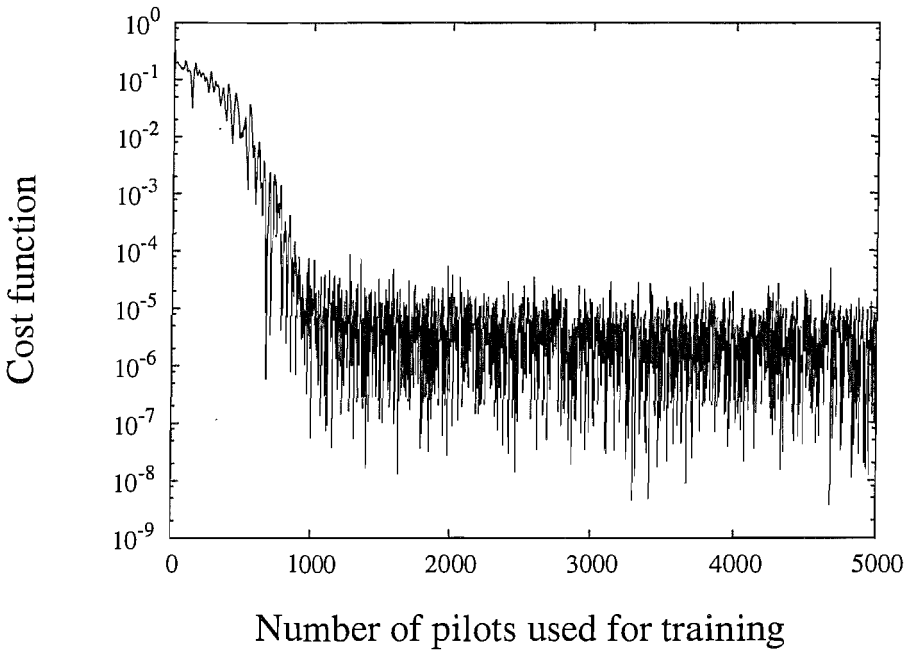


Figure 8.8: The cost function value of (8.25) versus the number of pilots used for training of the FCRTRL scheme of Eq.(8.30). The remaining parameters were summarized in Table.8.3.

Learning rate η	0.01
Number of activation N	1
Number of external inputs P	9
Number of training data samples	5000
Number of data samples for calculating MSE	10000
Maximum doppler frequency f_{dm}	120 Hz
Data rate f_b	4ksymbols/s
Sampling rate f_s	4kHz
SNR	10dB

Table 8.4: Parameters used for quantifying the prediction performances of both SCRTRL and FCRTRL

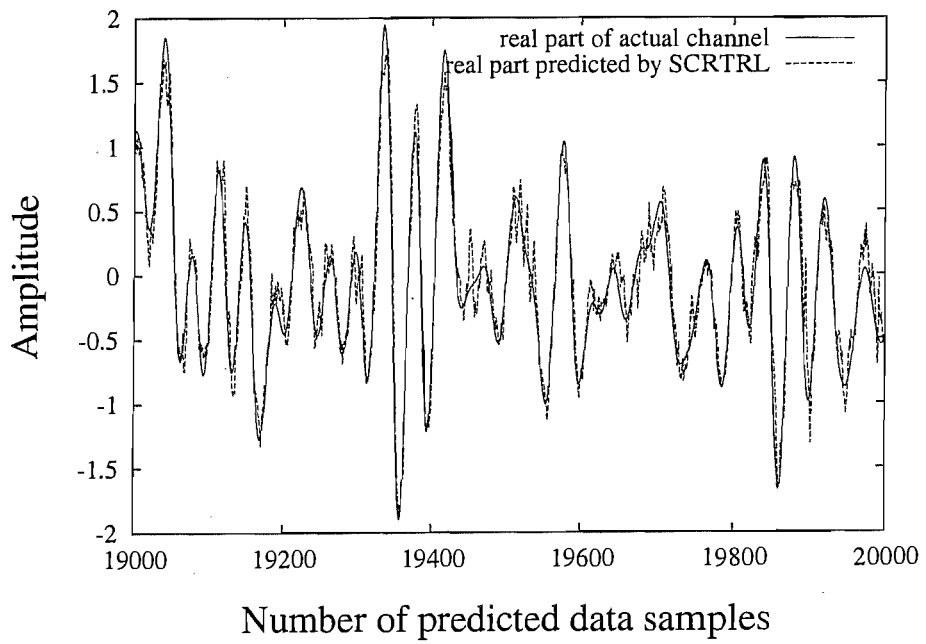


Figure 8.9: The real part of the complex-valued fading channel profile predicted by the SCRTRL scheme of (8.29) versus the number of predicted data. The remaining parameters were summarized in Table 8.4.

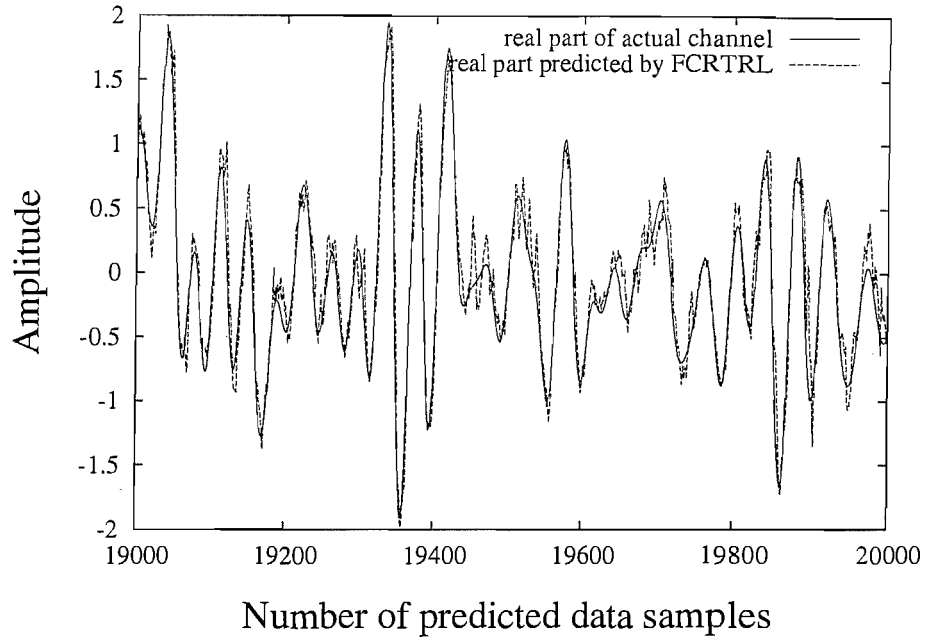


Figure 8.10: The real part of the complex-valued fading channel profile predicted by the FCRTRL scheme of (8.30) versus the number of predicted data. The remaining parameters were summarized in Table 8.4.

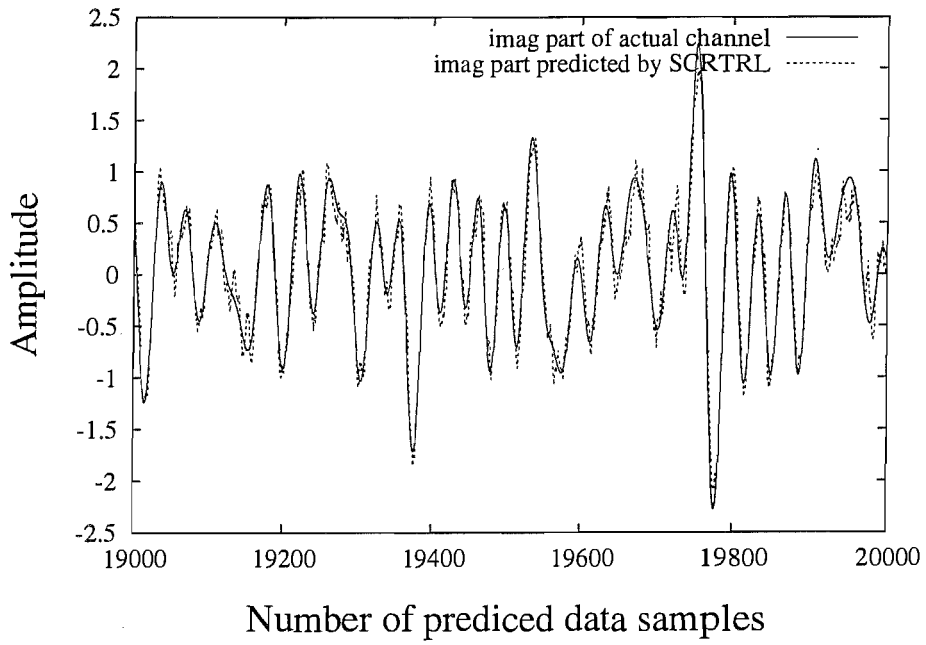


Figure 8.11: The imaginary part of the complex-valued fading channel profile predicted by the SCRTRL scheme of (8.29) versus the number of predicted data. The remaining parameters were summarized in Table 8.4.

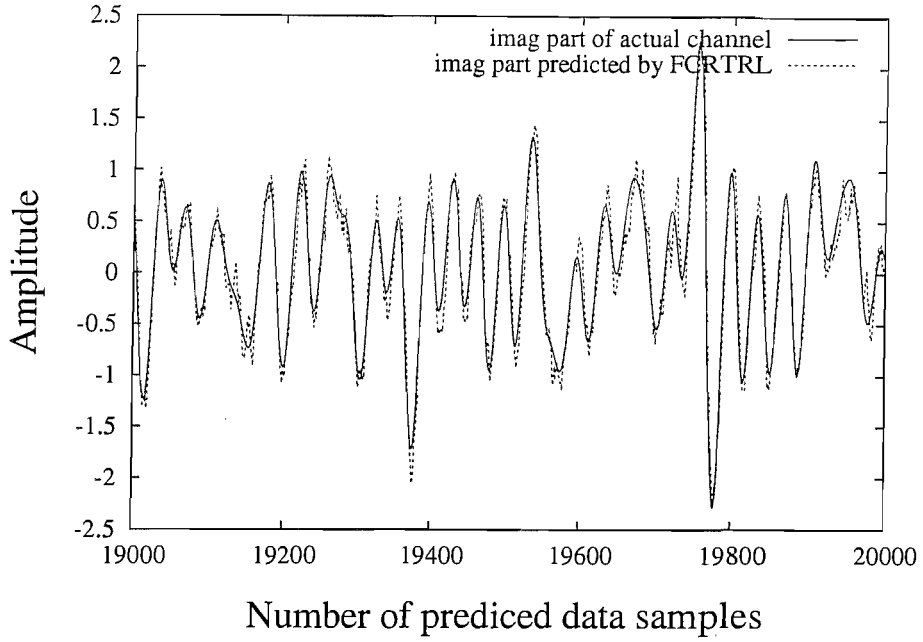


Figure 8.12: The imaginary part of the complex-valued fading channel profile predicted by the FCRTRL scheme of (8.30) versus the number of predicted data. The remaining parameters were summarized in Table 8.4.

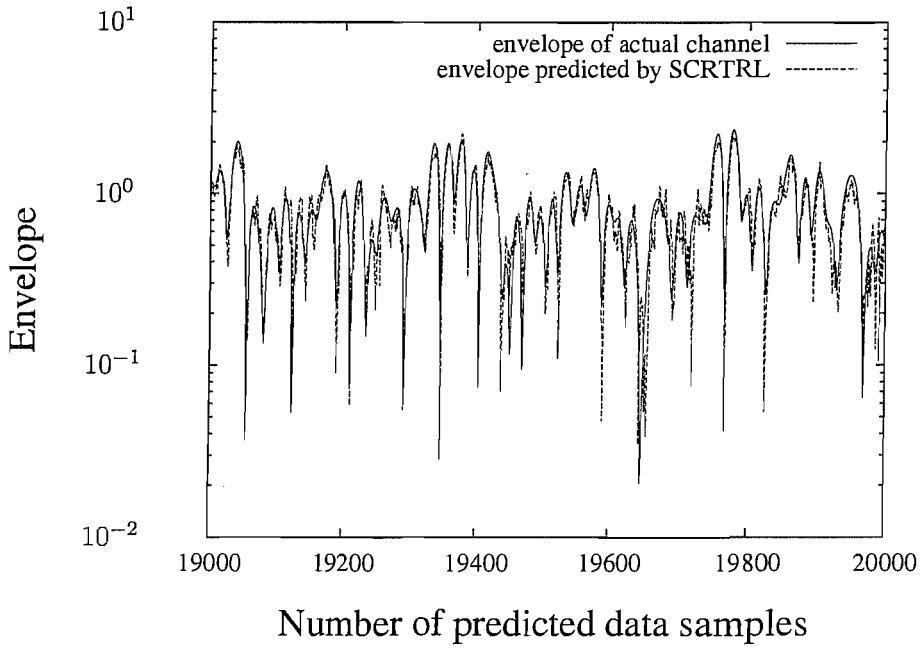


Figure 8.13: Magnitude of the complex-valued fading channel envelope predicted by the SCRTRL scheme of (8.29) versus the number of predicted data. The remaining parameters were summarized in Table 8.4.

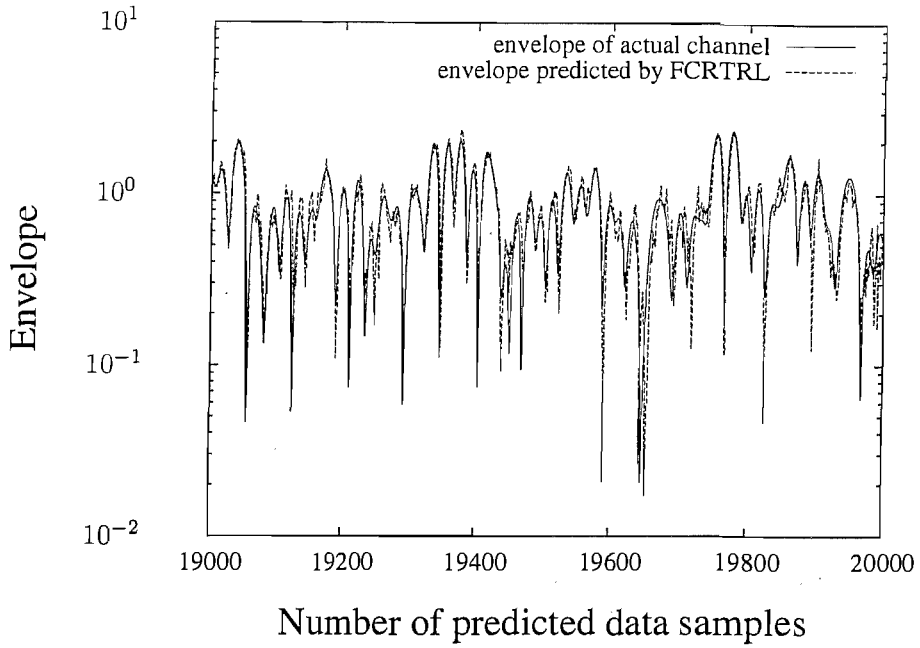


Figure 8.14: Magnitude of the complex-valued fading channel envelope predicted by the FCRTRL scheme of (8.30) versus the number of predicted data. The remaining parameters were summarized in Table 8.4.

Number of activation N	2
Number of external inputs P	9
Number of training data samples	5000
Number of data samples for calculating MSE	10000
Maximum doppler frequency f_{dm}	120 Hz
Data rate f_b	4ksymbols/s
Sampling rate f_s	4kHz
SNR	10dB

Table 8.5: Parameters for prediction performances of both GEFK and DEKF with the split activation function ans full activation

the envelope of the channel. For the specific initialization considered, the MSE performance of the SCRTRL scheme of (8.29) is better than that of the FCRTRL sheme of (8.30), but depending on the specific choise of system parameters, these trends may change.

In Figure 8.15 and Figure 8.16, the cost function value of (8.21) versus the number of pilots used for training is portrayed for the GEKF training scheme of Appendix A.2.2.3 and the DEKF training scheme of Appendix A.2.2.4, respectively, when using the split activation function of (8.15) in conjunction with the parameters of Table 8.5. The comparision of Figure 8.15 and Figure 8.16

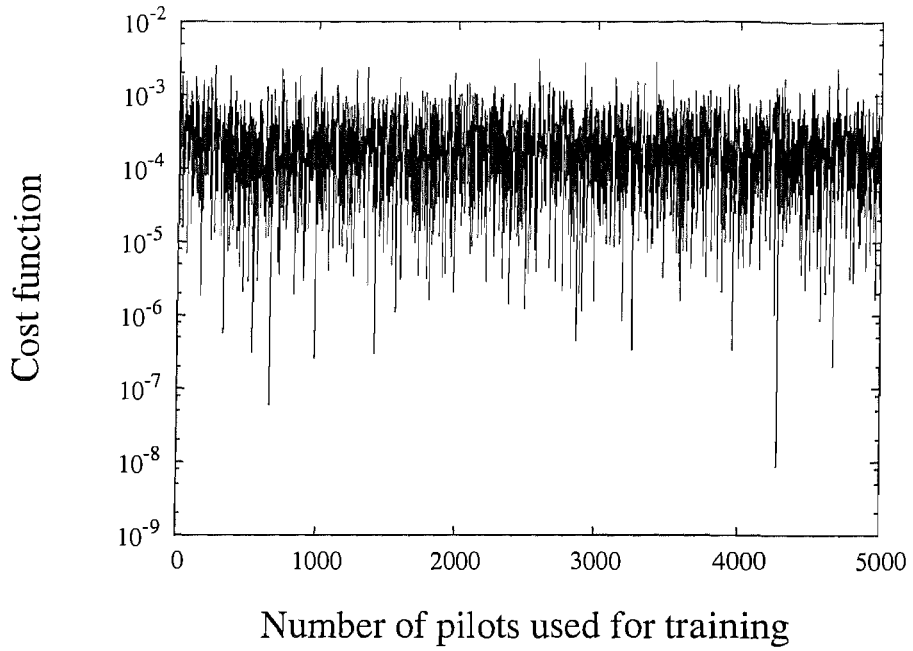


Figure 8.15: The cost function value of (8.21) versus the number of pilots used for training of the GEKF training scheme of Appendix A.2.2.3 when using the split activation function of (8.15). The remaining parameters were summarized in Table 8.5.

shows that the cost function value of (8.21) associated with the DEKF training scheme of Appendix A.2.2.4 is slightly better than that recorded for the GEKF training scheme of Appendix A.2.2.3, when using the split activation function of (8.15). Furthermore, by comparing Figure 8.15 and Figure 8.16 to Figure 8.7, we can see that the the cost function value of (8.21) recorded for both the GEKF training schemes of Appendix A.2.2.3 and the DEKF training scheme of Appendix A.2.2.4 in conjunction with the split activation function of (8.15) converges faster than that for the SCRTRL training scheme of Appendix A.2.1.1.

In Figures 8.17 and 8.18, the cost function value of (8.25) versus the number of pilots used for training is portrayed for the GEKF training scheme of Appendix A.2.2.3 and the DEKF training scheme of Appendix A.2.2.4, respectively, when using the full activation function of Eq.(8.17) in conjunction with the parameters of Table 8.5. Figure 8.18 shows that the cost function value of (8.25) associated with the DEKF training scheme of Appendix A.2.2.4, respectively, is identical to that recorded for the GEKF training scheme of Appendix A.2.2.3 when using the full activation function of (8.17). Furthermore, by comparing Figure 8.17 and Figure 8.18 to Figure 8.8, we can see that the the cost function value of (8.25) recorded for both the GEKF training schemes of Appendix A.2.2.3

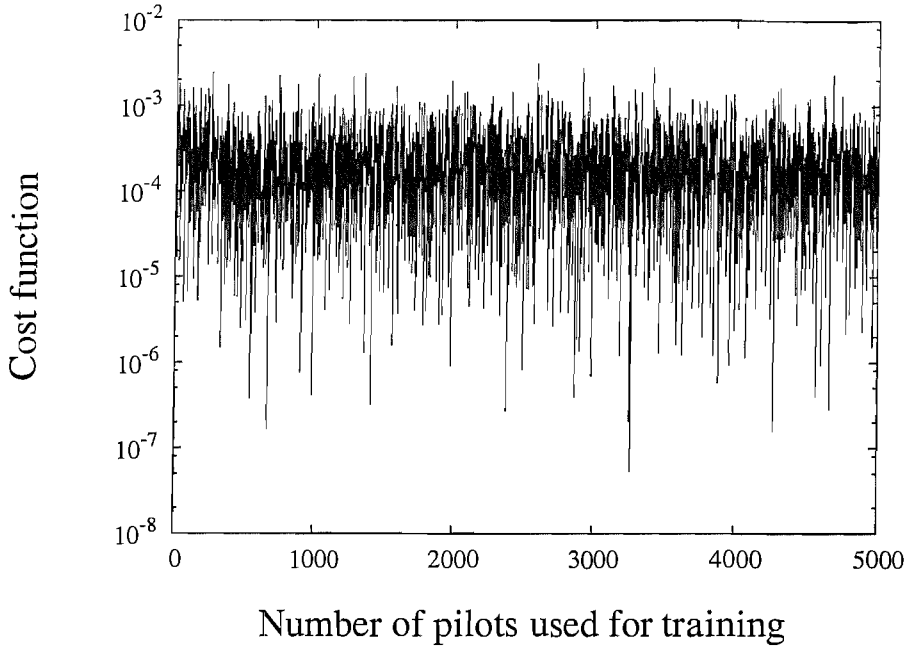


Figure 8.16: The cost function value of (8.21) versus the number of pikots used for training of the DEKF training scheme of Appendix A.2.2.4 when using the split activation function of (8.15). The remaining parameters were summarized in Table 8.5.

and the DEKF training scheme of Appendix A.2.2.4 in conjunction with the full activation function of (8.17) converges faster than that for the FCRTRL training scheme of Appendix A.2.1.2.

In Figures 8.19 to 8.22 the complex-valued channel profile prediction performance of both the GEKF and DEKF training schemes of Appendix A.2.2.3 and Appendix A.2.2.4 in conjunction with the split activation function of (8.15) and the full activation function of (8.17) is compared in terms of the magnitude of the channel profile using the parameters of Table 8.4. In this specific initialization scenario, we had an MSE of 0.055629872 for the GEKF training scheme of Appendix A.2.2.3 and an MSE of 0.055582135 for the DEKF training scheme of Appendix A.2.2.4 in conjunction with the split activation function of (8.15), while an MSE of 0.0560591 for both the GEKF and DEKF schemes of Appendix A.2.2.3 and A.2.2.4, when using the full activation function of (8.17). For the specific initialization considered, the MSE performance of the DEKF training scheme of Appendix A.2.2.4 recorded in conjucntion with the split activation function of (8.15) is the best. Furthermore, the MSE of the GEKF training scheme of Appendix A.2.2.3 employing the split activation function of (8.15) is the second best, while that of both the GEKF and DEKF schemes of Appendix A.2.2.3 and A.2.2.4 using the full activation function of (8.17) is the worst, although depending on the specific system

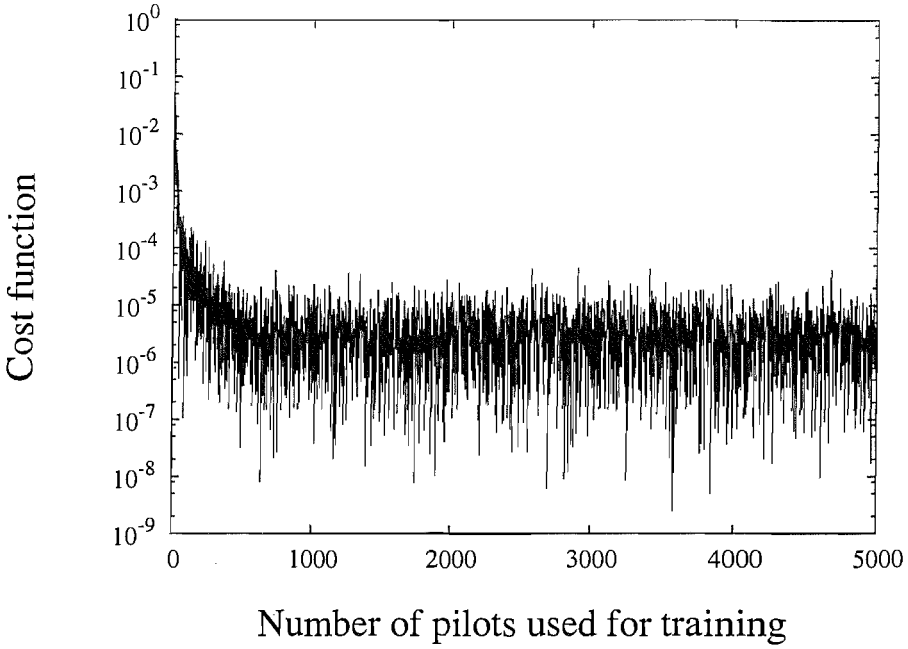


Figure 8.17: The cost function value of (8.25) versus the number of pilots used for training data of the GEKF training scheme of Appendix A.2.2.3 when using the full activation function of (8.17). The remaining parameters were summarized in Table 8.5.

parameters, these trends may change.

8.6.2 Prediction Performance for a Non-Gaussian Channel

In this section the performance of RNNs trained by the RTRL, GEKF and DEKF schemes for a non-Gaussian channel is characterized by simulations and compared to that of a linear MMSE preditor. During the training process, perfect CSI was assumed. For the linear MMSE preditor, the autocorrelation coefficients of both the predicted channel profile and of the prefect profile were estimated for the first 500 received data symbols.

In Figure 8.23 to Figure 8.26, the MSE performance versus the GSNR of (8.28) was recorded, when assuming $\alpha = 1.3$, 1.5, 1.7 and 1.9, respectively. The other parameters were summarized in Table 8.6. From the results of Figure 8.23 we can see that when we have $\alpha = 1.3$, the RNN-based predictor trained by the GEKF and DEKF is capable of achieving an acceptable MSE, provided that the GSNR is sufficient high, while the linear MMSE preditor is inferior. However, when α is increased to 1.5 and 1.7, both the RNN-based predictor trained by the GEKF and DEKF as well as the linear MMSE preditor attains an acceptable MSE upon increasing the GSNR. However, observe in

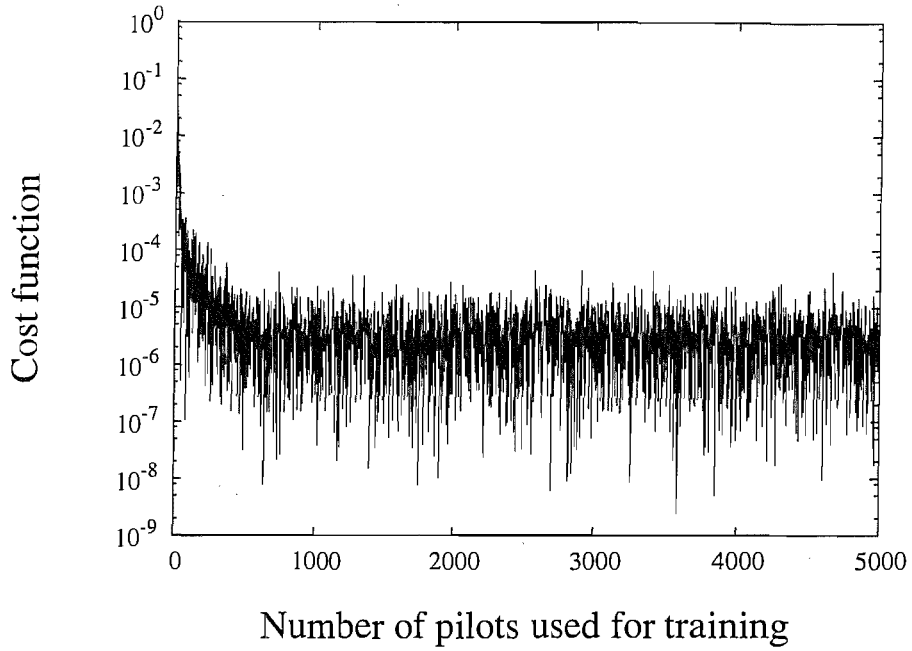


Figure 8.18: The cost function value of (8.25) versus the number of pilots used for training of the DEKF training scheme of Appendix A.2.2.4 when using the full activation function of (8.17). The remaining parameters were summarized in Table 8.5.

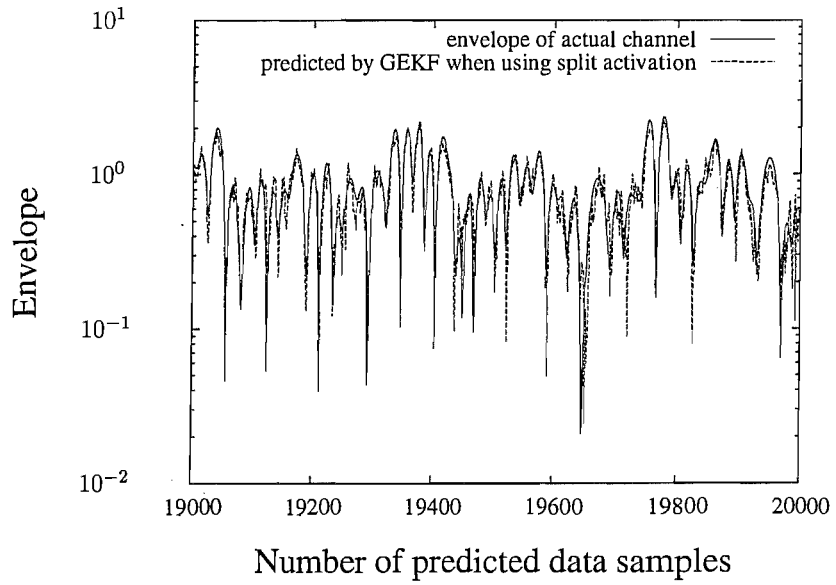


Figure 8.19: Magnitude of the complex-valued fading channel envelope predicted by the GEKF training scheme of Appendix A.2.2.3 versus the number of predicted data when using the split activation function of (8.15). The remaining parameters were summarized in Table 8.4.

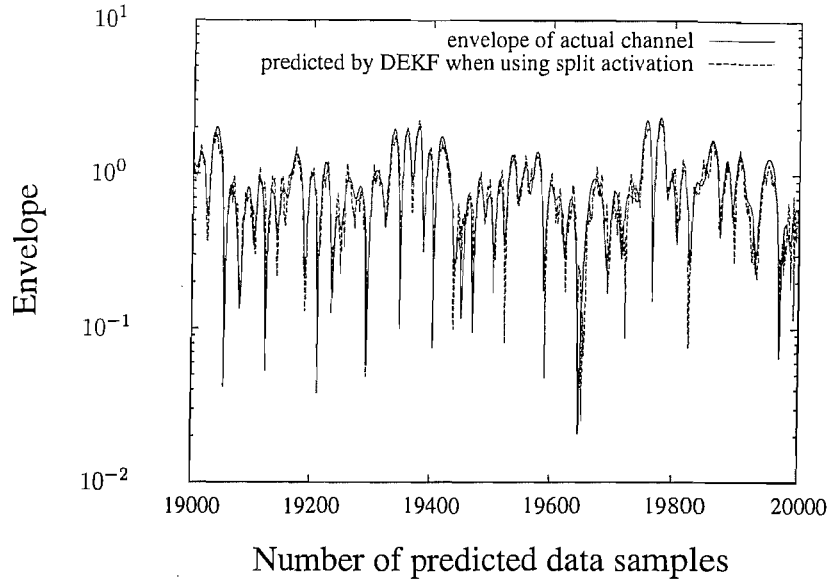


Figure 8.20: Magnitude of the complex-valued fading channel envelope predicted by the GEKF training scheme of Appendix A.2.2.4 versus the number of predicted data when using the split activation function of (8.15). The remaining parameters were summarized in Table 8.4.

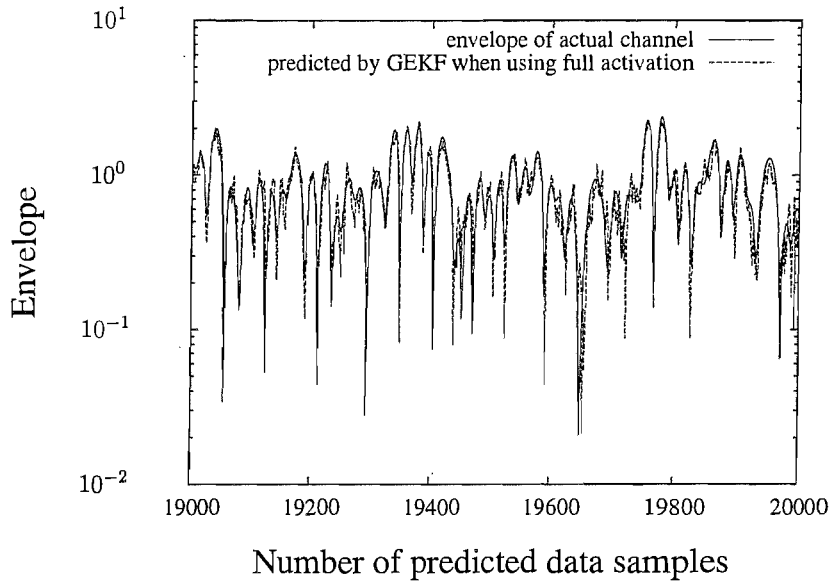


Figure 8.21: Magnitude of the complex-valued fading channel envelope predicted by the GEKF training scheme of Appendix A.2.2.3 versus the number of predicted data when using the full activation function of (8.17). The remaining parameters were summarized in Table 8.4.

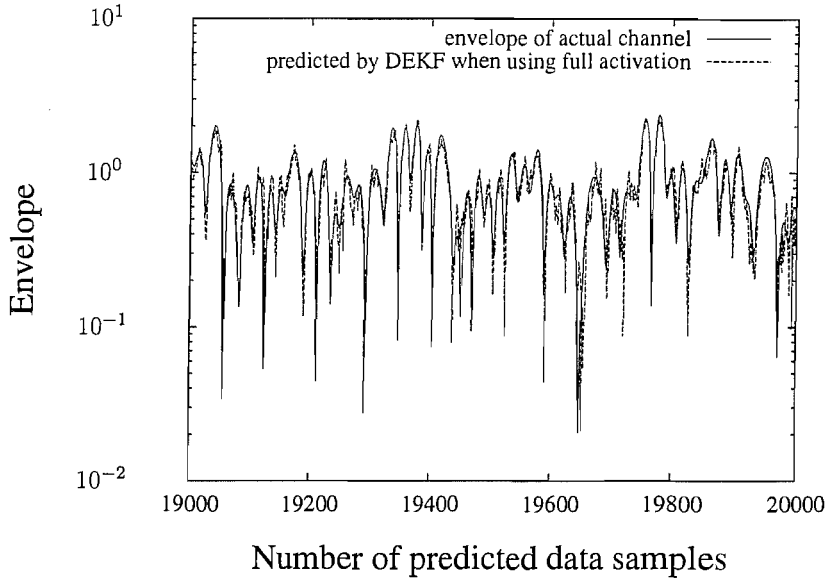


Figure 8.22: Magnitude of the complex-valued fading channel envelope predicted by the GEKF training scheme of Appendix A.2.2.4 versus the number of predicted data when using the full activation function of (8.17). The remaining parameters were summarized in Table 8.4.

Number of activation N	2
Number of external inputs P	10
Total number transmitted data symbols	40000
Number of training data samples	500
Number of data samples for calculating MSE	39500
Maximum doppler frequency f_{dm}	20 Hz
Data rate f_b	4ksymbols/s
Sampling rate f_s	4kHz
Learning rate for RTRLη	0.005
γ	1.0
Number of independent simulations	100

Table 8.6: Parameters used for non-Gaussian channel prediction

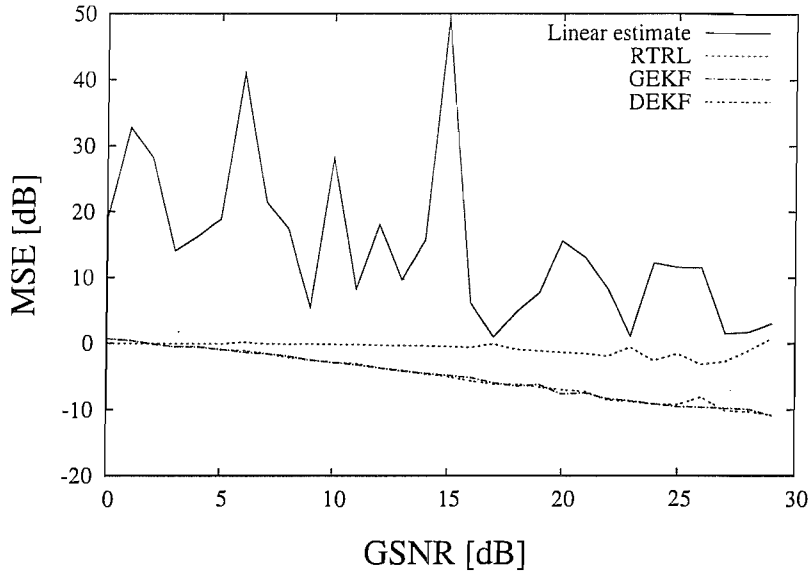


Figure 8.23: The MSE versus GSNR performance for $\alpha = 1.3$. The other parameters were summarized in Table.8.6.

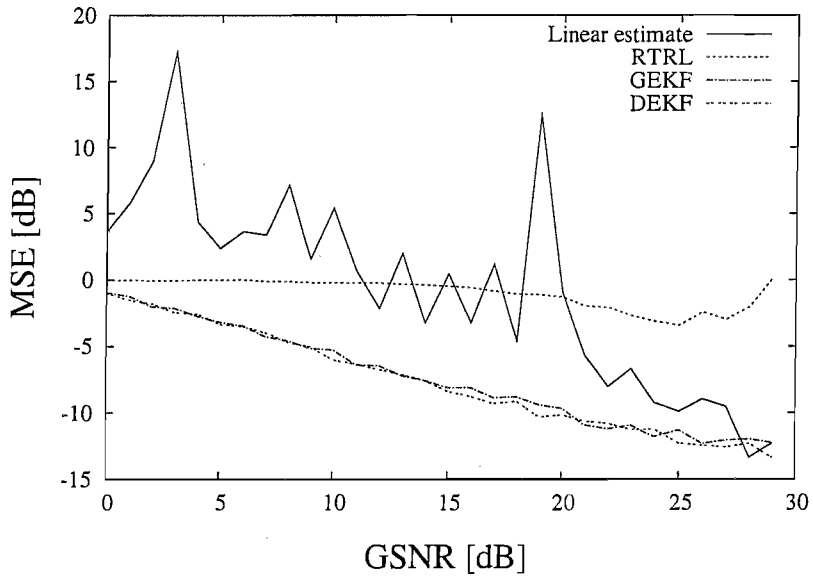


Figure 8.24: The MSE versus GSNR performance for $\alpha = 1.5$. The other parameters were summarized in Table.8.6.

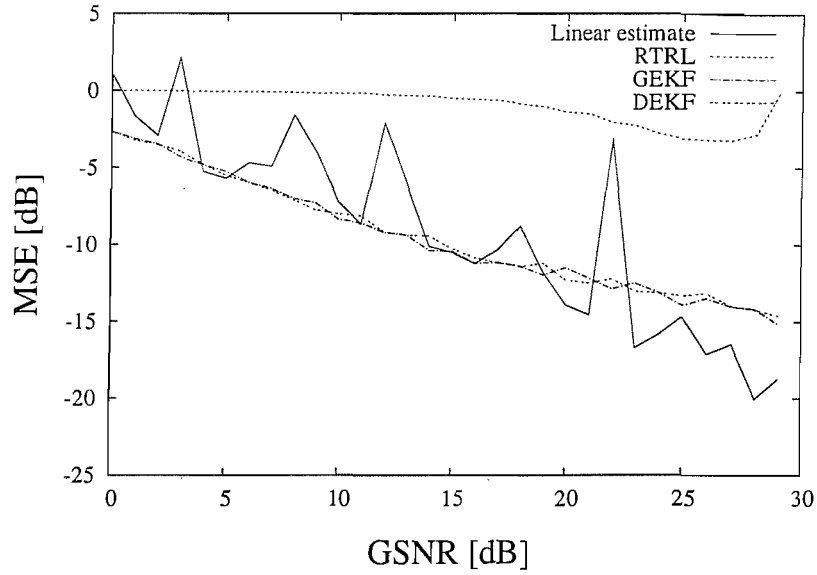


Figure 8.25: The MSE versus GSNR performance for $\alpha = 1.7$. The other parameters were summarized in Table.8.6.

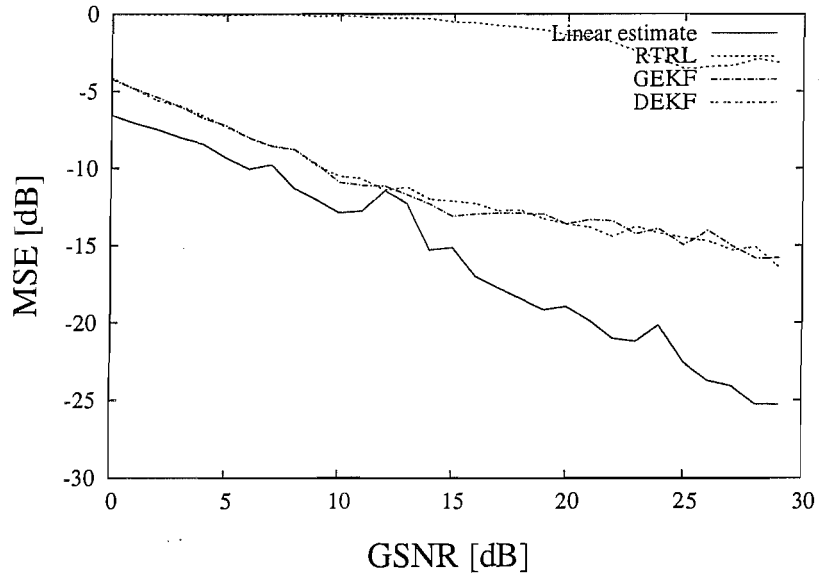


Figure 8.26: The MSE versus GSNR performance for $\alpha = 1.9$. The other parameters were summarized in Table.8.6.

Figure 8.25 and Figure 8.25 that the RNN-based predictor trained by the GEKF and DEKF converges faster than the linear MMSE predictor, provided that the GSNR is moderate. Futhermore, the performance of the RNN-based preditor is more stable than that of the linear MMSE predictor. Finally, when we have $\alpha = 1.9$, the linear MMSE predictor performs better than the RNN-based predictor. We may conclude that when α is sufficiently high, which implies that the impulsive noise is near-Gaussian, the linear MMSE predictor may achieve a better performance. However, when the PDF of the impulsive noise substantially deviates from the PDF of the Gaussian noise, which implies that α deviates from 2.0, the RNN-based predictor trained by the GEKF and DEKF becomes capable of achieving a better performance. Note that in all cases the RNN-based predictor trained by the RTRL fails to achieve an acceptable performance.

8.7 Conclusion

In this chapter, RNN-based narrowband channel predictors were proposed when assuming the presence of both AWGN and non-Guassion noise. Specially, two different activation functions were invoked, namely the split activation function and the full activation function of Section 8.3, respectively. Furthermore, for each of the activation functions, three different training algorithms, namely the RTRL, GEKF and DEKF schemes of Section A.2, were investigated. Again, the application of the RNN-based channel predictor was also extended to non-Gaussian channels. Our simulation results showed that the GEKF and DEKF training schemes converge faster than the RTRL training regime and obtain a better MSE performance in case of the AWGN channel. Furthermore, as far as non-Gaussian channels are concerned, the GEKF and DEKF trained RNN-based channnel predictors converge faster than the linear MMSE predictor and are also more stable.

Chapter 9

Conclusions and Future Work

In this concluding chapter, a summary of the thesis will be presented in Section 9.1. This will be followed by our suggestions for future work in Section 9.2.

9.1 Summary and Conclusions

9.1.1 Chapter 1

In Chapter 1, the importance of channel prediction in wireless communication systems was discussed and an overview of various prediction algorithms was presented. Following this, the contributions of the thesis were highlighted and its organization was provided.

9.1.2 Chapter 2

In Chapter 2, the characteristics of wireless channels were considered. Specifically, the two main factors of characterizing a wireless channel, namely the multipath propagation phenomenon and the Doppler frequency shift were introduced in Section 2.2. Following this, the classification of wireless channels into frequency-nonselective and frequency-selective subclasses or into fast-fading or slow-fading categories was discussed in Section 2.3. In Section 2.4 statistical models, such as as Rayleigh, Rician or Nakagami models were provided. The simulation of Rayleigh fading channels using IDFT, Jakes' model and an AR model was presented in Section 2.5.

SNR γ_b	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Effective SNR $\gamma_{(e,n)}$	16.4	17.4	18.4	19.4	20.3	21.3	22.3	23.3	24.3	25.3	26.3

Table 9.1: Effective SNR of (4.69) after frequency domain estimation

9.1.3 Chapter 3

In Chapter 3, the LRP philosophy was investigated in the context of narrowband wireless channels. Specifically, in Section 3.2 it was shown that a more accurate prediction can be obtained, despite using a lower sampling rate than the data rate based on MMSE LRP. Motivated by this discovery, Kalman filtering assisted LRP was investigated in Section 3.3.

9.1.4 Chapter 4

In Chapter 4 the channel estimation and prediction processes invoked for a single-carrier system encountering wideband propagation conditions were considered. The characterization of dispersive wideband channel was porvided in Section 4.2 and the single-carrier transmitter was discussed in Section 4.3. Two-dimensional channel estimation employed in the FD and TD was investigated in Section 4.4, while Kalman filtering assisted LRP was proposed for a single-carrier system in Section 4.5, when communicating over wideband channels. The simulation results provided in Section 4.6 showed the benefits of channel estimation, which was repeated for the reader's convenience in Figure 9.1. We can see from Figure 9.1 that the effective SNR was significantly improved with the aid of channel estimation, as summarized in Table 9.1, where an approximately 16.3 dB SNR gain can be observed after frequency domain estimation.

9.1.5 Chapter 5

In Chapter 5, subspace tracking based MIMO eigenmode transmit preprocessing was investigated in the context of a single user. MIMO eigenmode transmission requires the knowledge of the right-hand-side singular eigen vectors of the MIMO channel matrix seen in (5.3) at the transmitter side and that if the left-hand-side singular eigen vectors of the MIMO channel matrix at the receiver side, as shown in Figure 9.2. In order to acquire these singular eigen vectors, traditionally, channel estimation is required at the receiver side, followed by the SVD of the estimated channel matrix, as shown in Figure 9.3. Alternatively, a subspace tracking algorithm can be invoked for generating the required singular eigen vectors without performing channel estimation followed by its SVD, as shown in Figure 9.3, which reduces the computationaly complexity imposed. Specifically, the PASTD subspace

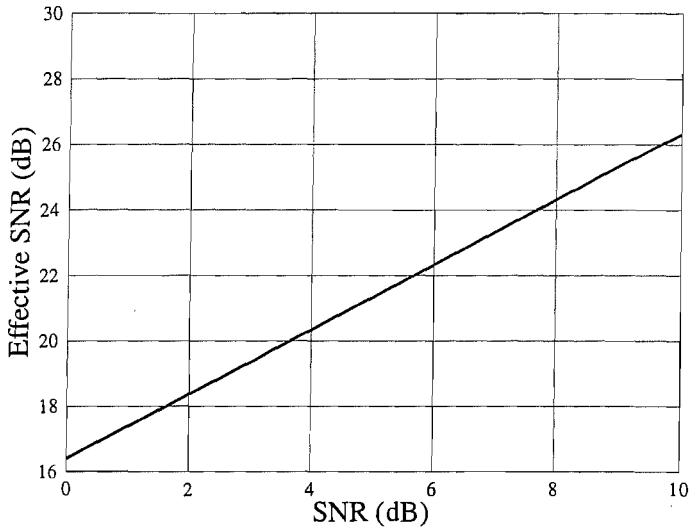


Figure 9.1: The effective SNR of Eq.(4.69) for LRP after CTF estimation in the frequency domain per symbol versus average SNR of Eq.(4.66) per symbol. The remaining parameters are summarized in Table 4.1.

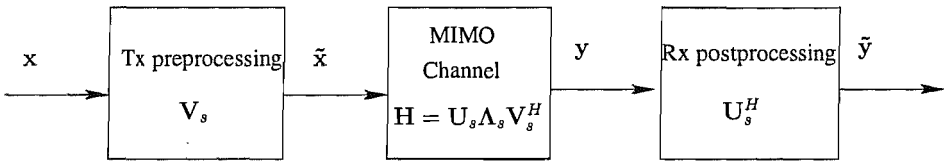


Figure 9.2: Schematic of MIMO eigenmode transmission.

tracking algorithm was invoked in the context of a TDD MIMO system in Chapter 5. Furthermore, the phase ambiguity imposed by the SVD of the channel's CIR matrix was resolved by the employment of DPSK modulation. The simulations characterize the achievable performance with the aid of PASTD subspace tracking.

9.1.6 Chapter 6

In Chapter 6, both zero forcing and MMSE based transmit preprocessing techniques were considered in the context of a MIMO aided multiple user scenario. Speficically, the transmit diversity gain and spatial multiplexing gain achieved by these transmit preprocessing techniques was investigated in the

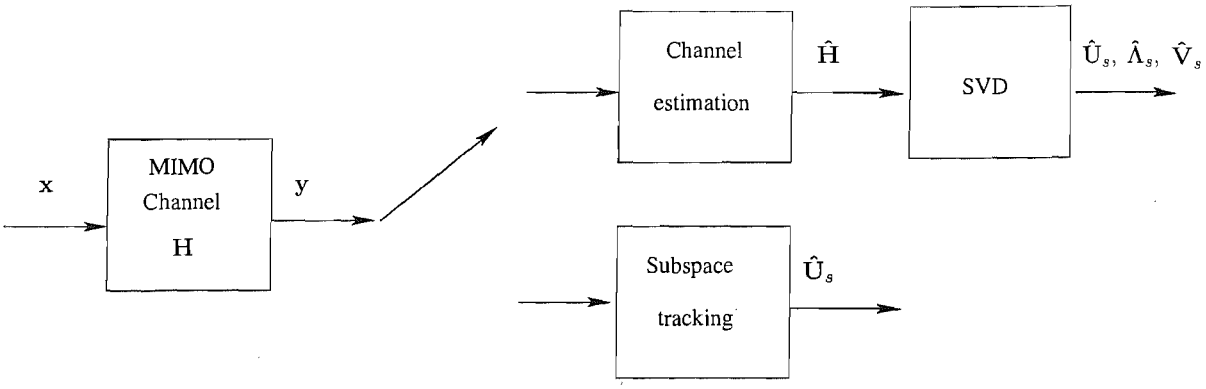


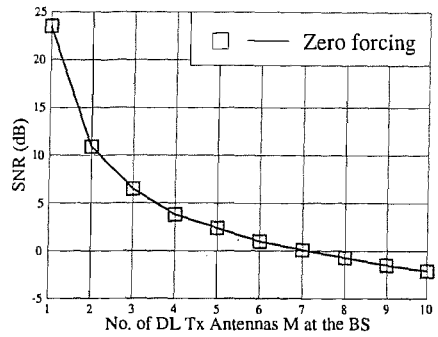
Figure 9.3: Schematic of obtaining the preprocessing and postprocessing.

	Scheme 1 [dB]	Scheme 2 [dB]		Scheme 3 [dB]		Scheme 4 [dB]	
Throughput	1 bit	2 bits		3 bits		4 bits	
	Zero forcing	Zero forcing	MMSE	Zero forcing	MMSE	Zero forcing	MMSE
From	Fig.6.6	Fig.6.7	Fig.6.10	Fig.6.8	Fig.6.11	Fig.6.9	Fig.6.12
M=1	23.5						
M=2	10.9	24.4	20.6				
M=3	6.5	10.7	9.1	23.8	16.2		
M=4	3.8	6.0	5.1	10.3	7.6	23.8	14.1
M=5	2.4	3.5	2.9	5.7	4.6	10.3	6.6
M=6	1.0	1.7	1.5	3.2	2.5	5.6	3.7
M=7	0.1	0.6	0.3	1.6	1.2	3.2	2.1
M=8	-0.7	-0.3	-0.4	0.6	0.1	1.6	0.7
M=9	-1.5	-0.9	-1.2	-0.3	-0.6	0.4	-0.2
M=10	-2.1	-1.5	-1.8	-1.0	-1.3	-0.6	-1.0

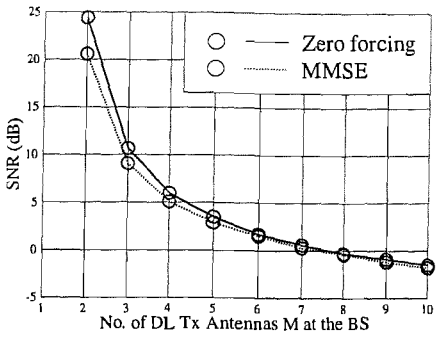
Table 9.2: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.2

		Scheme 1 [dB]		Scheme 2 [dB]		Scheme 3 [dB]		Scheme 4 [dB]	
		Zero forcing	MMSE	Zero forcing	MMSE	Zero forcing	MMSE	Zero forcing	MMSE
	Throughput	Fig.6.16	Fig.6.20	Fig.6.17	Fig.6.21	Fig.6.18	Fig.6.22	Fig.6.19	Fig.6.23
K=1	1 bits	0.1	0.1	-0.7	-0.7	-1.3	-1.3	-2.1	-2.1
K=2	2 bits	0.6	0.4	-0.3	-0.4	-1.0	-1.2	-1.5	-1.8
K=3	3 bits	1.8	1.2	0.6	0.3	-0.3	-0.7	-1.0	-1.3
K=4	4 bits	3.1	2.1	1.5	0.7	0.3	-0.3	-0.6	-1.0
K=5	5 bits	5.0	3.2	3.0	1.9	1.5	0.4	0.6	-0.4
K=6	6 bits	10.6	5.6	5.6	2.9	2.9	1.5	1.5	0.4
K=7	7 bits	25.7	11.5	10.3	5.6	5.6	3.2	2.9	1.8
K=8	8 bits			24.7	9.0	9.7	4.7	5.4	2.6
K=9	9 bits					24.4	7.8	10.0	4.7
K=10	10 bits							24.1	7.6

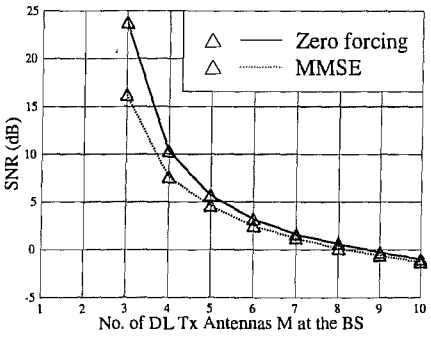
Table 9.3: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.5



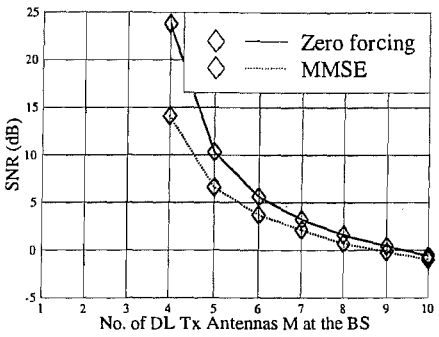
(a) Scheme 1



(b) Scheme 2



(c) Scheme 3



(d) Scheme 4

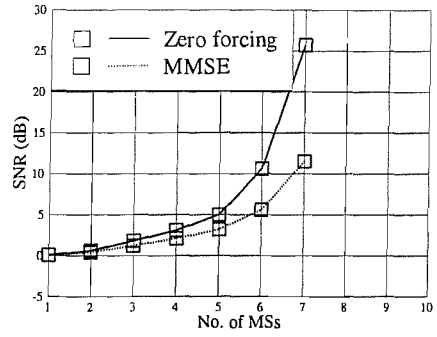
Figure 9.4: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.2

context of BPSK modulation and was summarized in Tables 9.2 and 9.3, which are also plotted in Figures 9.4 and 9.5. As we can see from Figures 9.4 and 9.5, MMSE preprocessing required a lower SNR for attaining the same BER performance than zero forcing preprocessing. This is due to the fact that MMSE preprocessing takes into accoount the effects of both the MUI and of noise jointly, while zero forcing only considers those of the MUI.

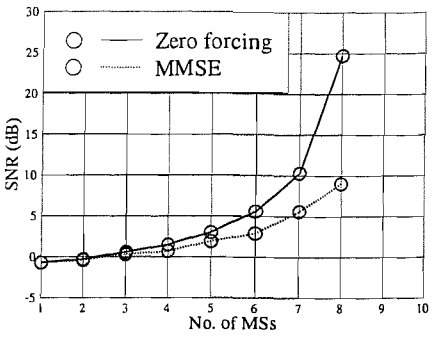
Additionally, Kalman filtering assisted channel prediction was invoked for supporting the zero forcing and MMSE based transmit preprocessing techniques in the context of a TDD system.

9.1.7 Chapter 7

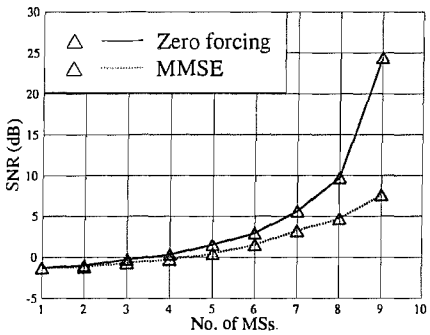
In Chapter 7, SVD based transmit preprocessing and postprocessing was proposed in the context of MIMO aided multiple users for both UL and DL transmissions. Furthermore, based on the proposed



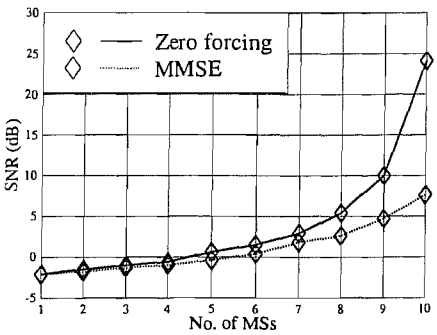
(a) Scheme 1



(b) Scheme 2



(c) Scheme 3



(d) Scheme 4

Figure 9.5: SNR [dB] required for maintaining BER=10^{-3} for both zero forcing and MMSE preprocessing for Schemes 1-4 of Table 6.5

algorithm, both the maximum information rate and the maximum SNR based power allocation policies of [64, 185, 221] were considered for uplink transmisson and both the attainable capacity as well as the BER performance was compared for both the uncorrelated and correlated noise scenarios. The simulation results of Figures 7.3 to 7.10 suggest that a performance loss is experienced due to the noise enhancement experienced in the correlated noise scenarios. For downlink transmission, the joint and the individual maximum information rate based power allocation policies of [64] were considered first. The simulation results of Figures 7.11 to 7.14 suggest that both of them achieve a similar performance, while and the power constraints of (7.74) and (7.79) have a significant influence on the attainable system performance. Furthermore, the maximum SNR and the equal-power allocation policies of [151, 221] were compared for downlink transmissions as well. The simulation results of Figures 7.15 to 7.17 suggest that the BER performance corresponding to the maximum SNR policy

is better than that of the equal-power policy, since the equal-power policy ingores the fact that the lower eigenvalues of the channel matrix result in a lower effective channel gain and simply allocate an equal power to each transmitted symbol. Hence in the context of the equal-power policy some of the lower-SNR data streams will degrade the overall BER performance, while the maximum-SNR policy is capable of allocating more power to the specific data streams having lower eigenvalues in order to achieve an improved balance among all data streams. The TDD mode may exploit that the UL and DL channels are similar and invoke this knowledge for downlink transmit preprocessing, as discussed in Section 7.4.

9.1.8 Chapter 8

In Chapter 8, nonlinear channel prediction using RNNs was considered. Speficically, the structure of RNN-based channel predictor was considered in Section 8.2, while two types of RNN activation functions, namely, the split-and the full activation function were presented in Section 8.3. Furthermore, three different training algorithms, namely, the RTRL, the GEKF and the DEKF techniques were introduced in Section 8.4, respectively. The application of RNN-based channel predictors was discussed in Section 8.5 in the context of non-Gaussian channels. Finally, our simulation results provided in Section 8.6 showed that the RNN-based channel predictor was capable of fulfilling the task of channel prediction and that the GEFK and DEKF training schemes exhibited a rapid convergence. Furthermore, when the dominant source of impariment was impulsive noise, rather than Gaussian noise, the RNN-based predictor achieved a better performance than the linear MMSE based predictor.

9.2 Future Work

Due to the emerging high demand for supporting novel multimedia applications, next generation wireless systems are expected to support high data rates. When employing multiple antennas at both the transmitter and receiver, multiple input multiple output (MIMO) systems have the potential of achieving a higher transmission rate or higher intergrity than their traditional single input single output (SISO) counterparts [60]. Accordingly, the investigation of application channel prediction in the context of MIMO systems constitutes our main research direction in the future.

9.2.1 Long-Range Prediction Based Scheduling

Since the wireless channel imposes adverse effects on wireless communications, amongst others techniques, antenna diversity has been employed for combatting its effects. However, instead of mitigating it, in future generation of wireless communication systems, we may opt for circumventing its effects with the aid of the so-called multiuser-diversity based scheduling [222, 223]. Numerous scheduling strategies rely on the availability of full or partial CSI in order to schedule the transmissions of specific users during the next time slot, using for example maximum SNR-based scheduling [224] or proportional fair scheduling [225]. Hence it is intitutively reasonable to employ channel prediction instead of channel estimation in the context of scheduling techniques [226], and hence channel prediction assisted scheduling will be one of out future reseach topics.

9.2.2 Long-Range Prediction Based Transmitter Antenna Selection

One of the drawbacks of MIMO systems is their increased complexity, and cost imposed by the employment of mutliple RF chains of the multiple antennas used at both the transmitter and receiver.

For this reason, there is now great interest in so-called antenna selection schemes, which aim for reducing the complexity of MIMO systems by activating either a limited subset of transmitter or receiver antennas or a combination of subsets of transmitter and receiver antennas (transmitter/receiver antenna selection) for communicating without compromising the achievable the performance [227, 228]. Specifically, when transmitter antenna selection is invoked, the CSI corresponding to next transmission instant is required at the transmitter. When using channel prediciton instead of channel estimation, transmitter antenna selection may achieve a better performance [229]. Hence, channel prediciton assisted transmitter antenna selection will constitute another future research field.

Bibliography

[1] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, Inc, fourth ed., 2002.

[2] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering: With Matlab Exercises And Solutions*. Wiley, third ed., 1997.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing:Estimation Theory*. New Jersey: Prentice Hall, Inc, 1993.

[4] L. Hanzo, C. H. Wong, and M. S. Yee, *Adaptive wireless transceivers: turbo-coded, turbo-equalized and space-time coded TDMA, CDMA, and OFDM Systems*. John Wiley & Sons, 2002.

[5] L. Hanzo, M. Münster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. John Wiley & Sons, 2003.

[6] L. Hanzo, S. X. Ng, W. T. Webb, and T. Keller, *Quadrature amplitude modulation: from basics to adaptive trellis-coded, turbo-equalised and space-time doded OFDM, CDMA and MC-CDMA Systems*. John Wiley, second ed., 2004.

[7] J. M. Torrance and L. Hanzo, "Upper bound performance of adaptive modulation in a slow Rayleigh fading channel," *Electronics Letters*, vol. 32, pp. 718–719, April 1996.

[8] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, pp. 1561 – 1571, September 2001.

[9] Z. Zhou and B. Vucetic, "MIMO systems with adaptive modulation," *IEEE Transactions on Vehicular Technology*, vol. 54, pp. 1828 – 1842, September 2005.

[10] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1285 – 1294, July 2004.

[11] M. Mohammad and R. M. Buehrer, "On the impact of SNR estimation error on adaptive modulation," *IEEE Communications Letters*, vol. 9, pp. 490–492, June 2005.

[12] J. F. Paris, M. C. Aguayo-Torres, and J. T. Entrambasaguas, "Impact of channel estimation error on adaptive modulation performance in flat fading," *IEEE Transactions on Communications*, vol. 52, pp. 716 – 720, May 2004.

[13] A. J. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, pp. 1218 – 1230, October 1997.

[14] S. Falahati, A. Svensson, T. Ekman, and M. Sternad, "Adaptive modulation systems for predicted wireless channels," *IEEE Transactions on Communications*, vol. 52, pp. 307 – 316, February 2004.

[15] G. E. Oien, H. Holm, and K. J. Hole, "Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 758 – 769, May 2004.

[16] X. Cai and G. B. Giannakis, "Adaptive PSAM accounting for channel estimation and prediction errors," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 246– 256, January 2005.

[17] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[18] J. K. Huang and J. H. Winters, "Sinusoidal modeling and prediction of fast fading processes," in *Global Telecommunications Conference*, vol. 2, (Sydney, NSW, Australia), pp. 892 – 897, 8 - 12 November 1998.

[19] J. B. Andersen, J. Jensen, S. H. Jensen, and F. Frederiksen, "Prediction of future fading based on past measurements," in *Proceedings of IEEE Vehicular Technology Conference, 1999, Fall.*, vol. 1, (Amsterdam, Netherlands), pp. 151 – 155, 19 - 22 September 1999.

[20] L. Dong, G. Xu, and H. Ling, "Prediction of fast fading mobile radio channels in wideband communication systems," in *Global Telecommunications Conference, 2001*, vol. 6, (San Antonio, TX, USA), pp. 3287 – 3291, 25 - 29 November 2001.

[21] S. Semmelrodt and R. Kattenbach, "Application of spectral estimation techniques to 2-D fading forecast of time-variant channels," in *COST 273 TD(01)034.*, (Bologna, Italy), 15 - 17 October 2001.

[22] S. Semmelrodt and R. Kattenbach, "A 2-D fading forecast of time-variant channels based on parametric modeling techniques," in *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2002.*, vol. 4, (Lisbon, Portugal), pp. 1640 – 1644, 15 - 18 September 2002.

[23] M. Chen and M. Viberg, "LMMSE channel prediction based on sinusoidal modeling," in *IEEE Sensor Array and Multichannel Signal Processing Workshop Proceedings*, (Sitges, Barcelona, Spain), pp. 377 – 381, 18 - 21 July 2004.

[24] I. C. Wong and B. L. Evans, "Joint channel estimation and prediction for OFDM systems," in *IEEE Global Telecommunications Conference*, (St. Louis, Missouri USA), pp. 2255 – 2259, 28 November - 2 December 2005.

[25] J. Liu and X. Liu, "Time-varying channel identification and prediction in OFDM systems using 2-D frequency estimation," in *Military Communications Conference*, (Washington, DC, USA), pp. 1 – 7, 23 - 25 October 2006.

[26] A. J. Barabell, "Improve the resolution performance of eigenstructure-based direction-finding algorithms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Boston, MA, USA), pp. 336–339, 14 - 16,April 1983.

[27] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters by rotational invariance techniques," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 37, pp. 984–995, July 1989.

[28] A. Forenza, *Exploitation of the fixed wireless channel through the link-adaptation algorithm and the channel prediction.* Master thesis, Instituts Eurocom, 2001.

[29] A. Duel-Hallen, S. Hu, and H. Hallen, "Long range prediction of fading signals: enabling adaptive transmission for mobile radio channels," *IEEE Signal Processing Magzine*, vol. 17, pp. 62–75, May 2000.

[30] M. Chen, M. Viberg, and T. Ekman, "Two new approaches to channel prediction based on sinusoidal modelling," in *IEEE/SP 13th Workshop on Statistical Signal Processing*, (Telecom Paris, Paris), pp. 697– 700, 17 - 20 July 2005.

[31] M. Chen, T. Ekman, and M. Viberg, "New approaches for channel prediction based on sinusoidal modeling," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 49393, 2007.

[32] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third-Generation Cellular and WATM Systems OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. John Wiley & Sons, 1999.

[33] T. Eyceoz, A. Duel-Hallen, and H. Hallen, "Deterministic channel modeling and long range prediction of fast fading mobile radio channels," *IEEE Communications Letters*, vol. 2, pp. 254–256, September 1998.

[34] T. Eyceoz, S. Hu, and A. Duel-Hallen, "Performance analysis of long range prediction for fast fading channels," in *Proceedings of the 33rd Annual Conference on Information Sciences & Systems (CISS'99)*, (Baltimore, MD, USA), pp. 656–661, 17 - 19 March 1999.

[35] T. Eyceoz, A. Duel-Hallen, and H. Hallen, "Prediction of fast fading farameters by resolving the interference pattern," in *Proceedings of the 31st ASILOMAR Conference on Signals, System, & Compurters*, (Pacific Grove, CA, USA), pp. 167–171, 2 - 5 November 1997.

[36] I. C. Wong, A. Forenza, R. W. Heath, and B. L. Evans, "Long range channel prediction for adaptive OFDM systems," in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), pp. 732 – 736, 7 - 10 November 2004.

[37] Z. Luo, H. Gao, Y. Liu, and J. Gao, "Robust pilot-symbol-aided MIMO channel estimation and prediction," in *IEEE Global Telecommunications Conference*, (Dallas, Texas, USA), pp. 3646 – 3650, 29 November - 3 December 2004.

[38] J. Akhtman and L. Hanzo, "Low-complexity channel estimation for OFDM and MC-CDMA," in *IEEE 59th Vehicular Technology Conference, Spring*, (Milan, Italy), pp. 1134 – 1138, 17 - 19 May 2004.

[39] Y. Chen, H. Tsai, and J. Peng, "Effects of adaptive prediction algorithms on adaptive QAM in flat Rayleigh fading channels," in *IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, (Beijing, China), pp. 2446 – 2451, 7 - 10 September 2003.

[40] D. Schafhuber and G. Matz, "MMSE and adaptive prediction of time-varying channels for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 593 – 602, March 2005.

[41] G. Xie, R. Zhang, and Y. Liu, "Simplified and adaptive prediction algorithm of time-varying wideband channels," in *International Symposium on Communications and Information Technologies*, (Bangkok, Thailand), pp. 264 – 267, 18 - 20 October 2006.

[42] J. Wen, C. Chang, G. Lee, and C. Huang, "OFDM channel prediction using fuzzy update LMS algorithm in time-variant mobile channel," in *IEEE 64th Vehicular Technology Conference Fall*, (Montreal, Canada), pp. 1 – 5, 15 - 28 September 2006.

[43] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Transactions on Communications*, vol. 46, pp. 902 – 915, July 1998.

[44] S. Semmelrodt and R. Kattenbach, "Investigation of different fading forecast schemes for flat fading radio channels," in *VTC 2003-Fall.*, (Orlando, Florida, USA), pp. 149 – 153, 6 - 9 October 2003.

[45] W. R. Braun and U. Dersch, "A physical mobile radio channel model," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 472–482, May 1991.

[46] S. Bug and R. Jakoby, "Modeling the mobile channel using theory of dynamics-first derivations and results," in *Proceedings of the European Wireless Conference*, (Barcelona, Spain), pp. 402–407, 24 - 27 February 2004.

[47] S. Bug and R. Jakoby, "Modeling the mobile radio channel using theory of dynamics - reconstruction of dynamics by differential dquations," in *Proceedings of VTC 2004-Fall*, (Los Angeles, USA), pp. 28–32, 26 - 29 September 2004.

[48] S. Bug, A. Nazarov, K. Kastell, and R. Jakoby, "Characterizing the mobile radio channel by a conservative dynamical system," in *VTC 2006-Spring*, (Melbourne, Australia), pp. 2772 – 2776, 7 - 10 May 2006.

[49] T. Ekman and G. Kubin, "Nonlinear prediction of mobile radio channels: measurements and MARS model designs," in *ICASSP'99*, (Phoenix, Arizona, USA), pp. 2667 – 2670, 15 - 19 March 1999.

[50] J. Sun, T. Zhang, and F. Liu, "Nonlinear prediction of fast fading channel parameters based on the chaotic attractor," in *Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication*, (Shanghai, China), pp. 181 – 184, 31 May - 2 June 2004.

[51] J. Sun, T. Zhang, and F. Liu, "Nonlinear prediction of mobile-radio fading channel using recurrent least squares support vector machines and embedding phase space," in *International Conference on Communications, Circuits and Systems*, (Chengdu, China), pp. 282 – 286, 27 - 29 June 2004.

[52] X. M. Gao, J. M. A. Tanskanen, and S. J. Ovaska, "Comparison of linear and neural network-based power prediction schemes for mobile DS/CDMA systems," in *Proceedings of IEEE Vehicular Technology Conference, 1996.*, (Atlanta, Georgia, USA), pp. 61 – 65, 28 April - 1 May 1996.

[53] X. M. Gao, X. Z. Gao, J. M. A. Tanskanen, and S. J. Ovaska, "Power prediction in mobile communication systems using an optimal neural-network structure," *IEEE Transactions on Neural Networks*, vol. 8, pp. 1446 – 1455, November 1997.

[54] B. Visweswaran and T. Kiran, "Channel prediction based power control in W-CDMA systems," in *First International Conference on 3G Mobile Communication Technologies*, (London, UK), pp. 41 – 45, 27 - 29 March 2000.

[55] J. Klaue and A. Aguiar, "Robust real-time channel prediction based on inaccurate instantaneous measurements: an approach," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, (New York City, USA), pp. 530 – 534, 5 - 8 June 2005.

[56] P. M. Castro, L. Castro, and J. Miguez, "Precoding in wireless communications systems using particle filtering for blind channel prediction," in *IEEE 5th Workshop on Signal Processing Advances in Wireless Communications*, (Lisboa, Portugal), pp. 97 – 101, 11 - 14 July 2004.

[57] T. K. Chee, C. Lim, and J. Choi, "Channel prediction using lumpable finite-state Markov channels in OFDMA systems," in *IEEE 63rd Vehicular Technology Conference Spring*, (Melbourne, Australia), pp. 1560 – 1564, 7 - 10 May 2006.

[58] T. Zemen, C. F. Mechlenbrauker, and B. H. Fleury, "Time-variant channel prediction using time-concentrated and band-limited sequences," in *IEEE International Conference on Communications*, (Istanbul, Turkey), pp. 5660 – 5665, 11 - 15 June 2006.

[59] Y. Lee, "Channel prediction with cascade AR modeling," in *Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services*, (Guadeloupe, French Caribbean), pp. 40 – 40, 19 - 25 February 2006.

[60] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bölcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, pp. 198– 218, February 2004.

[61] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 281– 302, April 2003.

[62] S. N. Diggavi, N. Al-dhahir, A. Stamoulis, and A. R. Calderbank, "Great expectations: the value of spatial diversity in wireless networks," *Proceedings of the IEEE*, vol. 92, pp. 219–270, February 2004.

[63] R. Irmer, *Multiuser transmission in code division multiple access mobile communications systems*. Ph.D. thesis, Technische Universitat Dresden, 2005.

[64] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, May 1999.

[65] W. Liu, S. X. Ng, L. L. Yang, and L. Hanzo, "Joint channel prediction aided differentially encoded TTCM and BICM-ID assisted eigen-beamforming," *IEE Electronics Letters*, vol. 43, pp. 232 – 234, February 2007.

[66] S. X. Ng, W. Liu, J. Wang, M. Tao, L. L. Yang, and L. Hanzo, "Performance analysis of iteratively decoded variable-length space-time coded modulation," in *IEEE International Conference on Communications*, (Glasgow, Scotland, UK), p. CDROM, 24 - 28 June 2007.

[67] S. X. Ng, W. Liu, L. L. Yang, and L. Hanzo, "Channel prediction aided coded modulation assisted eigen-beamforming," in *Proceedings of the IEEE Vehicular Technology Conference*, (Dublin, Ireland), pp. 1742 – 1746, 23 - 25 April 2007.

[68] A. Ahreans, W. Liu, S. X. Ng, V. Kuehn, L. L. Yang, and L. Hanzo, "SVD-aided, iteratively detected spatial division multiplexing using long-range channel prediction," *submitted to IEEE Workshop on Signal Processing Systems, (Shanghai, China)*, 17 - 19 October 2007.

[69] W. Liu, L. L. Yang, and L. Hanzo, "Wideband channel estimation and prediction in single-carrier wireless systems," in *Proceedings of the IEEE Vehicular Technology Conference*, (Stockholm, Sweden), pp. 543 – 547, 30 May - 1 June 2005.

[70] B. Hu, W. Liu, L. L. Yang, and L. Hanzo, "Multiuser decorrelating based long-range frequency-domain channel transfer function prediction in multicarrier DS-CDMA systems," in *Proceedings of IEEE International Symposium on Spread Spectrum Techniques and Applications*, (Manaus, Amazon, Brazil), pp. 163 – 167, 28 - 31 August 2006.

[71] W. Liu, L. L. Yang, and L. Hanzo, "Subspace tracking based blind MIMO transmit preprocessing," in *Proceedings of the IEEE Vehicular Technology Conference 2007 Spring*, (Dublin, Ireland), pp. 2228 – 2232, 23 - 25 April 2007.

[72] W. Liu, L. L. Yang, and L. Hanzo, "Channel prediction aided multiuser transmission in SDMA," *submitted to IEEE Workshop on Signal Processing Systems, (Shanghai, China)*, 17 - 19 October 2007.

[73] W. Liu, L. L. Yang, and L. Hanzo, "Recurrent neural network based narrowband channel prediction," in *Proceedings of the IEEE Vehicular Technology Conference*, (Melbourne, Australia), pp. 2173 – 2177, 7 - 10 May 2006.

[74] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 686 – 693, November 1991.

[75] T. Ekman, M. Sternad, and G. Kubin, "Unbiased power prediction of rayleigh fading channels," in *IEEE 56th Vehicular Technology Conference*, (Vancouver, Canada), pp. 280 – 284, 24 - 29 September 2002.

[76] W. Cao and W. Wang, "A frequency-domain channel prediction algorithm in wideband wireless communication systems," in *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications.*, (Barcelona, Spain), pp. 2402 – 2405, 5 - 8 September 2004.

[77] K. E. Baddour and N. C. Beaulieu, "Improved pilot-assisted prediction of unknown time-selective Rayleigh channels," in *IEEE International Conference on Communications*, (Istanbul, Turkey), pp. 5192 – 5199, 11 - 15 June 2006.

[78] K. E. Baddour and N. C. Beaulieu, "Autoregressive models for fading channel simulation," in *Global Telecommunications Conference*, (San Antonio, TX, USA), pp. 1187 – 1192, 25 - 29 November 2001.

[79] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 1650 – 1662, July 2005.

[80] G. L. Stüber, *Principles of Mobile Communication.* Norwell, MA, USA: Kluwer Academic Publishers, 1996.

[81] J. G. Proakis, *Digital Communications.* McGraw Hill, 4th ed., 2000.

[82] J. I. Smith, "A computer generated multipath fading simulation for mobile radio," *IEEE Transactions on Vehicular Techonology*, vol. 24, pp. 39–40, August 1975.

[83] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Transactions on Vehicular Techonology*, vol. 44, pp. 163–171, February 1995.

[84] Y. L. Guan and L. F. Turner, "Generalised FSMC model for radio channels with correlated fading," in *Communications, IEE Proceedings*, pp. 133–137, April 1999.

[85] T. S. Rappaport, *Wireless Communications: Principles and Practice.* New York: Prentice Hall, Inc, 1996.

[86] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, pp. 58 – 66, April 2002.

[87] D. D. Falconer and S. L. Ariyavisitakul, "Broadband wireless using single carrier and frequency domain equalization," in *The 5th International Symposium on Wireless Personal Multimedia Communications, 2002.*, (Sheraton Waikiki, Honolulu, Hawaii), pp. 27 – 36, 27 - 30 October 2002.

[88] M. Morelli and L. Sanguinetti, "A novel prefiltering technique for downlink transmissions in TDD MC-CDMA systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2064 – 2069, September 2005.

[89] L. U. Choi and R. D. Murch, "A transmit MIMO scheme with frequency domain pre-equalization for wireless frequency selective channels," *IEEE Transactions on Communications*, vol. 3, pp. 929– 938, May 2004.

[90] Y. Zhu and K. B. Letaief, "Frequency domain pre-equalization with precoding for broadband SDMA systems," in *WCNC*, (Hong Kong, China), pp. 1449 – 1454, 11-15 March 2007.

[91] L. Gong, Y. Du, J. Li, and J. Yuan, "A new channel state information utilization criterion in SC-FDE," in *IEEE Radio and Wireless Symposium*, (San Diego, California, USA), pp. 19 – 22, 17 - 19 Jan 2006.

[92] B. Fong, G. Y. Hong, and A. C. M. Fong, "Adaptive single carrier modulation scheme for high-speed mobile computing," in *International Conference on Consumer Electronics*, (Las Vegas, NV, USA), pp. 295 – 296, 8 - 12 Jan 2005.

[93] K. Takeda, H. Tomeba, and F. Adachi, "BER performance analysis of joint Tomlinson-Harashima precoding and frequency-domain equalization," in *WCNC*, (Hong Kong, China), pp. 1465 – 1469, 11-15 March 2007.

[94] L. Hanzo, L. L. Yang, E.-L. Kuan, and K. Yen, *Single and multi-carrier DS-CDMA: multi-user detection, space-time spreading, synchronisation, networking and standards*. John Wiley, 2003.

[95] R. Esmailzadeh, E. Sourour, and M. Nakagawa, "Prerake diversity combining in time-division duplex CDMA mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 795–801, May 1999.

[96] W. M. Jang, B. R. Vojcic, and R. L. Pickholtz, "Joint transmitter-receiver optimization in synchronous multiuser communications over multipath channel," *IEEE Transactions on Communications*, vol. 46, pp. 269–278, February 1998.

[97] M. K. Tsatasnis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing.*, (Atlanta, GA, USA), pp. 1093 – 1096, 7 - 10 May 1996.

[98] J. X. Wu and C. S. Xiao, "Time-varying and frequency-selective channel estimation with unequally spaced pilot symbols," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.*, (Hong Kong, China), pp. 620–623, 6 - 10 April 2003.

[99] T. Hwang and Y. Li, "Iterative cyclic prefix reconstruction for coded single-carrier systems with frequency-domain equalization (SC-FDE)," in *VTC 2003-Spring.*, (Jeju, South Korea), pp. 1841 – 1845, 23 - 25 April 2003.

[100] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time block-coded transmissions over frequency-selective fading channels," *IEEE Communications Letters*, vol. 5, pp. 304 – 306, July 2001.

[101] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time-coded transmissions over broadband wireless channels," in *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2001.*, (San Diego, CA, USA), pp. B–143–B146, 30 September - 3 October 2001.

[102] A. F. Naguib, "Combined interference suppression and frequency domain equalization for space-time block coded transmission," in *IEEE International Conference on Communications, 2003.*, (Anchorage, Alaska, USA), pp. 3261 – 3266, 11 - 15 May 2003.

[103] P. Hoeher, "TCM on frequency-selective land-mobile fading channels," in *5th Tirrenia International Workshop on Digital Communications*, (Tirrenia, Italy), pp. 317–328, 8 - 12 September 1991.

[104] S. Kaiser and P. Hoeher, "Performance of multi-carrier CDMA systems with channel estimation in two dimensions," in *The 8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1997.*, (Helsinki, Finland), pp. 115 – 119, 1 - 4 September 1997.

[105] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by wiener filtering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997.*, (Munich, Germany), pp. 1845 – 1848, 21 - 24 April 1997.

[106] P. Hoeher, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," in *Proc. Sixth Communication Theory Mini-Conference in conjunction with IEEE GLOBECOM '97*, (Phoenix, AZ, USA), pp. 90–96, 3 -8 November 1997.

[107] S. Kaiser, "Multi-carrier CDMA mobile radio systems - analysis and optimization of detection, decoding, and channel estimation," *Ph.D. Thesis*, 1998.

[108] W. G. Jeon, K. H. Paik, and y. S. Cho, "Two-dimensional MMSE channel estimation for OFDM systems with transmitter diversity," in *VTC 2001-Fall.*, (Atlantic City, New Jersey, USA), pp. 1682 – 1685, 7 - 11 October 2001.

[109] H. Witschnig, T. Mayer, A. Springer, L. Maurer, M. Huemer, and R. Weigel, "The advantages of a known sequence versus cyclic prefix in a SC/FDE system," in *The 5th International Symposium on Wireless Personal Multimedia Communications, 2002.*, (Sheraton Waikiki, Honolulu, Hawaii, USA), pp. 1328 – 1332, 27 - 30 October 2002.

[110] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "White paper: Frequency domain equalization for single-carrier broadband wireless systems," in *http://www.sce.carleton.ca/bbw/papers/whitepaper2.pdf.*

[111] C. Tellambura, M. G. Parker, Y. J. Guo, S. J. Shepherd, and S. K. Barton, "Optimal sequences for channel estimation using discrete Fourier transform techniques," *IEEE Transactions on Communications*, vol. 47, pp. 230 – 238, February 1999.

[112] Y. Zeng and T. S. Ng, "Pilot cyclic prefixed single carrier communication: channel estimation and equalization," *IEEE Signal Processing Letters*, vol. 12, pp. 56 – 59, January 2005.

[113] J. Li and Y. Du, "Channel estimation schemes for SC-FDE/FS system," in *IEEE Symposium on Computers and Communications*, (Pula-Cagliari, Sardinia, Italy), pp. 155 – 160, 26 - 29 June 2006.

[114] J. Siew, J. Coon, R. J. Piechocki, A. Dowler, A. Nix, M. Beach, S. Armour, and J. McGeehan, "A channel estimation algorithm for MIMO-SCFDE," *IEEE Communications Letters*, vol. 8, pp. 555 – 557, September 2004.

[115] J. Coon, M. Beach, and J. McGeehan, "Optimal training sequences for channel estimation in cyclic-prefix-based single-carrier systems with transmit diversity," *IEEE Signal Processing Letters*, vol. 11, pp. 729– 732, September 2004.

[116] A. Osseiran, A. Logothetis, and S. B. Slimane, "Pilot design criteria for single carrier frequency domain equalization on a real-time DSP-based MIMO test-bed," in *IEEE Vehicular Technology Conference*, (Montreal, Canada), pp. 1 – 4, 25 - 28 September 2006.

[117] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Transactions on Information Theory*, vol. 18, pp. 531– 532, July 1972.

[118] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, October 1998.

[119] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment using multi-element arrays," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, Autumn 1996.

[120] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *International Symposium on Signals, Systems, and Electronics*, (Pisa, Italy), pp. 295–300, 29 September - 2 October 1998.

[121] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using v-blast space-time communication architecture," *Electronics Letters*, vol. 35, pp. 14–16, January 1999.

[122] A. S. Dakdouki, V. L. Banket, N. K. Mykhaylov, and A. A. Skopa, "Downlink processing algorithms for multi-antenna wireless communications," *IEEE Communications Magazine*, vol. 43, pp. 122 – 127, January 2005.

[123] J. S. Blogh and L. Hanzo, *Third-generation systems and intelligent wireless networking: smart antennas and adaptive modulation.* John Wiley-IEEE Press, 2002.

[124] M. Haardt, A. Klein, R. Koehn, S. Oestreich, M. Purat, V. Sommer, and T. Ulrich, "The TD-CDMA based UTRA TDD mode," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1375 – 1385, August 2000.

[125] M. Peng and W. Wang, "A framework for investigating radio resource management algorithms in TD-SCDMA systems," *IEEE Communications Magazine*, vol. 43, pp. S12 – S18, June 2005.

[126] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted mmse criterion," *IEEE Transactions on Communications*, vol. 49, pp. 2198 – 2206, December 2001.

[127] G. Lebrun, J. Gao, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 757– 764, March 2005.

[128] G. Lebrun, S. Spiteri, and M. Faulkner, "Channel estimation for an SVD-MIMO system," in *IEEE International Conference on Communications*, (Paris, France), pp. 3025 – 3029, 20 - 24 June 2004.

[129] A. Cano-Gutierrez, M. Stojanovic, and J. Vidal, "Effect of channel estimation error on the performance of SVD-based MIMO communication systems," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, (Barcelona, Spain), pp. 508 – 512, 5 - 8 September 2004.

[130] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *IEEE Transactions on Signal Processing*, vol. 52, pp. 2424–2431, September 2004.

[131] A. S. Y. Poon, D. N. C. Tse, and R. W. Brodersen, "An adaptive multiantenna transceiver for slowly flat fading channels," *IEEE Transactions on Communications*, vol. 51, pp. 1820 – 1827, November 2003.

[132] Y. Tang, B. Vucetic, and Y. Li, "An iterative singular vectors estimation scheme for beamforming transmission and detection in MIMO systems," *IEEE Communications Letters*, vol. 9, pp. 505 – 507, June 2005.

[133] Y. Tan, G. Lebrun, and M. Faulkner, "An adaptive channel SVD tracking strategy in time-varying TDD system," in *IEEE Semiannual Vehicular Technology Conference*, (Jeju, Korea), pp. 769 – 773, 22 - 25 April 2003.

[134] H. Z. Jafarian and G. Gulak, "Iterative MIMO channel SVD estimation," in *IEEE International Conference on Communications*, (Seoul, Korea), pp. 1157 – 1161, 16 - 20 May 2005.

[135] H. Z. Jafarian and G. Gulak, "Adaptive channel SVD estimation for MIMO-OFDM systems," in *IEEE Vehicular Technology Conference, Spring*, (Stockholm, Sweden), pp. 552– 556, 30 May - 1 June 2005.

[136] T. Bianchi, C. F. Micheli, and F. Argenti, "SVD tracking algorithm for zero padded block transmission over fading channels," in *IEEE Global Telecommunications Conference,*, (San Francisco, CA, USA), pp. 2125 – 2129, 1 - 3 December 2003.

[137] T. J. Willink, "An efficient SVD update algorithm and application to MIMO communications," in *Proceedings of European Signal Processing Conference*, (Antalya, Turkey), 4 - 8 September 2005.

[138] F. Schafer, M. Stege, C. Michalke, and G. Fettweis, "Efficient tracking of eigenspaces and its application to MIMO-systems," in *Proceedings of the IST Mobile & Wireless Communications Summit*, (Aveiro, Portugal), 15 - 18 June 2003.

[139] C. Michalke, M. Stege, F. Schafer, and G. Fettweis, "Efficient tracking of eigenspaces and its application to eigenbeamforming," in *IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, (Beijing, China), pp. 2847 – 2851, 7 - 10 September 2003.

[140] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Wireless Communications*, vol. 43, pp. 95 – 107, January 1995.

[141] X. Wang and H. V. Poor, "Blind multiuser detection: a subspace approach," *IEEE Transactions on Information Theory*, vol. 44, pp. 677–690, March 1998.

[142] C. Li and X. Wang, "Performance comparisons of MIMO techniques with application to WCDMA systems," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 649–661, May 2004.

[143] H. T. Nguyen, G. Leus, and N. Khaled, "Prediction of the eigenvectors for spatial multiplexing MIMO systems in time-varying channels," in *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology*, (Athens, Greece), pp. 119– 124, 18 - 21 December 2005.

[144] D. Yang, L. L. Yang, and L. Hanzo, "Performance of SDMA systems using transmitter preprocessing based on noisy feedback of vector-quantized channel impulse responses," in *Proceedings of the IEEE Vehicular Technology Conference 2007 Spring*, (Dublin, Ireland), pp. 2119 – 2123, 23 - 25 April 2007.

[145] S. Attallah and K. Abed-Meraim, "Fast algorithms for subspace tracking," *IEEE Signal Processing Letters*, vol. 8, pp. 203 – 206, July 2001.

[146] W. Utschick, "Tracking of signal subspace projectors," *IEEE Signal Processing Letters*, vol. 50, pp. 769 – 778, April 2002.

[147] B. Vucetic and J. Yuan, *Space-Time Coding*. Wiley, 2003.

[148] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge University Press, 2005.

[149] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Communications Magazine*, vol. 42, pp. 60– 67, October 2004.

[150] W. Ajib and D. Haccoun, "An overview of scheduling algorithms in MIMO-based fourth-generation wireless systems," *IEEE Network*, vol. 19, pp. 43– 48, September 2005.

[151] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2700– 2712, August 2005.

[152] L. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 20– 24, January 2004.

[153] R. L. Choi and R. D. Murch, "New transmit schemes and simplified receivers for MIMO wireless communication systems," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 1217– 1230, November 2003.

[154] R. Choi and R. Murch, "MIMO transmit optimization for wireless communication systems," in *The First IEEE International Workshop on Electronic Design, Test and Applications*, (Christchurch, New Zealand), pp. 33 – 37, 29 - 31 January 2002.

[155] Y. Wu, J. Zhang, S. Zhou, and X. Xu, "Precoding in the multiuser MIMO downlink based on subspace tracking techniques," in *IEEE Vehicular Technology Conference*, (Dallas, TX, USA), pp. 2382 – 2386, 25 - 28 September 2005.

[156] L. Choi and R. D. Murch, "Transmit-preprocessing techniques with simplified receivers for the downlink of MISO TDD-CDMA systems," *IEEE Transactions on Wireless Communications*, vol. 35, pp. 285– 295, March 2004.

[157] D. V. Duong, B. Holter, and G. E. Oien, "Optimal pilot spacing and power in rate-adaptive MIMO diversity systems with imperfect transmitter CSI," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, (New York, USA), pp. 47 – 51, 5 - 8 June 2005.

[158] H. T. Nguyen, G. Leus, and N. Khaled, "Precoder and decoder prediction in time-varying MIMO channels," in *Proceedings of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, (Puerto Vallarta, Mexico), pp. 153 – 156, 13 - 15 December 2005.

[159] S. Spiteri, G. Lebnin, and M. Faulkner, "Prediction for time-varying SVD systems," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, (Barcelona, Spain), pp. 1608 – 1612, 5 - 8 September 2004.

[160] T. Svantesson, "A performance bound for prediction of MIMO channels," *IEEE Transactions on Signal Processing*, vol. 54, pp. 520 – 529, February 2006.

[161] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1065 – 1076, May 2002.

[162] D. H. Brandwood, "A complex gradient operator and its apphcation in adaptive array theory," *IEE Proceedings, Part F - Communications, Radar and Signal Processing*, vol. 130, pp. 11 – 16, February 1983.

[163] H. H. Chen, C. X. Fan, and W. W. Lu, "China's perspectives on 3G mobile commuunications and beyond: TD-SCDMA technology," *IEEE Wireless Communications*, vol. 9, pp. 48 – 59, April 2002.

[164] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 461 – 471, February 2004.

[165] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 528–541, March 2006.

[166] R. F. H. Fisher, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," in *4th ITG Conference on Source and Channel Coding*, (Berlin, Germany), pp. 139 – 147, 28 - 30 January 2002.

[167] L. U. Choi and R. D. Murch, "A pre-BLAST-DFE technique for the downlink of frequency-selective fading MIMO channels," *IEEE Transactions on Communications*, vol. 52, pp. 737 – 743, May 2004.

[168] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Transactions on Signal Processing*, vol. 54, pp. 161– 176, January 2006.

[169] S. Serberli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Transactions on Signal Processing*, vol. 52, pp. 214 – 226, January 2004.

[170] Y. Tokgoz and B. D. Rao, "Performance analysis of maximum ratio transmission based multi-cellular MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 83 – 89, January 2006.

[171] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Transactions on Communications*, vol. 47, pp. 1458 – 1461, October 1999.

[172] J. C. Roh and B. D. Rao, "Adaptive modulation for multiple antenna channels," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), pp. 526 – 530, 3 - 6 November 2002.

[173] S. Zhou and G. B. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Transactions on Vehicular Technology*, vol. 3, pp. 1626– 1636, September 2004.

[174] Z. Zhou and B. Vucetic, "The effect of CSI imperfection on the performance of SVD based adaptive modulation in MIMO systems," in *International Symposium on Information Theory*, (Chicago, IL, USA), pp. 317 – 317, 27 June - 2 July 2003.

[175] Z. Zhou and B. Vucetic, "Design of adaptive modulation using imperfect CSI in MIMO systems," *Electronics Letters*, vol. 40, pp. 1073 – 1075, August 2004.

[176] Z. Zhou, B. Vucetic, Z. Chen, and Y. Li, "Design of adaptive modulation in MIMO systems using outdated CSI," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, (Berlin, Germany), pp. 1101 – 1105, 11 - 14 September 2005.

[177] Z. Wang, C. He, and A. He, "Robust AM-MIMO based on minimized transmission power," *IEEE Communications Letters*, vol. 10, pp. 432 – 434, June 2006.

[178] J. H. Sung and J. R. Barry, "Space-time processing with channel knowledge at the transmitter," in *International Conference on International Conference on Trends in Communications EUROCON 2001*, (Bratislava, Slovaki), pp. 26 – 29, 4 - 7 July 2001.

[179] S. X. Ng, B. L. Yeap, and L. Hanzo, "Full-rate, full-diversity adaptive space time block coding for transmission over Rayleigh fading channels," in *IEEE Vehicular Technology Conference*, (Stockholm, Sweden), pp. 1210 – 1214, 30 May - 1 June 2005.

[180] T. J. Willink, "An adaptive algorithm for V-BLAST," in *IEEE Vehicular Technology Conference*, (Stockholm, Sweden), pp. 2044 – 2048, 26 - 29 September 2004.

[181] J. Kim and J. M. Cioffi, "Spatial multiuser access with antenna diversity using singular value decomposition," in *IEEE International Conference on Communications*, (New Orleans, LA, USA), pp. 1253 – 1257, 18-22 June 2000.

[182] K. Kim, S. Lee, and K. Chang, "An efficient multiuser access scheme combining the transmit diversity with the modified SVD methods for MIMO channels," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, California, USA), pp. 1719 – 1721, 3 - 6 November 2002.

[183] K. E. Dawui and D. T. M. Slock, "Multiuser-MIMO downlink TX-RX design based on SVD channel diagonalization and multiuser diversity," in *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), pp. 1493 – 1497, October 28 - November 1 2005.

[184] S. Verdu, *Multiuser Detection.* Cambridge University Press, 1998.

[185] B. Holter, "On the capacity of the MIMO channel: A tutorial introduction," in *IEEE Norwegian Symposium on Signal Processing*, (Trondheim, Norway), pp. 167 – 172, 18 - 20 October 2001.

[186] R. W. Heath, M. Airy, and A. J. Paulraj, "Multiuser diversity for mimo wireless systems with linear receivers," in *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), pp. 1194 – 1199, 4 - 7 Nov 2001.

[187] Y. Ding, T. N. Davidson, Z. Q. Luo, and K. M. Wong, "Minimum ber block precoders for zero-forcing equalization," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2410– 2423, September 2003.

[188] A. Yasotharan, "Multirate zero-forcing Tx-Rx design for MIMO channel under BER constraints," *IEEE Transactions on Signal Processing*, vol. 54, pp. 2288– 2301, June 2006.

[189] K. Zhang and Z. Niu, "Joint transmit rate, power and antenna allocation for MIMO systems with multimedia traffic," *IEICE TRANSACTIONS on Communications*, vol. E89-B, pp. 1939–1942, June 2006.

[190] G. Cocchi and A. Uncini, "Subband neural networks prediction for on-line audio signal recovery," *IEEE Transactions on Neural Networks*, vol. 13, pp. 867–876, July 2002.

[191] A. Uncini and G. Cocchi, "Subband neural networks for noisy signal forecasting and missing data reconstruction," in *International Joint Conference on Neural Networks*, (Honolulu, Hawaii, USA), pp. 438 – 441, 12 - 17 May 2002.

[192] G. Cocchi and A. Uncini, "Subbands audio signal recovering using neural nonlinear prediction," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Salt Lake City, Utah, USA), pp. 1289 – 1292, 7 - 11 May 2001.

[193] S. Haykin and L. Li, "Adaptive digital communication receivers," *IEEE Communications Magazine*, vol. 38, pp. 106 – 114, December 2000.

[194] M. Ibnkahla, "Applications of neural networks to digital communications: A survey," *Signal Processing*, vol. 80, pp. 1185–1215, July 2000.

[195] S. Haykin, *Neural Networks: A Comprehensive Foundation.* Prentice Hall, Inc, 2rd ed., 1999.

[196] S. Ong, C. You, S. Choi, and D. Hong, "A decision feedback recurrent neural equalizer as an infiniteimpulse response filter," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2851 – 2858, November 1997.

[197] E. Varoglu and K. Hacioglu, "Speech prediction using recurrent neural networks," *Electronics Letters*, vol. 35, pp. 1353 – 1355, August 1999.

[198] J. A. Pérez-Ortiz, J. Calera-Rubio, and M. L. Forcada, "A comparison between recurrent neural architectures for real-time nonlinear prediction of speech signals," in *IEEE Signal Processing Society Workshop Neural Networks for Signal Processing*, (Falmouth, Massachusetts, USA), pp. 73 – 81, 10 - 12 September 2001.

[199] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 43, pp. 526 – 535, February 1995.

[200] J. Baltersee and J. A. Chambers, "Nonlinear adaptive prediction of speech with a pipelined recurrent neural network," *IEEE Transactions on Signal Processing*, vol. 46, pp. 2207 – 2216, August 1998.

[201] D. P. Mandic and J. A. Chambers, "Toward an optimal PRNN-based nonlinear predictor," *IEEE Transactions on Neural Networks*, vol. 10, pp. 1435 – 1442, November 1999.

[202] D. P. Mandic and J. A. Chambers, "Advanced PRNN based nonlinear prediction/system identification," in *IEE Colloquium on Non-Linear Signal and Image Processing*, (Savoy Place, London, UK), pp. 11/1 – 11/6, 22 May 1998.

[203] D. P. Mandic and J. A. Chambers, "On the choice of parameters of the cost function in nested modular RNN's," *IEEE Transactions on Neural Networks*, vol. 11, pp. 315 – 322, March 2000.

[204] D. P. Mandic and J. A. Chambers, "From an a priori RNN to an a posteriori PRNN nonlinear predictor," in *IEEE Signal Processing Society Workshop Neural Networks for Signal Processing VIII*, (Cambridge, UK), pp. 174 – 183, 31 August - 2 September 1998.

[205] J. D. Ortiz-Fuentes and M. L. Forcada, "A comparison between recurrent neural network architectures for digital equalization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Munich, Germany), pp. 3281 – 3284, 21 - 24 April 1997.

[206] P. H. G. Coelho, "A new state space model for a complex RTRL neural network," in *International Joint Conference on Neural Networks*, (Washington DC, USA), pp. 1756 – 1761, 15 - 19 July 2001.

[207] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. John Willy & Sons, 2001.

[208] N. Benvenuto, M. Marchesi, F. Piazza, and A. Uncini, "A comparison between real and complex valued neural networks in communication applications," in *Proceedings of the International Conference on Artificial Neural Networks*, (Espoo, Finland), pp. 1177 – 1180, 24 - 28 June 1991.

[209] S. L. Goh and D. P. Mandic, "A complex-valued RTRL algorithm for recurrent neural networks," *Neural Computation*, vol. 16, pp. 2699–2713, December 2004.

[210] G. Kechriotis and E. S. Manolakos, "Training fully recurrent neural networks with complex weights," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 41, pp. 235 – 238, March 1994.

[211] G. Kechriotis, E. Zervas, and E. S. Manolakos, "Using recurrent neural networks for adaptive communication channel equalization," *IEEE Transactions on Neural Networks*, vol. 5, pp. 267 – 278, March 1994.

[212] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989.

[213] S. L. Goh and D. P. Mandic, "Nonlinear adaptive prediction using a complex-valued PRNN," in *IEEE 13th Workshop on Neural Networks for Signal Processing*, (Toulouse, France), pp. 779 – 788, 17 - 19 September 2003.

[214] S. L. Goh, D. Popović, and D. P. Mandic, "Complex-valued estimation of wind profile and wind power," in *Proceedings of the 12th IEEE Mediterranean Electrotechnical Conference*, (Dubrovnik, Croatia), pp. 1037 – 1040, 9 - 12 May 2004.

[215] S. L. Goh and D. P. Mandic, "A data-reusing gradient descent algorithm for complex-valued recurrent neural networks," in *Proceedings of the Knowledge-Based Intelligent Information and Engineering Systems*, (Oxford, UK), pp. 340–350, 3 - 5 September 2003.

[216] G. V. Puskorius and L. A. Feldkamp, "Decoupled extended Kalman filter training of feedforward layered networks," in *International Joint Conference on Neural Networks*, (Seattle, WA, USA), pp. 771 – 777, 8 - 14 July 1991.

[217] X. Wang and H. V. Poor, "Robust multiuser detection in non-gaussian channels," *IEEE Transactions on Signal Processing*, vol. 47, pp. 289 – 305, February 1999.

[218] J. Choi, M. Bouchard, and T. H. Yeap, "Recurrent neural equalization for communication channels in impulsive noise environments," in *IEEE International Joint Conference on Neural Networks*, (Montreal, Canada), pp. 3232 – 3237, 31 July - 4 August 2005.

[219] G. A. Tsihrintzis and C. L. Nikias, "Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process," *IEEE Transactions on Communications*, vol. 43, pp. 904 – 914, February/March/April 1995.

[220] J. G. Gonzalez, *Robust techniques for wireless communications in non-Gaussian environments.* Ph.D. thesis, University of Delaware, USA, 1997.

[221] M. Morelli, M. Pun, and C. J. Kuo, "Frequency-domain pre-equalization for single-carrier space-division multiple-access downlink transmissions," in *VTC 2006-Spring.*, (Melbourne, Australia), pp. 2418– 2422, 7 - 10 May 2006.

[222] A. Gyasi-Agyei, "Cross-layer multiservice opportunistic scheduling for wireless networks," *Communications Magazine*, vol. 44, pp. 50 – 57, June 2006.

[223] C. Anton-Haro, "Cross-layer scheduling for multi-user MIMO systems," *Communications Magazine*, vol. 44, pp. 39 – 45, September 2006.

[224] T. Bonald, "Flow-level performance analysis of some opportunistic scheduling algorithms," *European transactions on telecommunications*, vol. 16, pp. 65 – 75, January 2004.

[225] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277 – 1294, June 2002.

[226] H. J. Bang, T. Ekman, and D. Gesbert, "A channel predictive proportional fair scheduling algorithm," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, (New York City, USA), pp. 620 – 624, 5 - 8 June 2003.

[227] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Communications Magazine*, vol. 42, pp. 68 – 73, October 2004.

[228] A. F. Molisch, "MIMO systems with antenna selection," *IEEE Microwave Magazine*, vol. 5, pp. 46 – 56, March 2004.

[229] Y. Takei and T. Ohtsuki, "Throughput maximization transmission control scheme using channel prediction for MIMO systems," in *IEEE Global Telecommunications Conference*, (St. Louis, MO, USA), pp. 2344 – 2348, 28 November - 2 December 2003.

[230] S. Haykin, *Kalman Filtering and Neural Networks.* John Wiley & Sons, 2001.

[231] S. Singhal and L. Wu, "Training multilayer perceptrons with the extended Kalman algorithm," in *Advances in Neural Information Processing Systems 1*, pp. 133–140, 1989.

[232] G. V. Puskorius and L. A. Feldkamp, "A signal processing framework based on dynamic neural networks with application to problems in adaptation, filtering, and classification," *Proceedings of the IEEE*, vol. 86, pp. 2259 – 2277, November 1998.

[233] L. A. Feldkamp and G. V. Puskorius, "Training of robust neural controllers," in *IEEE Conference on Decision and Control*, (Orlando, FL, USA), pp. 2754 – 2759, 14 - 16 December 1994.

[234] G. V. Puskorius and L. A. Feldkamp, "Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks," *IEEE Transactions on Neural Networks*, vol. 5, pp. 279 – 297, March 1994.

[235] J. Choi, A. C. Lima, and S. Haykin, "Kalman filter-trained recurrent neural equalizers for time-varying channels," *IEEE Transactions on Communications*, vol. 53, pp. 472–480, March 2005.

[236] J. Choi, M. Bouchard, and T. H. Yeap, "Decision feedback recurrent neural equalization with fast convergence rate," *IEEE Transactions on Neural Networks*, vol. 16, pp. 699 – 708, May 2005.

[237] P. H. G. Coelho, "A complex EKF-RTRL neural network," in *International Joint Conference on Neural Networks*, (Washington DC, USA), pp. 120 – 125, 15 - 19 July 2001.

[238] P. H. G. Coelho, "Adaptive channel equalization using EKF-CRTRL neural networks," in *International Joint Conference on Neural Networks*, (Honolulu, Hawaii, USA), pp. 1195 – 1199, 12 - 17 May 2002.

Appendix A

Fully Connected Recurrent Neural Networks

A.1 Introduction

An FCRNN structure has been shown in Figure 8.2, which consists of N activation neurons having P external inputs as well as a fixed bias input b_{bias}. The network has two distinct layers consisting of the external input-feedback layer and a layer of processing elements. Let the $(N \times 1)$-dimensional vector $\mathbf{y}(k) = [y_l(k), \cdots, y_N(k)]^T$ denote the complex-valued output vector of all the neurons at time index k, where $y_n(k)$ is the complex-valued output of the nth neuron, $n = 1, \cdots, N$, and the $(P \times 1)$-dimensional vector $\mathbf{s}(k) = [s(k-1), \cdots, s(k-P)]^T$ denotes the complex-valued external input vector at time index k. Furthermore, let the bias be $b_{bias} = (1 + j)$, where $j = \sqrt{-1}$. Then the $(P+1+N) \times 1$-dimensional combined input vector $\boldsymbol{\rho}(k)$ of the network seen in Figure 8.2 represents the concatenation of the vectors $\mathbf{s}(k)$, b_{bias} and $\mathbf{y}(k-1)$, which is given by

$$\begin{aligned} \boldsymbol{\rho}(k) &= \begin{bmatrix} \mathbf{s}(k) \\ b_{bias} \\ \mathbf{y}(k-1) \end{bmatrix} \\ &= [s(k-1), \cdots, s(k-P), 1+j, y_1(k-1), \cdots, y_N(k-1)]^T \\ &= \boldsymbol{\rho}^{(r)}(k) + j\boldsymbol{\rho}^{(i)}(k), \end{aligned} \tag{A.1}$$

where the superscripts $(\cdot)^{(r)}$ and $(\cdot)^{(i)}$ denote the real and imaginary parts of the argument, respectively. Assuming that the outputs of the first M neurons are the outputs of the FCRNN, then the

$(M \times 1)$-dimensional output vector of the FCRNN $\mathbf{y}_o(k)$ can be expressed as

$$\mathbf{y}_o(k) = [y_1(k), \cdots, y_M(k)]^T, \ 1 \leq M \leq N. \tag{A.2}$$

Let $w_{n,l}$ denote the complex-valued weight of the FCRNN, which connects the nth neuron and lth input, where we have $1 \leq n \leq N$ and $1 \leq l \leq (P+1+N)$. Then the input of the nth node at time index k is given by

$$\begin{aligned} n_{et_n}(k) &= \sum_{l=1}^{P+1+N} w_{n,l}(k)\rho_l(k) \\ &= n_{et_n}^{(r)}(k) + jn_{et_n}^{(i)}(k), \end{aligned} \tag{A.3}$$

where $n_{et_n}^{(r)}(k)$ and $n_{et_n}^{(i)}(k)$ are the real and imaginary parts of $n_{et_n}(k)$, respectively, which can be expressed as [209]

$$\begin{aligned} n_{et_n}^{(r)}(k) &= \sum_{l=1}^{P+1+N} [w_{n,l}^{(r)}(k)\rho_l^{(r)}(k) - w_{n,l}^{(i)}(k)\rho_l^{(i)}(k)] \\ &= \sum_{l=1}^{P} [w_{n,l}^{(r)}(k)s^{(r)}(k-l) - w_{n,l}^{(i)}(k)s^{(i)}(k-l)) + (w_{n,P+1}^{(r)}(k) - w_{n,P+1}^{(i)}(k)] + \\ &\quad \sum_{q=1}^{N} [w_{n,P+1+q}^{(r)}(k)y_q^{(r)}(k-1) - w_{n,P+1+q}^{(i)}(k)y_q^{(i)}(k-1)], \end{aligned} \tag{A.4}$$

$$\begin{aligned} n_{et_n}^{(i)}(k) &= \sum_{l=1}^{P+1+N} [w_{n,l}^{(r)}(k)\rho_l^{(i)}(k) + w_{n,l}^{(i)}(k)\rho_l^{(r)}(k)] \\ &= \sum_{l=1}^{P} [w_{n,l}^{(r)}(k)s^{(i)}(k-l) + w_{n,l}^{(i)}(k)s^{(r)}(k-l)) + (w_{n,P+1}^{(r)}(k) + w_{n,P+1}^{(i)}(k)] + \\ &\quad \sum_{q=1}^{N} [w_{n,P+1+q}^{(r)}(k)y_q^{(i)}(k-1) + w_{n,P+1+q}^{(i)}(k)y_q^{(i)}(k-1)]. \end{aligned} \tag{A.5}$$

The output of the nth activation neuron can be formulated as [209]

$$\begin{aligned} y_n(k) &= \Phi(n_{et_n}(k)) \\ &= \Phi^{(r)}(n_{et_n}(k)) + j\Phi^{(i)}(n_{et_n}(k)) \\ &= y_n^{(r)}(k) + jy_n^{(i)}(k), \ n = 1, \cdots, N, \end{aligned} \tag{A.6}$$

where Φ is a complex-valued nonlinear activation function, while $y_n^{(r)}(k)$ and $y_n^{(i)}(k)$ are the real and imaginary parts of $y_n(k)$, respectively.

Let $d_m(k), 1 \leq m \leq M$ be the correponding desired output of the mth neuron. Then the $(M \times 1)$-dimensional desired output vector of the FCRNN $\mathbf{d}_o(k)$ can be expressed as

$$\mathbf{d}_o(k) = [d_1(k), \cdots, d_M(k)]^T, \ 1 \leq M \leq N. \tag{A.7}$$

A.2 Training Algorithms for FCRNN

In this section, the real time recurrent learning (RTRL) technique [209–215], the global extended Kalman filter (GEKF) aided training [216] and the decoupled extended Kalman filter (DEKF) [216] will be discussed in detail. When $\mathbf{d}_o$ is the desired output vector, the $(M \times 1)$-dimensional error vector of the FCRNN $\mathbf{e}_o(k)$ can be written as

$$\begin{aligned} \mathbf{e}_o(k) &= \mathbf{d}_o(k) - \mathbf{y}_o(k) \\ &= [e_1(k), \cdots, e_M(k)]^T, \ 1 \leq M \leq N, \end{aligned} \tag{A.8}$$

where $e_m(k)$ is the error corresponding to the mth neuron, which is given by

$$\begin{aligned} e_m(k) &= d_m(k) - y_m(k) \\ &= e_m^{(r)}(k) + j e_m^{(i)}(k), \ 1 \leq m \leq M \end{aligned} \tag{A.9}$$

where $e_m^{(r)}(k)$ and $e_m^{(i)}(k)$ are the real part and imaginary part of $e_m(k)$ and are expressed as

$$e_m^{(r)}(k) = d_m^{(r)}(k) - y_m^{(r)}(k), \ 1 \leq m \leq M. \tag{A.10}$$

$$e_m^{(i)}(k) = d_m^{(i)}(k) - y_m^{(i)}(k), \ 1 \leq m \leq M. \tag{A.11}$$

A.2.1 Complex-Valued Real Time Recurrent Learning Algorithm for FCRNN

Real Time Recurrent Learning (RTRL) constitutes the most widely used algorithm for training RNNs. It was proposed in [212] for real-valued cases, where the inputs, output, weights and activation functions are assumed to be real-valued. However, in many applications the inputs and outputs of a practical system are best described as complex valued signals. In these cases, the real-valued RTRL

algorithm has to be extended to the complex-valued RTRL (CRTRL) [210]. There are two different CRTRL algorithms proposed for training FCRNNs in the literature, namely the split CRTRL (SCRTRL) [210, 211] and the full CRTRL (FCRTRL) [209, 213–215]. Let us first consider the SCRTRL algorithm.

A.2.1.1 Split-Complex-Valued Real Time Recurrent Learning Algorithm for FCRNN

For the SCRTRL, (A.6) is first expressed as

$$\begin{aligned} y_n(k) &= y_n^{(r)}(k) + jy_n^{(i)}(k) \\ &= f(n_{et_n}^{(r)}(k)) + jf(n_{et_n}^{(i)}(k)),\ n = 1, \cdots, N, \end{aligned} \tag{A.12}$$

where $f(\cdot)$ is a function having real-valued inputs and real-valued outputs. For real-time applications, the cost function of the recurrent network is given by

$$\begin{aligned} E(k) &= \frac{1}{2}\sum_{m=1}^{M} |e_m(k)|^2 \\ &= \frac{1}{2}\sum_{m=1}^{M} e_m(k)e_m^*(k). \end{aligned} \tag{A.13}$$

Upon substituting Eq.(A.9) into Eq.(A.13), we obtain

$$E(k) = \frac{1}{2}\sum_{m=1}^{M}[(e_m^{(r)}(k))^2 + (e_m^{(i)}(k))^2],\ 1 \leq M \leq N. \tag{A.14}$$

The SCRTRL algorithm minimizes the cost function $E(k)$ by recursively updating the weight coefficients based on the gradient descent scheme, which is given by

$$\begin{aligned} w_{n,l}(k+1) &= w_{n,l}(k) + \triangle w_{n,l}(k) \\ &= w_{n,l}(k) - \eta \bigtriangledown_{w_{n,l}} E(k)|_{w_{n,l}=w_{n,l}(k)},\ 1 \leq n \leq N,\ 1 \leq l \leq P+1+N \end{aligned} \tag{A.15}$$

where η controls the learning rate, which is chosen to be a small, positive constant, and $\bigtriangledown_{w_{n,l}} E(k)$ is the gradient of $E(k)$ with respect to the complex weight $w_{n,l}(k)$. Notice that $E(k)$ is a real-valued

function, and $\nabla_{w_{n,l}} E(k)$ is hence given by

$$\nabla_{w_{n,l}} E(k) = \frac{\partial E(k)}{\partial w_{n,l}^{(r)}} + j\frac{\partial E(k)}{\partial w_{n,l}^{(i)}}, 1 \leq n \leq N,\ 1 \leq l \leq P+1+N. \tag{A.16}$$

where $\frac{\partial E(k)}{\partial w_{n,l}^{(r)}}$ and $\frac{\partial E(k)}{\partial w_{n,l}^{(i)}}$ are the gradients of $E(k)$ with respect to the real and imaginary part of the complex weight $w_{n,l}$, respectively. According to (A.14) and also due to the fact that $E(k)$ is an indirect function of $w_{n,l}(k)$ through the variable $y_m(k)$, it can be shown that $\frac{\partial E(k)}{\partial w_{n,l}^{(r)}}$ can be calculated as

$$\begin{aligned}\frac{\partial E(k)}{w_{n,l}^{(r)}(k)} &= \sum_{m=1}^{M}\left[\frac{\partial E(k)}{y_m^{(r)}(k)}\left(\frac{\partial y_m^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)}\right) + \frac{\partial E(k)}{y_m^{(i)}(k)}\left(\frac{\partial y_m^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}\right)\right] \\ &= -\sum_{m=1}^{M} e_m^{(r)}(k)\frac{\partial y_m^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} - \sum_{m=1}^{M} e_m^{(i)}(k)\frac{\partial y_m^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}.\end{aligned} \tag{A.17}$$

Similarly, $\frac{\partial E(k)}{\partial w_{n,l}^{(i)}}$ can be calculated as

$$\begin{aligned}\frac{\partial E(k)}{w_{n,l}^{(i)}(k)} &= \sum_{m=1}^{M}\left[\frac{\partial E(k)}{y_m^{(r)}(k)}\left(\frac{\partial y_m^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)}\right) + \frac{\partial E(k)}{y_m^{(i)}(k)}\left(\frac{\partial y_m^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}\right)\right] \\ &= -\sum_{m=1}^{M} e_m^{(r)}(k)\frac{\partial y_m^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} - \sum_{m=1}^{M} e_m^{(i)}(k)\frac{\partial y_m^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}.\end{aligned} \tag{A.18}$$

In (A.17) and (A.18) the factors $\frac{\partial y_u(k)}{\partial w_{n,l}^{(r)}(k)} = \frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} + j\frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}$ and $\frac{\partial y_u(k)}{\partial w_{n,l}^{(i)}(k)} = \frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} + j\frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}$ provide measures for the output sensitivity of the uth neuron at time index k in response to a small variation in the value of $w_{n,l}(k)$. According to (A.12), these sensitivities can be evaluated as

$$\frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} = f'(n_{et_u}^{(r)}(k))\frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)}, \tag{A.19}$$

$$\frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} = f'(n_{et_u}^{(r)}(k))\frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)}, \tag{A.20}$$

$$\frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)} = f'(n_{et_u}^{(i)}(k)) \frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}, \tag{A.21}$$

$$\frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)} = f'(n_{et_u}^{(i)}(k)) \frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}, \tag{A.22}$$

where $f'(\cdot)$ is the first derivative of $f(\cdot)$ with respect to its argument. Following the derivation of the real-valued RTRL in [212], after differentiating (A.4) and (A.5) these sensitivities can be computed as

$$\frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} = \left[\sum_{q=1}^{N} \left(\frac{\partial y_q^{(r)}(k-1)}{\partial w_{n,l}^{(r)}(k)} w_{u,P+1+q}^{(r)}(k) - \frac{\partial y_q^{(i)}(k-1)}{\partial w_{n,l}^{(r)}(k)} w_{u,P+1+q}^{(i)}(k) \right) \right] + \delta_{nu} \rho_l^{(r)}(k), \tag{A.23}$$

$$\frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)} = \left[\sum_{q=1}^{N} \left(\frac{\partial y_q^{(i)}(k-1)}{\partial w_{n,l}^{(r)}(k)} w_{u,P+1+q}^{(r)}(k) + \frac{\partial y_q^{(r)}(k-1)}{\partial w_{n,l}^{(r)}(k)} w_{u,P+1+q}^{(i)}(k) \right) \right] + \delta_{nu} \rho_l^{(i)}(k), \tag{A.24}$$

$$\frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} = \left[\sum_{q=1}^{N} \left(\frac{\partial y_q^{(r)}(k-1)}{\partial w_{n,l}^{(i)}(k)} w_{u,P+1+q}^{(r)}(k) - \frac{\partial y_q^{(i)}(k-1)}{\partial w_{n,l}^{(i)}(k)} w_{u,P+1+q}^{(i)}(k) \right) \right] - \delta_{nu} \rho_l^{(i)}(k), \tag{A.25}$$

$$\frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)} = \left[\sum_{q=1}^{N} \left(\frac{\partial y_q^{(i)}(k-1)}{\partial w_{n,l}^{(i)}(k)} w_{u,P+1+q}^{(r)}(k) + \frac{\partial y_q^{(r)}(k-1)}{\partial w_{n,l}^{(i)}(k)} w_{u,P+1+q}^{(i)}(k) \right) \right] + \delta_{nu} \rho_l^{(r)}(k), \tag{A.26}$$

where

$$\delta_{nu} = \begin{cases} 1, & if\ u = n \\ 0, & otherwise \end{cases} \tag{A.27}$$

is the Kronecker delta function. For convenience, we denote the sensitivities as $\pi_{n,l}^{u,(rr)}(k) = \frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)}$, $\pi_{n,l}^{u,(ir)}(k) = \frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}$, $\pi_{n,l}^{u,(ri)}(k) = \frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)}$, $\pi_{n,l}^{u,(ii)}(k) = \frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}$. Then, with the aid of (A.19) to (A.26), these sensitivity items can be recursively updated by the following matrix equations [210,211]

$$\begin{bmatrix} \pi_{n,l}^{u,(rr)}(k+1) & \pi_{n,l}^{u,(ri)}(k+1) \\ \pi_{n,l}^{u,(ir)}(k+1) & \pi_{n,l}^{u,(ii)}(k+1) \end{bmatrix} = \begin{bmatrix} f'(n_{et_u}^{(r)}(k+1)) & 0 \\ 0 & f'(n_{et_u}^{(i)}(k+1)) \end{bmatrix} \times \left(\sum_{q=1}^{N} \begin{bmatrix} w_{u,P+1+q}^{(r)}(k) & -w_{u,P+1+q}^{(i)}(k) \\ w_{u,P+1+q}^{(i)}(k) & w_{u,P+1+q}^{(r)}(k) \end{bmatrix} \begin{bmatrix} \pi_{n,l}^{q,(rr)}(k) & \pi_{n,l}^{q,(ri)}(k) \\ \pi_{n,l}^{q,(ir)}(k) & \pi_{n,l}^{q,(ii)}(k) \end{bmatrix} + \delta_{nu} \begin{bmatrix} \rho_l^{(r)}(k) & -\rho_l^{(i)}(k) \\ \rho_l^{(i)}(k) & \rho_l^{(r)}(k) \end{bmatrix} \right), \quad 1 \le u \le N. \tag{A.28}$$

According to (A.17) to (A.26), (A.16) can also be expressed as

$$\nabla_{w_{n,l}} E(k) = -\sum_{m=1}^{M} \left([e_m^{(r)}(k)\ e_m^{(i)}(k)] \begin{bmatrix} \pi_{n,l}^{m,(rr)}(k) & \pi_{n,l}^{m,(ri)}(k) \\ \pi_{n,l}^{m,(ir)}(k) & \pi_{n,l}^{m,(ii)}(k) \end{bmatrix} \begin{bmatrix} .1 \\ j \end{bmatrix} \right) \tag{A.29}$$

Finally, upon substituting (A.29) into (A.15), the weight update equation becomes

$$\begin{aligned} w_{n,l}(k+1) &= w_{n,l}(k) - \eta \nabla_{w_{n,l}} E(k) \\ &= w_{n,l}(k) + \eta \sum_{m=1}^{M} \left([e_m^{(r)}(k)\ e_m^{(i)}(k)] \begin{bmatrix} \pi_{n,l}^{m,(rr)}(k) & \pi_{n,l}^{m,(ri)}(k) \\ \pi_{n,l}^{m,(ir)}(k) & \pi_{n,l}^{m,(ii)}(k) \end{bmatrix} \begin{bmatrix} 1 \\ j \end{bmatrix} \right), \end{aligned} \tag{A.30}$$

where $1 \le n \le N$ and $1 \le l \le P+1+N$. Having derived the SCRTRL algorithm for the RNN, let us now derive the FCRTRL algorithm.

A.2.1.2 Full Complex-Valued Real Time Recurrent Learning Algorithm for FCRNN

In the context of the FCRTRL algorithm, (A.6) can be expressed as

$$\begin{aligned} y_n(k) &= y_n^{(r)}(k) + j y_n^{(i)}(k) \\ &= \Phi^{(r)}(n_{et_n}^{(r)}(k), n_{et_n}^{(i)}(k)) + j\Phi^{(i)}(n_{et_n}^{(r)}(k), n_{et_n}^{(i)}(k)), \quad n = 1, \cdots, N. \end{aligned} \tag{A.31}$$

Accordingly, the sensitivities defined for FCRTRL can be evaluated as

$$\frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} = \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(r)}(k)} \frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} + \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(i)}(k)} \frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}, \tag{A.32}$$

$$\frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} = \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(r)}(k)} \frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} + \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(i)}(k)} \frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}, \tag{A.33}$$

$$\frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)} = \frac{\partial y_u^{(i)}(k)}{\partial n_{et_u}^{(r)}(k)} \frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} + \frac{\partial y_u^{(i)}(k)}{\partial n_{et_u}^{(i)}(k)} \frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}, \tag{A.34}$$

$$\frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)} = \frac{\partial y_u^{(i)}(k)}{\partial n_{et_u}^{(r)}(k)} \frac{\partial n_{et_u}^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} + \frac{\partial y_u^{(i)}(k)}{\partial n_{et_u}^{(i)}(k)} \frac{\partial n_{et_u}^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}. \tag{A.35}$$

According to (A.23) to (A.26) and (A.32) to (A.35), the sensitivities defined for for FCRTRL can be recursively updated by the following matrix equations [209]

$$\begin{bmatrix} \pi_{n,l}^{u,(rr)}(k+1) & \pi_{n,l}^{u,(ri)}(k+1) \\ \pi_{n,l}^{u,(ir)}(k+1) & \pi_{n,l}^{u,(ii)}(k+1) \end{bmatrix} = \begin{bmatrix} \frac{\partial y_u^{(r)}(k+1)}{\partial n_{et_u}^{(r)}(k+1)} & \frac{\partial y_u^{(r)}(k+1)}{\partial n_{et_u}^{(i)}(k+1)} \\ \frac{\partial y_u^{(i)}(k+1)}{\partial n_{et_u}^{(r)}(k+1)} & \frac{\partial y_u^{(i)}(k+1)}{\partial n_{et_u}^{(i)}(k+1)} \end{bmatrix} \times$$
$$\left(\sum_{q=1}^{N} \begin{bmatrix} w_{u,P+1+q}^{(r)}(k) & -w_{u,P+1+q}^{(i)}(k) \\ w_{u,P+1+q}^{(i)}(k) & w_{u,P+1+q}^{(r)}(k) \end{bmatrix} \begin{bmatrix} \pi_{n,l}^{q,(rr)}(k) & \pi_{n,l}^{q,(ri)}(k) \\ \pi_{n,l}^{q,(ir)}(k) & \pi_{n,l}^{q,(ii)}(k) \end{bmatrix} \right.$$
$$\left. + \delta_{nu} \begin{bmatrix} \rho_l^{(r)}(k) & -\rho_l^{(i)}(k) \\ \rho_l^{(i)}(k) & \rho_l^{(r)}(k) \end{bmatrix} \right), \; 1 \le n, u \le N, \; 1 \le l \le P+1+N. \tag{A.36}$$

For a complex function to be analytic at a point in $Z \in C$, where C encompasses the entire complex plane, it has to satisfy the Cauchy-Riemann equations [209]. To arrive at the Cauchy-Riemann equations, the partial derivatives (sensitivities) along the real and imaginary axes should be equal, requiring,

$$\frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} + j \frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)} = \frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)} - j \frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)}. \tag{A.37}$$

Equating the real and imaginary parts in (A.37), we obtain

$$\frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} = \frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)}, \tag{A.38}$$

$$\frac{\partial y_u^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} = -\frac{\partial y_u^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)}. \tag{A.39}$$

Equivalently, we have

$$\pi_{n,l}^{u,(rr)}(k) = \pi_{n,l}^{u,(ii)}(k), \tag{A.40}$$

$$\pi_{n,l}^{u,(ri)}(k) = -\pi_{n,l}^{u,(ir)}(k). \tag{A.41}$$

With the aid of the Cauchy-Riemann equations, a more compact representation of the gradient $\nabla_{w_{n,l}} E(k)$ can be obtained, which is given by

$$\begin{aligned}
\nabla_{w_{n,l}} E(k) &= -\sum_{m=1}^{M} \left(e_m^{(r)}(k) \frac{\partial y_m^{(r)}(k)}{\partial w_{n,l}^{(r)}(k)} + e_m^{(i)}(k) \frac{\partial y_m^{(i)}(k)}{\partial w_{n,l}^{(r)}(k)} + j e_m^{(r)}(k) \frac{\partial y_m^{(r)}(k)}{\partial w_{n,l}^{(i)}(k)} + j e_m^{(i)}(k) \frac{\partial y_m^{(i)}(k)}{\partial w_{n,l}^{(i)}(k)} \right) \\
&= -\sum_{m=1}^{M} \left(e_m^{(r)}(k) \pi_{n,l}^{m,(rr)}(k) + e_m^{(i)}(k) \pi_{n,l}^{m,(ir)} + j e_m^{(r)}(k) \pi_{n,l}^{m,(ri)}(k) + j e_m^{(i)}(k) \pi_{n,l}^{m,(ii)}(k) \right) \\
&= -\sum_{m=1}^{M} \left(\left[e_m^{(r)}(k) + j e_m^{(i)}(k) \right] \left[\pi_{n,l}^{m,(rr)}(k) + j \pi_{n,l}^{m,(ri)}(k) \right] \right) \\
&= -\sum_{m=1}^{M} e_m(k) \left[\pi_{n,l}^{m,(rr)}(k) - j \pi_{n,l}(k)^{m,(ir)}(k) \right] \\
&= -\sum_{m=1}^{M} e_m(k) \left(\pi_{n,l}^{m} \right)^* (k),
\end{aligned} \tag{A.42}$$

where $e_m(k) = e_m^{(r)}(k) + j e_m^{(i)}(k)$ and $\left(\pi_{n,l}^{m} \right)(k) = \left(\pi_{n,l}^{m,(rr)}(k) + j \pi_{n,l}^{m,(ri)}(k) \right)$. Finally, the weight update is given by

$$\triangle w_{n,l}(k) = \eta \sum_{m=1}^{M} e_m(k) \left(\pi_{n,l}^{m} \right)^* (k), \; 1 \leq M, n \leq N, 1 \leq l \leq P + N + 1. \tag{A.43}$$

As in the context of the RTRL algorithm of [212] under the assumption that for a sufficiently small learning rate η, we have [209]

$$\frac{\partial y_u^{(r)}(k-1)}{\partial w_{n,l}(k)} \approx \frac{\partial y_u^{(r)}(k-1)}{\partial w_{n,l}(k-1)}, \; 1 \leq u,n \leq N, 1 \leq l \leq P+N+1, \tag{A.44}$$

$$\frac{\partial y_u^{(i)}(k-1)}{\partial w_{n,l}(k)} \approx \frac{\partial y_u^{(i)}(k-1)}{\partial w_{n,l}(k-1)}, \; 1 \leq u,n \leq N, 1 \leq l \leq P+N+1, \tag{A.45}$$

the update of the sensitivities $\left(\pi_{n,l}^{u}\right)^{*}(k)$ becomes

$$
\begin{aligned}
\left(\pi_{n,l}^{u}\right)^{*}(k) &= \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(r)}(k)}\Bigg[\delta_{nu}\rho_l^{(r)}(k) - j\delta_{nu}\rho_l^{(i)}(k) + \\
&\sum_{q=1}^{N}\Bigg[\left(w_{u,P+1+q}^{(r)}(k) - jw_{u,P+1+q}^{(i)}(k)\right)\pi_{n,l}^{q,(rr)}(k-1) + \\
&\left(jw_{u,P+1+q}^{(r)}(k) + w_{u,P+1+q}^{(i)}(k)\right)\pi_{n,l}^{q,(ri)}(k-1)\Bigg]\Bigg] \\
&+\frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(i)}(k)}\Bigg[\delta_{nu}\rho_l^{(i)}(k) + j\delta_{nu}\rho_l^{(r)}(k) + \\
&\sum_{q=1}^{N}\Bigg[\left(-w_{u,P+1+q}^{(r)}(k) + jw_{u,P+1+q}^{(i)}(k)\right)\pi_{n,l}^{q,(ri)}(k-1) + \\
&\left(w_{u,P+1+q}^{(i)}(k) + jw_{u,P+1+q}^{(r)}(k)\right)\pi_{n,l}^{q,(rr)}(k-1)\Bigg]\Bigg] \\
&= \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(r)}(k)}\Bigg[\delta_{nu}\rho_l^{*}(k) + \sum_{q=1}^{N}\Bigg[w_{u,P+1+q}^{*}(k)\pi_{n,l}^{q,(rr)}(k-1) + \\
&jw_{u,P+1+q}^{*}(k)\pi_{n,l}^{q,(ri)}(k-1)\Bigg]\Bigg] + \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(i)}(k)}\Bigg[j\delta_{nu}\rho_l^{*}(k) + \\
&\sum_{q=1}^{N}\Bigg[-w_{u,P+1+q}^{*}(k)\pi_{n,l}^{q,(ri)}(k-1) + jw_{u,P+1+q}^{*}(k)\pi_{n,l}^{q,(rr)}(k-1)\Bigg]\Bigg] \\
&= \left(\frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(r)}(k)} + \frac{\partial y_u^{(r)}(k)}{\partial n_{et_u}^{(i)}(k)}\right)\Bigg[\delta_{nu}\rho_l^{*}(k) + \\
&\sum_{q=1}^{N}w_{u,P+1+q}^{*}(k)\left(\pi_{n,l}^{q,(rr)}(k-1) + j\pi_{n,l}^{q,(ri)}(k-1)\right)\Bigg] \\
&= \{\Phi'(n_{et_u}^{(r)}, n_{et_u}^{(i)})(k)\}^{*}\left[\mathbf{w}_{u,\pi}^{H}\boldsymbol{\pi}_{n,l}^{*}(k-1) + \delta_{nu}I_t^{*}(k)\right], \qquad \text{(A.46)}
\end{aligned}
$$

where $\Phi'(\cdot)$ is the first derivative of $\Phi(\cdot)$ with respect to its argument, while $\mathbf{w}_{u,\pi} = [w_{u,P+1+1}, \cdots, w_{u,P+1+N}]^T$ and $\boldsymbol{\pi}_{n,l}(k-1) = [\pi_{n,l}^{1}(k-1), \cdots, \pi_{n,l}^{N}(k-1)]^T$. Finally, by substituting (A.42) into (A.15), the weight update equation becomes

$$
\begin{aligned}
w_{n,l}(k+1) &= w_{n,l}(k) + \eta\,\nabla_{w_{n,l}}E(k) \\
&= w_{n,l}(k) + \eta\sum_{m=1}^{M}e_m(k)\left(\pi_{n,l}^{m}\right)^{*}(k), \qquad \text{(A.47)}
\end{aligned}
$$

where $\left(\pi_{n,l}^{m}\right)^{*}(k)$ is given by (A.46).

A.2.2 Complex Parameter-Based Kalman Filter Training for FCRNNs

Although the RTRL algorithm is popular owing to its reasonable complexity, it is based on the gradient method using first-order derivatives. Hence, it may exhibit an inferior convergence speed in comparison to the family of more sophisticated learning techniques using second-order derivatives [230].

The extended Kalman filter (EKF) [230] forms the basis of a second-order neural network training method. The essence of the recursive EKF procedure is that an approximate covariance matrix that constitutes second-order information about the training problem considered is generated and evolved during the training process. Since Singhal and Wu firstly introduced the EKF training algorithm in [231] for static forward oriented neural networks (FNNs), the EKF has constituted the basis for the enhancement of computationally effective neural network training methods that enable the application of FNNs and RNNs to diverse problems such as pattern classification [216, 232], control [233, 234], channel equalization [235–238], etc.

In this section, the so-called global extended Kalman filter (GEKF) [216] and the decoupled extended Kalman filter (DEKF) [216] aided training algorithms are investigated by using bith the split activation function of (A.12) and the full activation function of (A.31), respectively.

A.2.2.1 Principles of Extended Kalman Filters

The Kalman filtering problem considered in Chapter 3 has focused on the prediction of a state vector in a linear model of a dynamic system. If, however, the model is nonlinear, we may extend the use of Kalman filtering through a linearization procedure. The resultant filter is referred to as the extended Kalman filter (EKF) [1–3].

In order to derive the extended Kalman filter, we commence our discussion by considering a nonlinear dynamical system described by the state-space model of [230]

$$\mathbf{w}(k+1) = \mathbf{f}(k, \mathbf{w}(k)) + \boldsymbol{\omega}(k), \tag{A.48}$$

$$\mathbf{y}(k) = \mathbf{h}(k, \mathbf{w}(k)) + \boldsymbol{\upsilon}(k), \tag{A.49}$$

where $\mathbf{w}(k)$ is a $(L \times 1)$-dimensional state vector and $\mathbf{y}(k)$ is a $(M \times 1)$-dimensional measure-

ment vector while $\omega(k)$ and $\upsilon(k)$ are independent, zero-mean, white Gaussian noise processes with covariance matrices of $\mathbf{Q}_\omega(k)$ and $\mathbf{Q}_\upsilon(k)$, respectively. Furthermore, in (A.48) the functional $\mathbf{f}(k, \mathbf{w}(k))$ denotes a nonlinear transition matrix function that might be time-variant, while the functional $\mathbf{h}(k, \mathbf{w}(k)$ in (A.49) denotes a nonlinear measurement matrix that may be time-variant as well.

The basic principle behind the extended Kalman filter is the linearization of the statespace model of (A.48) and (A.49) at each time instant associated with the most recent state estimate. The most recently estimated state is expressed as either $\hat{\mathbf{w}}(k|k)$ or as $\hat{\mathbf{w}}(k|k-1)$, depending on which particular functional is being considered, as will be detailed in our forthcoming discourse. Once a linear model is obtained, the conventional linear model based Kalman filter equations can then be used.

More specifically, the approximation can be carried out in two stages. In the first stage, the following two matrices are constructed [195, 230]:

$$\mathbf{F}(k+1,k) = \frac{\partial \mathbf{f}(k,\mathbf{w})}{\partial \mathbf{w}}|_{\mathbf{w}=\hat{\mathbf{w}}(k|k)}, \tag{A.50}$$

$$\mathbf{H}(k) = \frac{\partial \mathbf{h}(k,\mathbf{w})}{\partial \mathbf{w}}|_{\mathbf{w}=\hat{\mathbf{w}}(k|k-1)}, \tag{A.51}$$

where $\mathbf{F}(k+1,k)$ is an $(L \times L)$-dimensional matrix, whose ijth entry is equal to the partial derivative of the ith component of $\mathbf{f}(k,\mathbf{w})$ with respect to the jth component of $\mathbf{w}$, while $\mathbf{H}(k)$ is a $(M \times L)$-dimensional matrix whose ijth entry is equal to the partial derivative of the ith component of $\mathbf{h}(k,\mathbf{w})$ with respect to the jth component of $\mathbf{w}$. Note that in (A.50) at point of the derivatives are evaluated at $\mathbf{w} = \hat{\mathbf{w}}(k|k)$ while in (A.51), the derivatives are evaluated at $\mathbf{w} = \hat{\mathbf{w}}(k|k-1)$. The entries of the matrices $\mathbf{F}(k+1,k)$ and $\mathbf{H}(k)$ are all known (i.e., they are computable), since $\hat{\mathbf{w}}(k)$ and $\hat{\mathbf{w}}(k|k-1)$ are available at time k.

In the context of the second stage, once the matrices $\mathbf{F}(k+1,k)$ and $\mathbf{H}(k)$ are evaluated, the first-order Taylor approximation of the nonlinear functions $\mathbf{f}(k,\mathbf{w}(k))$ and $\mathbf{h}(k,\mathbf{w}(k))$ can be employed around the point of $\mathbf{w} = \hat{\mathbf{w}}(k|k)$ and $\mathbf{w} = \hat{\mathbf{w}}(k|k-1)$, respectively. Specifically, $\mathbf{f}(k,\mathbf{w}(k))$ and $\mathbf{h}(k,\mathbf{w}(k))$ can be approximated as [195, 230]

$$\mathbf{f}(k,\mathbf{w}(k)) \approx \mathbf{F}(k,\hat{\mathbf{w}}(k|k)) + \mathbf{F}(k+1,k)[\mathbf{w}(k) - \hat{\mathbf{w}}(k|k)], \tag{A.52}$$

$$\mathbf{h}(k,\mathbf{w}(k)) \approx \mathbf{H}(k,\hat{\mathbf{w}}(k|k-1)) + \mathbf{H}(k)[\mathbf{w}(k) - \hat{\mathbf{w}}(k|k-1)]. \tag{A.53}$$

Given the above approximations, we may now proceed to approximate the nonlinear expression of Eqs.(A.48) and (A.49) using their corresponding linear models, which can be expressed as [230]

$$\mathbf{w}(k+1) \approx \mathbf{F}(k+1,k)\mathbf{w}(k) + \boldsymbol{\omega}(k) + \mathbf{x}(k), \tag{A.54}$$

$$\bar{\mathbf{y}}(k) \approx \mathbf{H}(k)\mathbf{w}(k) + \boldsymbol{\upsilon}(k), \tag{A.55}$$

where we have introduced the new quantities [230]:

$$\bar{\mathbf{y}}(k) = \mathbf{y}(k) - [\mathbf{h}(k, \hat{\mathbf{w}}(k|k-1)) - \mathbf{H}(k)\hat{\mathbf{w}}(k|k-1)], \tag{A.56}$$

$$\mathbf{x}(k) = \mathbf{f}(k, \hat{\mathbf{w}}(k|k)) - \mathbf{F}(k+1,k)\hat{\mathbf{w}}(k|k). \tag{A.57}$$

Since the entries of $\bar{\mathbf{y}}(k)$ are all known at time k, $\bar{\mathbf{y}}(k)$ can be regarded as an observation vector at time k. Similarly, the entries in the term $\mathbf{x}(k)$ are all known at time k.

A.2.2.2 State-Space Model of FCRNN

In order to apply the extended Kalman filtering principles for the training of the FCRNN, the FCRNN has to be firstly formulated in terms of the state-space model. The state space model for the FCRNN can be derived either from the split activation model of (A.12) or from the full activation model of A.31). Let us now derive them in detail.

A.2.2.2.1 State-Space Model of FCRNN Using Split Activation

For the FCRNN depicted in Figure 8.2, which consists of N neurons and has P external inputs, the dynamic behavior of the network which is assumed to be noise-free, can be described by [230]

$$\begin{aligned} \mathbf{y}(k+1) &= \boldsymbol{\Phi}(\mathbf{W}_u(k)\mathbf{u}(k) + \mathbf{W}_y(k)\mathbf{y}(k)) \\ &= \boldsymbol{\Phi}(\mathbf{W}(k)\boldsymbol{\rho}(k)), \end{aligned} \tag{A.58}$$

$$\mathbf{y}_o(k) = \mathbf{C}\mathbf{y}(k). \tag{A.59}$$

In (A.58) and (A.59) the $(N \times 1)$-dimensional vector $\mathbf{y}(k) = [y_1(k), \cdots, y_N(k)]^T$ denotes

the state of the network, the $((P+1)\times 1)$-dimensional vector $\mathbf{u}(k) = [s(k-1),\cdots,s(k-P),b_{bias}]^T$ represents the input applied to the network, the $(M\times 1)$-dimensional vector $\mathbf{y}_o(k) = [y_1(k),\cdots,y_M(k)]^T$ desribes the output of the network, $\mathbf{W}_u(k)$ is a $(N\times(P+1))$-dimensional weight matrix corresponding to the input vector $\mathbf{u}(k)$, while $\mathbf{W}_y(k)$ is a $(N\times N)$-dimensional matrix corresponding to the state vector $\mathbf{y}(k)$. Furthermore, $\mathbf{W}(k)$ is a $(N\times(N+p+1))$-dimensional weight matrix, which is given by

$$\mathbf{W}(k) = [\mathbf{W}_u(k)\ \mathbf{W}_y(k)] \tag{A.60}$$

and $\boldsymbol{\rho}(k) = [s(k-1),\cdots,s(k-P),1+j,y_1(k-1),\cdots,y_N(k-1)]^T$ and $\Phi : R^q \to R^q$ is a diagonal map. Finally, in Eq.(A.59) $\mathbf{C}$ is a $(M\times N)$-dimensional matrix. The dimensionality of the state space, namely N, is the order of the system. Therefore, the state-space model of Figure 8.2 is a P-input, M-output recurrent model of order N. Eqs (A.58) and (A.59) are the process equation and measurement equation of the model, respectively. The process equation in the state-space description of Eq.(A.58) can be rewritten in the following form:

$$\mathbf{y}(k+1) = \begin{bmatrix} \Phi(\mathbf{w}_1^T(k)\boldsymbol{\rho}(k)) \\ \Phi(\mathbf{w}_2^T(k)\boldsymbol{\rho}(k)) \\ \vdots \\ \Phi(\mathbf{w}_N^T(k)\boldsymbol{\rho}(k)) \end{bmatrix}, \tag{A.61}$$

where the $((P+1+N)\times 1)$-dimensional weight vector $\mathbf{w}_n$ is given by $\mathbf{w}_n = [w_{n,1},\cdots,w_{n,P+1+N}]^T$, where $n = 1,2,\cdots,N$. The weight vector $\mathbf{w}_n$ belongs to the nth neuron in the recurrent network seen in Figure 8.2 and corresponds to the nth row of the weight matrix $\mathbf{W}(k)$.

Letting $\mathbf{n_{et}}(k+1) = \mathbf{W}(k)\boldsymbol{\rho}(k)$, we obtain [236]

$$\mathbf{n_{et}}^{(r)}(k+1) = \mathbf{W}^{(r)}(k)\boldsymbol{\rho}^{(r)}(k) - \mathbf{W}^{(i)}(k)\boldsymbol{\rho}^{(i)}(k), \tag{A.62}$$

$$\mathbf{n_{et}}^{(i)}(k+1) = \mathbf{W}^{(r)}(k)\boldsymbol{\rho}^{(i)}(k) + \mathbf{W}^{(i)}(k)\boldsymbol{\rho}^{(r)}(k). \tag{A.63}$$

For the sake of convenient presentation, we introduce a range of new matrices in the spirit of [236], which are defined as follows.

- The derivative matrix of the state vector $\mathbf{y}(k)$ with respect to the weight vector $\mathbf{w}_n$ is defined

as [236]

$$\begin{aligned}\mathbf{\Lambda}_n^{(AB)}(k) &= \frac{\partial \mathbf{y}^{(A)}}{\partial \mathbf{w}_n^{(B)}} \\ &= \begin{bmatrix} \frac{\partial y_1^{(A)}(k)}{\partial w_{n,1}^{(B)}} & \frac{\partial y_1^{(A)}(k)}{\partial w_{n,2}^{(B)}} & \cdots & \frac{\partial y_1^{(A)}(k)}{\partial w_{n,N+p+1}^{(B)}} \\ \frac{\partial y_2^{(A)}(k)}{\partial w_{n,1}^{(B)}} & \frac{\partial y_2^{(A)}(k)}{\partial w_{n,2}^{(B)}} & \cdots & \frac{\partial y_2^{(A)}(k)}{\partial w_{n,N+p+1}^{(B)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_N^{(A)}(k)}{\partial w_{n,1}^{(B)}} & \frac{\partial y_N^{(A)}(k)}{\partial w_{n,2}^{(B)}} & \cdots & \frac{\partial y_N^{(A)}(k)}{\partial w_{n,N+p+1}^{(B)}} \end{bmatrix},\end{aligned} \tag{A.64}$$

where $A, B \in \{r, i\}$ and $\mathbf{\Lambda}_n^{(AB)}(k)$ is a $(N \times (N+p+1))$-dimensional matrix. According to the definition of the derivative matrix given in (A.64), the overall derivative matrix can be represented as [236]

$$\mathbf{\Lambda}_n(k) = \begin{bmatrix} \mathbf{\Lambda}_n^{(rr)}(k) & \mathbf{\Lambda}_n^{(ri)}(k) \\ \mathbf{\Lambda}_n^{(ir)}(k) & \mathbf{\Lambda}_n^{(ii)}(k) \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{y}^{(r)}(k)}{\partial \mathbf{w}_n^{(r)}} & \frac{\partial \mathbf{y}^{(r)}(k)}{\partial \mathbf{w}_n^{(i)}} \\ \frac{\partial \mathbf{y}^{(i)}(k)}{\partial \mathbf{w}_n^{(r)}} & \frac{\partial \mathbf{y}^{(i)}(k)}{\partial \mathbf{w}_n^{(i)}} \end{bmatrix}, \tag{A.65}$$

where $n = 1, 2, \cdots, N$.

- $\mathbf{Z}_n^{(A)}(k)$ is defined to be a $(N \times (N+p+1))$-dimensional matrix whose rows are all zero, except for the nth row, which is equal to the transpose of the vector $\boldsymbol{\rho}(k)$, i.e., $\mathbf{Z}_n^{(A)}(k)$ is defined as [236]

$$\mathbf{Z}_n^{(A)}(k) = \begin{bmatrix} \mathbf{0}^T \\ (\boldsymbol{\rho}(k)^{(A)})^T(k) \\ \mathbf{0}^T \end{bmatrix} \leftarrow n\text{th row}, \; n = 1, 2, \cdots, N, \tag{A.66}$$

where $A \in \{r, i\}$.

- $\mathbf{\Gamma}^{(A)}(k)$ is defined as the $(N \times N)$-dimensional diagonal matries, which are given by [236]

$$\mathbf{\Gamma}^{(r)}(k+1) = \text{diag}\left\{ \frac{dy_1^{(r)}(k+1)}{dn_{et_1}^{(r)}(k+1)}, \frac{dy_2^{(r)}(k+1)}{dn_{et_2}^{(r)}(k+1)}, \cdots, \frac{dy_N^{(r)}(k+1)}{dn_{et_n}^{(r)}(k+1)} \right\}, \tag{A.67}$$

$$\mathbf{\Gamma}^{(i)}(k+1) = \text{diag}\left\{ \frac{dy_1^{(i)}(k+1)}{dn_{et_1}^{(i)}(k+1)}, \frac{dy_2^{(i)}(k+1)}{dn_{et_2}^{(i)}(k+1)}, \cdots, \frac{dy_N^{(i)}(k+1)}{dn_{et_n}^{(i)}(k+1)} \right\}, \tag{A.68}$$

where $\frac{dy_n^{(r)}(k+1)}{dn_{et_n}^{(r)}(k+1)}$ denotes the first derivative of $y_n^{(r)}(k+1)$ with $n_{et_n}^{(r)}(k+1)$.

With the aid of the definitions given above, upon differentiating (A.61) and using the chain rule of calculus [236], $\mathbf{\Lambda}_n^{(rr)}(k)$ is defined as [236]

$$\begin{aligned}\mathbf{\Lambda}_n^{(rr)}(k+1) &= \frac{\partial \mathbf{y}^{(r)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)} \\ &= \frac{d\mathbf{y}^{(r)}(k+1)}{d\mathbf{n_{et}}^{(r)}(k+1)} \frac{\partial \mathbf{n_{et}}^{(r)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)},\end{aligned} \tag{A.69}$$

where the two derivatives are represented as [236]

$$\begin{aligned}\frac{d\mathbf{y}^{(r)}(k+1)}{d\mathbf{n_{et}}^{(r)}(k+1)} &= \text{diag}\left\{\frac{dy_1^{(r)}(k+1)}{dn_{et_1}^{(r)}(k+1)}, \frac{dy_2^{(r)}(k+1)}{dn_{et_2}^{(r)}(k+1)}, \cdots, \frac{dy_N^{(r)}(k+1)}{dn_{et_n}^{(r)}(k+1)}\right\} \\ &= \mathbf{\Gamma}^{(r)}(k+1),\end{aligned} \tag{A.70}$$

$$\frac{\partial \mathbf{n_{et}}^{(r)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)} = \mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(rr)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{Z}_n^{(r)}(k). \tag{A.71}$$

Substituting (A.70) and (A.71) into (A.69), $\mathbf{\Lambda}_n^{(rr)}(k)$ can then be expressed as [236]

$$\mathbf{\Lambda}_n^{(rr)}(k+1) = \mathbf{\Gamma}^{(r)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(rr)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{Z}_n^{(r)}(k)]. \tag{A.72}$$

Similarly, the terms of $\mathbf{\Lambda}_n^{(ri)}(k+1)$, $\mathbf{\Lambda}_n^{(ii)}(k+1)$, $\mathbf{\Lambda}_n^{(ir)}(k+1)$ can respectively be expressed as [236]

$$\mathbf{\Lambda}_n^{(ri)}(k+1) = \mathbf{\Gamma}^{(r)}(k+1)\left[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ri)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ii)}(k) - \mathbf{Z}_n^{(i)}(k)\right], \tag{A.73}$$

$$\mathbf{\Lambda}_n^{(ii)}(k+1) = \mathbf{\Gamma}^{(i)}(k+1)\left[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ii)}(k) + \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ri)}(k) + \mathbf{Z}_n^{(r)}(k)\right], \tag{A.74}$$

$$\mathbf{\Lambda}_n^{(ir)}(k+1) = \mathbf{\Gamma}^{(i)}(k+1)\left[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_{nl}^{(rr)}(k) + \mathbf{Z}_n^{(i)}(k)\right]. \tag{A.75}$$

Finally, when expressing (A.72) to (A.75) in an compact form, we can obtain the following recursive

equation [236]

$$\begin{bmatrix} \boldsymbol{\Lambda}_n^{(rr)}(k+1) & \boldsymbol{\Lambda}_n^{(ri)}(k+1) \\ \boldsymbol{\Lambda}_n^{(ir)}(k+1) & \boldsymbol{\Lambda}_n^{(ii)}(k+1) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}^{(r)}(k+1) & 0 \\ 0 & \boldsymbol{\Gamma}^{(i)}(k+1) \end{bmatrix} \times \left(\begin{bmatrix} \mathbf{W}_y^{(r)}(k) & -\mathbf{W}_y^{(i)}(k) \\ \mathbf{W}_y^{(i)}(k) & \mathbf{W}_y^{(r)}(k) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_n^{(rr)}(k) & \boldsymbol{\Lambda}_n^{(ri)}(k) \\ \boldsymbol{\Lambda}_n^{(ir)}(k) & \boldsymbol{\Lambda}_n^{(ii)}(k) \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_n^{(r)}(k) & -\mathbf{Z}_n^{(i)}(k) \\ \mathbf{Z}_n^{(i)}(k) & \mathbf{Z}_n^{(r)}(k) \end{bmatrix} \right). \quad \text{(A.76)}$$

A.2.2.2.2 State-Space Model of FCRNN Using Full Activation

In the case of using the full activation in the FCRNN, to the best of our knowledge, no previous work has been done in this field. Hence, we will derive the state-space model for FCRNN in the context of full activation fucntion below. Firstly, we introduce the following new matrices.

- $\boldsymbol{\Gamma}^{(AB)}(k)$ is a $(N \times N)$ diagonal matrix, which is defined as

$$\boldsymbol{\Gamma}^{(rr)}(k+1) = \text{diag}\left\{ \frac{\partial y_1^{(r)}}{\partial n_{et_1}^{(r)}}, \frac{\partial y_2^{(r)}}{\partial n_{et_2}^{(r)}}, \cdots, \frac{\partial y_N^{(r)}}{\partial n_{et_n}^{(r)}} \right\}, \quad \text{(A.77)}$$

$$\boldsymbol{\Gamma}^{(ri)}(k+1) = \text{diag}\left\{ \frac{\partial y_1^{(r)}}{\partial n_{et_1}^{(i)}}, \frac{\partial y_2^{(r)}}{\partial n_{et_2}^{(i)}}, \cdots, \frac{\partial y_N^{(r)}}{\partial n_{et_n}^{(i)}} \right\}, \quad \text{(A.78)}$$

$$\boldsymbol{\Gamma}^{(ir)}(k+1) = \text{diag}\left\{ \frac{\partial y_1^{(i)}}{\partial n_{et_1}^{(r)}}, \frac{\partial y_2^{(i)}}{\partial n_{et_2}^{(r)}}, \cdots, \frac{\partial y_N^{(i)}}{\partial n_{et_n}^{(r)}} \right\}, \quad \text{(A.79)}$$

$$\boldsymbol{\Gamma}^{(ii)}(k+1) = \text{diag}\left\{ \frac{\partial y_1^{(i)}}{\partial n_{et_1}^{(i)}}, \frac{\partial y_2^{(i)}}{\partial n_{et_2}^{(i)}}, \cdots, \frac{\partial y_N^{(i)}}{\partial n_{et_n}^{(i)}} \right\}. \quad \text{(A.80)}$$

Similarly, $\boldsymbol{\Lambda}_n^{(rr)}(k)$ is defined as

$$\begin{aligned} \boldsymbol{\Lambda}_n^{(rr)}(k+1) &= \frac{\partial \mathbf{y}^{(r)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)} \\ &= \frac{\partial \mathbf{y}^{(r)}(k+1)}{\partial \mathbf{n_{et}}^{(r)}(k+1)} \frac{\partial \mathbf{n_{et}}^{(r)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)} + \frac{\partial \mathbf{y}^{(r)}(k+1)}{\partial \mathbf{n_{et}}^{(i)}(k+1)} \frac{\partial \mathbf{n_{et}}^{(i)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)}, \end{aligned} \quad \text{(A.81)}$$

where the derivatives are further represented as

$$\begin{aligned}\frac{\partial \mathbf{y}^{(r)}(k+1)}{\partial \mathbf{n_{et}}^{(r)}(k+1)} &= \text{diag}\left\{\frac{\partial y_1^{(r)}}{\partial n_{et_1}^{(r)}}, \frac{\partial y_2^{(r)}}{\partial n_{et_2}^{(r)}}, \cdots, \frac{\partial y_N^{(r)}}{\partial n_{et_n}^{(r)}}\right\} \\ &= \mathbf{\Gamma}^{(rr)}(k+1), \end{aligned} \tag{A.82}$$

$$\begin{aligned}\frac{\partial \mathbf{y}^{i}(k+1)}{\partial \mathbf{n_{et}}^{(i)}(k+1)} &= \text{diag}\left\{\frac{\partial y_1^{(r)}}{\partial n_{et_1}^{(i)}}, \frac{\partial y_2^{(r)}}{\partial n_{et_2}^{(i)}}, \cdots, \frac{\partial y_N^{(r)}}{\partial n_{et_n}^{(i)}}\right\} \\ &= \mathbf{\Gamma}^{(ri)}(k+1), \end{aligned} \tag{A.83}$$

$$\frac{\partial \mathbf{n_{et}}^{(r)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)} = \mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(rr)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{Z}_n^{(r)}(k), \tag{A.84}$$

$$\frac{\partial \mathbf{n_{et}}^{(i)}(k+1)}{\partial \mathbf{w}_n^{(r)}(k)} = \mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(rr)}(k) + \mathbf{Z}_n^{(i)}(k). \tag{A.85}$$

By substituting (A.82) to (A.85) into (A.81), $\mathbf{\Lambda}_n^{(rr)}(k+1)$ may be expressed as

$$\begin{aligned}\mathbf{\Lambda}_n^{(rr)}(k+1) &= \mathbf{\Gamma}^{(rr)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(rr)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{Z}_n^{(r)}(k)] + \\ &\quad \mathbf{\Gamma}^{(ri)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(rr)}(k) + \mathbf{Z}_n^{(i)}(k)]. \end{aligned} \tag{A.86}$$

Upon following similar steps as above, the matrices $\mathbf{\Lambda}_n^{(ri)}(k+1)$, $\mathbf{\Lambda}_n^{(ii)}(k+1)$, $\mathbf{\Lambda}_n^{(ir)}(k+1)$ can be respectively expressed as

$$\begin{aligned}\mathbf{\Lambda}_n^{(ri)}(k+1) &= \mathbf{\Gamma}^{(rr)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ri)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ii)}(k) - \mathbf{Z}_n^{(i)}(k)] + \\ &\quad \mathbf{\Gamma}^{(ri)}(k+1)[\mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ii)}(k) + \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ri)}(k) + \mathbf{Z}_n^{(r)}(k)], \end{aligned} \tag{A.87}$$

$$\begin{aligned}\mathbf{\Lambda}_n^{(ii)}(k+1) &= \mathbf{\Gamma}^{(ir)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ri)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ii)}(k) - \mathbf{Z}_n^{(i)}(k)] + \\ &\quad \mathbf{\Gamma}^{(ii)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ii)}(k) + \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ri)}(k) + \mathbf{Z}_n^{(r)}(k)], \end{aligned} \tag{A.88}$$

$$\begin{aligned}\mathbf{\Lambda}_n^{(ir)}(k+1) &= \mathbf{\Gamma}^{(ir)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(rr)}(k) - \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{Z}_n^{(r)}(k)] + \\ &\quad \mathbf{\Gamma}^{(ii)}(k+1)[\mathbf{W}_y^{(r)}(k)\mathbf{\Lambda}_n^{(ir)}(k) + \mathbf{W}_y^{(i)}(k)\mathbf{\Lambda}_n^{(rr)}(k) + \mathbf{Z}_n^{(i)}(k)]. \qquad \text{(A.89)}\end{aligned}$$

Finally, combining (A.86) to (A.89) into a single matrix, we arrive at a recursive equation, which is expressed as

$$\begin{aligned}\begin{bmatrix} \mathbf{\Lambda}_n^{(rr)}(k+1) & \mathbf{\Lambda}_n^{(ri)}(k+1) \\ \mathbf{\Lambda}_n^{(ir)}(k+1) & \mathbf{\Lambda}_n^{(ii)}(k+1) \end{bmatrix} &= \begin{bmatrix} \mathbf{\Gamma}^{(rr)}(k+1) & \mathbf{\Gamma}_\tau^{(ri)}(k+1) \\ \mathbf{\Gamma}^{(ir)}(k+1) & \mathbf{\Gamma}_\tau^{(ii)}(k+1) \end{bmatrix} \times \left(\begin{bmatrix} \mathbf{W}_y^{(r)}(k) & -\mathbf{W}_y^{(i)}(k) \\ \mathbf{W}_y^{(i)}(k) & \mathbf{W}_y^{(r)}(k) \end{bmatrix} \right. \\ &\quad \left. \times \begin{bmatrix} \mathbf{\Lambda}_n^{(rr)}(k) & \mathbf{\Lambda}_n^{(ri)}(k) \\ \mathbf{\Lambda}_n^{(ir)}(k) & \mathbf{\Lambda}_n^{(ii)}(k) \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_n^{(r)}(k) & -\mathbf{Z}_n^{(i)}(k) \\ \mathbf{Z}_n^{(i)}(k) & \mathbf{Z}_n^{(r)}(k) \end{bmatrix} \right). \qquad \text{(A.90)}\end{aligned}$$

Note that the full activation function is analytic, hence according to Cauchy-Riemann equation we can arrive at

$$\begin{aligned}\begin{bmatrix} \mathbf{\Lambda}_n^{(rr)}(k+1) & \mathbf{\Lambda}_n^{(ri)}(k+1) \\ \mathbf{\Lambda}_n^{(ir)}(k+1) & \mathbf{\Lambda}_n^{(ii)}(k+1) \end{bmatrix} &= \begin{bmatrix} \mathbf{\Gamma}^{(rr)}(k+1) & \mathbf{\Gamma}^{(ri)}(k+1) \\ -\mathbf{\Gamma}^{(ri)}(k+1) & \mathbf{\Gamma}^{(rr)}(k+1) \end{bmatrix} \times \left(\begin{bmatrix} \mathbf{W}_y^{(r)}(k) & -\mathbf{W}_y^{(i)}(k) \\ \mathbf{W}_y^{(i)}(k) & \mathbf{W}_y^{(r)}(k) \end{bmatrix} \right. \\ &\quad \left. \begin{bmatrix} \mathbf{\Lambda}_n^{(rr)}(k) & \mathbf{\Lambda}_n^{(ri)}(k) \\ \mathbf{\Lambda}_n^{(ir)}(k) & \mathbf{\Lambda}_n^{(ii)}(k) \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_n^{(r)}(k) & -\mathbf{Z}_n^{(i)}(k) \\ \mathbf{Z}_n^{(i)}(k) & \mathbf{Z}_n^{(r)}(k) \end{bmatrix} \right). \qquad \text{(A.91)}\end{aligned}$$

A.2.2.3 Global Extended Kalman Filter Training for FCRNNs

In order to employ the EKF for training the FCRNN, the network's behavior has to be recast in the form of the following nonlinear discrete-time system: [230]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \boldsymbol{\omega}(k), \qquad \text{(A.92)}$$

$$\mathbf{y}_o(k) = \mathbf{C}\boldsymbol{\Phi}(\mathbf{w}(k), \boldsymbol{\rho}(k)) + \boldsymbol{\upsilon}(k), \qquad \text{(A.93)}$$

where $\mathbf{w}(k)$ is a $(L \times 1)$-dimensional vector, which is defined by

$$\mathbf{w}(k) = \begin{bmatrix} \mathbf{w}_1(k) \\ \mathbf{w}_2(k) \\ \vdots \\ \mathbf{w}_N(k) \end{bmatrix}, \qquad \text{(A.94)}$$

where $L = (P + 1 + N) \times N$ is the total number of the weights,while $\mathbf{w}_n(k)(n = 1, 2, \cdots, N)$ is the nth column of the transposed weight matrix $\mathbf{W}^T(k)$ in (A.60). (A.92) is the process equation, suggesting that the state of the system is given by the network's weight parameter values $\mathbf{w}(k)$ corrupted by the process noise $\boldsymbol{\omega}(k)$, where $\mathbf{w}(k)$ is a stationary process. By contrast, (A.93) is the measurement equation, which formulates the network's output vector $\mathbf{y}_o(k)$ as a nonlinear function $\boldsymbol{\Phi}(\cdot)$ of the weight vector $\mathbf{w}(k)$ and the vector $\boldsymbol{\rho}(k)$. As seen in (A.1) $\boldsymbol{\rho}(k)$ includes both the input vector $\mathbf{u}(k)$ and the recurrent node activations $\mathbf{y}(k)$. Furthermore, in (A.93) $\boldsymbol{\upsilon}(k)$ represents a random measurement noise vector. Note that the process noise $\boldsymbol{\omega}(k)$ is typically characterized as zero-mean white noise having a covariance given by $E[\boldsymbol{\omega}(i)\boldsymbol{\omega}^H(j)] = \delta_{ij}\mathbf{Q}_{\boldsymbol{\omega}}(j)$. The measurement noise $\boldsymbol{\upsilon}(k)$ is usually also characterized as zero-mean white noise having a covariance given by $E[\boldsymbol{\upsilon}(i)\boldsymbol{\upsilon}^H(j)] = \delta_{ij}\mathbf{Q}_{\boldsymbol{\upsilon}}(j)$.

In order to apply the GEKF to the state-space model of (A.54) and (A.55), the nonlinear function $\boldsymbol{\Phi}(\mathbf{w}(k), \boldsymbol{\rho}(k))$ seen in (A.93) must be linearized. According to (A.51), the linearization process can be described as [230]

$$\begin{aligned}
\boldsymbol{\Lambda}(k+1) &= \left.\frac{\partial \boldsymbol{\Phi}(\mathbf{w}(k+1), \boldsymbol{\rho}(k+1))}{\partial \mathbf{w}(k)}\right|_{\hat{\mathbf{w}}(k|k-1)} \\
&= \left.\nabla_{\mathbf{w}} \mathbf{y}(k+1)\right|_{\hat{\mathbf{w}}(k|k-1)} \\
&= \left.\left[\frac{\partial \mathbf{y}(k+1)}{\partial \mathbf{w}^{(r)}(k)} - j\frac{\partial \mathbf{y}(k+1)}{\partial \mathbf{w}^{(i)}(k)}\right]\right|_{\hat{\mathbf{w}}(k|k-1)} \\
&= \left.[\boldsymbol{\Lambda}^{(rr)}(k+1) + \boldsymbol{\Lambda}^{(ii)}(k+1)) + j(\boldsymbol{\Lambda}^{(ir)}(k+1) - \boldsymbol{\Lambda}^{(ri)}(k+1)]\right|_{\hat{\mathbf{w}}(k|k-1)},
\end{aligned} \tag{A.95}$$

where we have [236]

$$\begin{aligned}
\boldsymbol{\Lambda}^{(AB)}(k) &= \left.\frac{\partial \mathbf{y}^{(A)}(k)}{\partial \mathbf{w}^{(B)}}\right|_{\hat{\mathbf{w}}(k|k-1)} \\
&= \left.\begin{bmatrix}
\frac{\partial y_1^{(A)}(k)}{\partial w_{1,1}^{(B)}} & \cdots & \frac{\partial y_1^{(A)}(k)}{\partial w_{1,P+1+N}^{(B)}} & \cdots & \frac{\partial y_1^{(A)}(k)}{\partial w_{N,1}^{(B)}} & \cdots & \frac{\partial y_1^{(A)}(k)}{\partial w_{N,P+1+N}^{(B)}} \\
\frac{\partial y_2^{(A)}(k)}{\partial w_{1,1}^{(B)}} & \cdots & \frac{\partial y_2^{(A)}(k)}{\partial w_{1,P+1+N}^{(B)}} & \cdots & \frac{\partial y_2^{(A)}(k)}{\partial w_{N,1}^{(B)}} & \cdots & \frac{\partial y_2^{(A)}(k)}{\partial w_{N,P+1+N}^{(B)}} \\
\vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\
\frac{\partial y_N^{(A)}(k)}{\partial w_{1,1}^{(B)}} & \cdots & \frac{\partial y_N^{(A)}(k)}{\partial w_{1,P+1+N}^{(B)}} & \cdots & \frac{\partial y_N^{(A)}(k)}{\partial w_{N,1}^{(B)}} & \cdots & \frac{\partial y_N^{(A)}(k)}{\partial w_{N,P+1+N}^{(B)}}
\end{bmatrix}\right|_{\hat{\mathbf{w}}(k|k-1)} \\
&= \left.[\boldsymbol{\Lambda}_1^{(AB)}(k)\ \boldsymbol{\Lambda}_2^{(AB)}(k)\ \cdots\ \boldsymbol{\Lambda}_N^{(AB)}(k)]\right|_{\hat{\mathbf{w}}(k|k-1)},
\end{aligned} \tag{A.96}$$

where $A, B \in \{r, i\}$ and $\hat{\mathbf{w}}(k|k-1)$ represents the estimate of the state of the system at time k,

given the observed data up to the time instant of $(k-1)$. The resultant linearized equations can be expressed as [236]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \boldsymbol{\omega}(k), \tag{A.97}$$

$$\bar{\mathbf{y}}_o(k) = \mathbf{C}\boldsymbol{\Lambda}(k+1)\mathbf{w}(k) + \boldsymbol{\upsilon}(k), \tag{A.98}$$

where we have [230]

$$\bar{\mathbf{y}}_o(k) = \mathbf{y}_o(k) - \mathbf{C}[\boldsymbol{\Phi}(\mathbf{w}(k), \mathbf{z}(k))|_{\hat{\mathbf{w}}(k|k-1)} - \boldsymbol{\Lambda}(k+1)\hat{\mathbf{w}}(k|k-1)]. \tag{A.99}$$

Finally, with the aid of the linearized equations shown in (A.97) and (A.98) the training problem using the global extended Kalman filter can now be described as that of finding the minimum mean-squared error estimate of the state vector $\mathbf{w}(k)$ using all the observed data available. Specifically, let $\mathbf{H}(k) = \mathbf{C}\boldsymbol{\Lambda}(k+1)$. Then the GEKF solution to the training problem is given by the following recursion [1]:

$$\boldsymbol{\Upsilon}(k) = [\mathbf{Q}_{\boldsymbol{\upsilon}}(k) + \mathbf{H}(k)\mathbf{K}(k|k-1)\mathbf{H}^H(k)]^{-1}, \tag{A.100}$$

$$\mathbf{G}(k) = \mathbf{K}(k|k-1)\mathbf{H}^H(k)\boldsymbol{\Upsilon}(k), \tag{A.101}$$

$$\boldsymbol{\alpha}(k) = \mathbf{y}_o(k) - \hat{\mathbf{y}}_o(k|k-1), \tag{A.102}$$

$$\hat{\mathbf{w}}(k+1|k) = \hat{\mathbf{w}}(k|k-1) + \mathbf{G}(k)\boldsymbol{\alpha}(k), \tag{A.103}$$

$$\mathbf{K}(k+1|k) = \mathbf{K}(k|k-1) - \mathbf{G}(k)\mathbf{H}(k)\mathbf{K}(k|k-1) + \mathbf{Q}_{\boldsymbol{\omega}}(k), \tag{A.104}$$

where $\boldsymbol{\Upsilon}(k)$ represents a $(M \times M)$-dimensional global scaling matrix, $\mathbf{G}(k)$ is the $(L \times M)$-dimensional Kalman gain matrix and $\boldsymbol{\alpha}(k)$ is the $(M \times 1)$-dimensional innovation process. Furthermore, $\mathbf{y}_o(k)$ is the $(M \times 1)$-dimensional desired vector while $\hat{\mathbf{y}}_o(k|k-1)$ is the $(M \times 1)$-dimensional estimate of $\mathbf{y}_o(k)$ based on the input data available at time $(k-1)$ and it is represented by the actual

output vector $\mathbf{C}\Phi(\hat{\mathbf{w}}(k|k-1), \rho(k))$, when residing in state $\hat{\mathbf{w}}(k|k-1)$, while $\mathbf{K}(k|k-1)$ is the approximate error covariance matrix. Notice that the cost function $E(k)$ can be expressed as

$$E(k) = \frac{1}{2}\alpha^H(k)\alpha(k). \tag{A.105}$$

A.2.2.4 Decoupled Extended kalman Filter Training for FCRNNs

The computational requirements of GEKF are dominated by the need to store and update the approximate error covariance matrix $\mathbf{K}(k|k-1)$ at each time instant. For a neural network architecture having M outputs and L weights, the computational complexity of the GEKF algorithm is proportional to $O(ML^2)$ and its storage requirement is proportional to $O(L^2)$ [230]. In order to reduce the complexity of the GEKF algorithm, the DEKF algorithm has been proposed [216], which was derived from the GEKF by assuming that the interactions between certain weight estimates can be ignored. Explicitly, this simplification introduces many zeros into the error covariance matrix $\mathbf{K}(k|k-1)$. Furthermore, if the weights of GEKF are decoupled so that they can be divided into groups and the groups are mutually exclusive of one another, then the error covariance matrix $\mathbf{K}(k|k-1)$ can be arranged into a block-diagonal matrix form. Let g represent the number of these weight groups. Then, for group $n, 1 \leq n \leq g$, the vector $\hat{\mathbf{w}}^n(k|k-1)$ represents to the estimated weight parameters, $\mathbf{H}^n(k)$ is the submatrix of derivatives of network outputs with respect to the nth group's weights, $\mathbf{K}^n(k|k-1)$ is the weight group's approximate error covariance matrix, and $\mathbf{G}^n(k)$ is its Kalman gain matrix. The concatenation of the vectors $\hat{\mathbf{w}}^n(k|k-1)$ forms the vector $\hat{\mathbf{w}}(k|k-1)$. Similarly, the global derivative matrix $\mathbf{H}(k)$ of (A.51) is composed by the concatenation of the individual submatrices $\mathbf{H}^n(k)$. The DEKF algorithm for the nth weight group is given by [230]

$$\Upsilon(k) = \left[\mathbf{Q}_{\upsilon}(k) + \sum_{l=1}^{g} \mathbf{H}^l(k)\mathbf{K}^l(k|k-1)(\mathbf{H}^l(k)^l)^H\right]^{-1}, \tag{A.106}$$

$$\mathbf{G}^n(k) = \mathbf{K}^n(k|k-1)(\mathbf{H}^n(k))^H\Upsilon(k), \tag{A.107}$$

$$\alpha(k) = \mathbf{y}_o(k) - \hat{\mathbf{y}}_o(k|k-1), \tag{A.108}$$

$$\hat{\mathbf{w}}^n(k+1|k) = \hat{\mathbf{w}}^n(k|k-1) + \mathbf{G}^n(k)\alpha(k), \tag{A.109}$$

$$\mathbf{K}^n(k+1|k) = \mathbf{K}^n(k|k-1) - \mathbf{G}^n(k)\mathbf{H}^n(k)\mathbf{K}^n(k|k-1) + \mathbf{Q}^n(k). \quad \text{(A.110)}$$

Note that the DEKF algorithm reduces to the GEKF algorithm, when we have $g = 1$.

The computational complexity and storage requirements of the DEKF can be significantly lower than those of the GEKF. It can be shown that for g disjoint weight groups, the computational complexity of the DEKF becomes $O(M^2L + M\sum_{n=1}^{g} L_n^2)$, where L_n is the number of weights in group n, while the total storage requirements become $O(\sum_{n=1}^{g} L_n^2)$ [230].

Although many different approaches have been proposed for decoupling in RNNs [230], in this chapter only that specific scenario is considered [230], when the weights corresponding to different nodes are assumed to be independent of each other. Furthermore, we note that the cost function $E(k)$ used for the DEKF is the same as that of the GEKF.

A.2.2.5 Parameter Settings

The measurement and process noise covariance matrices, $\mathbf{Q}_{\boldsymbol{\upsilon}}(k)$ and $\mathbf{Q}_{\boldsymbol{\omega}}(k)$ have to be specified for all training patterns for both the GEKF and the DEKF training algorithm. Let $\eta(k)$ be the measurement noise coefficient to be specified by the user at the time index k. Then $\mathbf{Q}_{\boldsymbol{\upsilon}}(k)$ can be expressed for both the GEKF and the DEKF training algorithm as [236]

$$\mathbf{Q}_{\boldsymbol{\upsilon}}(k) = \eta^{-1}(k)(\mathbf{I} + j\mathbf{I}), \quad \text{(A.111)}$$

where $\mathbf{Q}_{\boldsymbol{\upsilon}}(k)$ is a $(M \times M)$-dimensional matrix. Normally, the measurement noise coefficient is assigned a low value at the beginning of training, which is increased gradually to a value no larger than unity during the course of training [230]. Specifically, in this chapter the measurement noise coefficent is assumed to be

$$\eta(k) = \eta(0) + \frac{0.8}{\mathcal{K} - k}, \quad k > 0, \quad \text{(A.112)}$$

where $\eta(0)$ is the initial value of the measurement noise coefficient, while $\mathcal{K}$ is the total number of training data samples. Furthermore, let $\varsigma(k)$ be the process noise coefficient, then for the GEKF, $\mathbf{Q}_{\boldsymbol{\omega}}(k)$ can be expressed as

$$\mathbf{Q}_{\boldsymbol{\omega}}(k) = \varsigma(k)(\mathbf{I} + j\mathbf{I}) \quad \text{(A.113)}$$

and for the DEKF $\mathbf{Q}_{\omega}^{n}(k)$ can be expressed as

$$\mathbf{Q}_{\omega}^{n}(k) = \varsigma(k)(\mathbf{I}^{n} + j\mathbf{I}^{n}), \tag{A.114}$$

where $\mathbf{Q}_{\omega}(k)$ is a $(L \times L)$-dimensional matrix and $\mathbf{Q}_{\omega}^{n}(k)$ is a $(L_n \times L_n)$-dimensional matrix, respectively. The process noise coefficient $\varsigma(k)$ generally has a relatively large initial value on the order of 10^{-2} and then it is tapered to a small value of the order of 10^{-6} [230]. Specifically, in this chapter the process noise coefficent used is given by

$$\varsigma(k) = \frac{\varsigma(0) - k * 0.000004}{2.0}, \quad k > 0. \tag{A.115}$$

Furthermore, if $\varsigma(k) < 0.000001$, we set $\varsigma(k) = 0.000001$. In (A.115) $\varsigma(0)$ is the initial value of the process noise coefficient.

As shown in (A.100) and (A.106), there is an approximate error covariance matrix $\mathbf{K}(1|0)$, which must be initialized at the beginning of the training process to reflect the fact that no a-priori knowledge was used to initialize the weights. Specifically, for the GEKF $\mathbf{K}(1|0)$ can be expressed as [236]

$$\mathbf{K}(1|0) = \vartheta^{-1}(\mathbf{I} + j\mathbf{I}), \tag{A.116}$$

where $\mathbf{K}(1|0)$ is a $(L \times L)$-dimensional matrix and for the DEKF $\mathbf{K}^{n}(1|0)$ can be formulated as [236]

$$\mathbf{K}^{n}(1|0) = \vartheta^{-1}(\mathbf{I}^{n} + j\mathbf{I}^{n}), \tag{A.117}$$

where $\mathbf{K}(1|0)$ is a $(L_n \times L_n)$-dimensional matrix, and ϑ is used to initialize both $\mathbf{K}(1|0)$ and $\mathbf{K}^{n}(1|0)$.

A.3 Activation Function

For the split activation function of (A.12) the real-valued input and real-valued output function $f(\cdot)$ was chosen as the logistic sigmoid function [209, 213–215]

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{A.118}$$

where x is real-valued. The first derivative of $f(\cdot)$ with respect to its argument can be expressed as

$$\begin{aligned} f'(x) &= \left(\frac{1}{1+e^{-x}}\right)' \\ &= f(x)\left(1-f(x)\right). \end{aligned} \tag{A.119}$$

By subsituting (A.118) into (A.12), we obtain

$$\begin{aligned} y_n(k) &= f(n_{etn}^{(r)}(k)) + jf(n_{etn}^{(i)}(k)) \\ &= \frac{1}{1+e^{-\beta n_{etn}^{(r)}(k)}} + j\frac{1}{1+e^{-\beta n_{etn}^{(i)}(k)}}. \end{aligned} \tag{A.120}$$

For our full activation function, the complex activation function was chosen to be the logistic sigmoid function directly. Then (A.31) can be rewritten as

$$\begin{aligned} y_n(k) &= \frac{1}{1+e^{-(n_{etn}^{(r)}(k)+jn_{etn}^{(i)}(k))}} \\ &= \frac{1}{1+e^{-n_{etn}^{(r)}(k)}\cos n_{etn}^{(i)}(k) - je^{-n_{etn}^{(r)}(k)}\sin n_{etn}^{(i)}(k)} \\ &= \frac{1+e^{-n_{etn}^{(r)}(k)}\cos n_{etn}^{(i)}(k)}{1+2e^{-n_{etn}^{(r)}(k)}\cos n_{etn}^{(i)}(k)+e^{-2n_{etn}^{(r)}(k)}} + j\frac{e^{-n_{etn}^{(r)}(k)}\sin n_{etn}^{(i)}(k)}{1+2e^{-n_{etn}^{(r)}}\cos n_{etn}^{(i)}+e^{-2n_{etn}^{(r)}}} \\ &= y_n^{(r)}(n_{etn}^{(r)}(k), n_{etn}^{(i)}(k)) + jy_n^{(i)}(n_{etn}^{(r)}(k), n_{etn}^{(i)}(k)), \end{aligned} \tag{A.121}$$

where we have

$$y_n^{(r)}(n_{etn}^{(r)}(k), n_{etn}^{(i)}(k)) = \frac{1+e^{-n_{etn}^{(r)}(k)}\cos n_{etn}^{(i)}(k)}{1+2e^{-n_{etn}^{(r)}(k)}\cos n_{etn}^{(i)}(k)+e^{-2n_{etn}^{(r)}(k)}}, \tag{A.122}$$

$$y_n^{(i)}(n_{etn}^{(r)}(k), n_{etn}^{(i)}(k)) = \frac{e^{-n_{etn}^{(r)}(k)}\sin n_{etn}^{(i)}(k)}{1+2e^{-n_{etn}^{(r)}}\cos n_{etn}^{(i)}+e^{-2n_{etn}^{(r)}}}. \tag{A.123}$$

The partial derivative of $y_n^{(r)}(n_{etn}^{(r)}(k), n_{etn}^{(i)}(k))$ with respect to $n_{etn}^{(r)}(k)$ and $n_{etn}^{(i)}(k)$ can be expressed,

respectively, as

$$\frac{\partial y_n^{(r)}(n_{et_n}^{(r)}(k), n_{et_n}^{(i)}(k))}{\partial n_{et_n}^{(r)}(k)} = \frac{e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + 2e^{-2n_{et_n}^{(r)}(k)}}{(1 + 2e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + e^{-2n_{et_n}^{(r)}(k)})^2} + \frac{e^{-3n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k)}{(1 + 2e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + e^{-2n_{et_n}^{(r)}(k)})^2} \quad \text{(A.124)}$$

$$\frac{\partial y_n^{(r)}(n_{et_n}^{(r)}(k), n_{et_n}^{(i)}(k))}{\partial n_{et_n}^{(i)}(k)} = \frac{e^{-n_{et_n}^{(r)}(k)} \sin n_{et_n}^{(i)}(k) - e^{-3n_{et_n}^{(r)}(k)} \sin n_{et_n}^{(i)}(k)}{(1 + 2e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + e^{-2n_{et_n}^{(r)}(k)})^2}. \quad \text{(A.125)}$$

Furthermote, the partial derivative of $y_n^{(i)}(n_{et_n}^{(r)}(k), n_{et_n}^{(i)}(k))$ with respect to $n_{et_n}^{(r)}(k)$ and $n_{et_n}^{(i)}(k)$ can be formulated, respectively, as

$$\frac{\partial y_n^{(i)}(n_{et_n}^{(r)}(k), n_{et_n}^{(i)}(k))}{\partial n_{et_n}^{(r)}(k)} = \frac{e^{-3n_{et_n}^{(r)}(k)} \sin n_{et_n}^{(i)}(k) - e^{-n_{et_n}^{(r)}(k)} \sin n_{et_n}^{(i)}(k)}{(1 + 2e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + e^{-2n_{et_n}^{(r)}(k)})^2}, \quad \text{(A.126)}$$

$$\frac{\partial y_n^{(i)}(n_{et_n}^{(r)}(k), n_{et_n}^{(i)}(k))}{\partial n_{et_n}^{(i)}(k)} = \frac{e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + 2e^{-2n_{et_n}^{(r)}(k)}}{(1 + 2e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + e^{-2n_{et_n}^{(r)}(k)})^2} + \frac{e^{-3n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k)}{(1 + 2e^{-n_{et_n}^{(r)}(k)} \cos n_{et_n}^{(i)}(k) + e^{-2n_{et_n}^{(r)}(k)})^2}. \quad \text{(A.127)}$$

From (A.124) to (A.127), we can see that the complex sigmoid function satisfies the Cauchy-Riemann Equation, which means that it is analytic in $\mathcal{C}$, except for the points $n_{et_n}(k) = (2l + 1)\pi$, where (A.121) is undefined.

A.4 Initialization

For both the SCRTRL and FCRTRL learning algorithms, at the begining of the training, in order to update the weights at time index 1, the value of $\pi_{n,l}^{u,(rr)}(0)$, $\pi_{n,l}^{u,(rr)}(0)$, $\pi_{n,l}^{u,(rr)}(0)$ and $\pi_{n,l}^{u,(rr)}(0)$ has to be known, as seen in (A.30) and (A.47). Since we assume that the initial state of the network has no funtional dependence on the weights, for both the SCRTRL and FCRTRL learning algorithms we

have [209]

$$\pi_{n,l}^{u,(rr)}(0) = 0, 1 \leq u, n \leq N, 1 \leq l \leq P + 1 + N, \tag{A.128}$$

$$\pi_{n,l}^{u,(ri)}(0) = 0, 1 \leq u, n \leq N, 1 \leq l \leq P + 1 + N, \tag{A.129}$$

$$\pi_{n,l}^{u,(ii)}(0) = 0, 1 \leq u, n \leq N, 1 \leq l \leq P + 1 + N, \tag{A.130}$$

$$\pi_{n,l}^{u,(ir)}(0) = 0, 1 \leq u, n \leq N, 1 \leq l \leq P + 1 + N. \tag{A.131}$$

In a similar way, for the EKF training algorithm we also have $\mathbf{\Lambda}_n^{(rr)}(0) = 0$, $\mathbf{\Lambda}_n^{(ri)}(0) = 0$, $\mathbf{\Lambda}_n^{(ir)}(0) = 0$ and $\mathbf{\Lambda}_n^{(ii)}(0) = 0$ [236]. Furthermore, for all the SCRTRL, FCRTRL and EKF training algorithms, each element $w_{n,l}(0)$ is initialized by letting both the real part and imaginary part of $w_{n,l}(0)$ equal a small value that satifies $|w_{n,l}^{(r)(0)}| < 0.0001$ and $|w_{n,l}^{(i)}(0)| < 0.0001$ [236]. Both the real and imaginary parts of $w_{n,l}(0)$ are drawn from a random uniform distribution. Moreover, the activations of each neuron $y_u(0), 1 \leq u \leq N$ are initialized in the same way as the weights matrix. Furthermore, for the EKF training algorithm the measurement noise coefficient and process noise coefficient are initialized as $\eta(0) = 0.1$ and $\varsigma(0) = 0.01$, respectively, while ϑ is initialized as $\vartheta = 0.01$.

Appendix B

Water-Filling Based Power Allocation

Assume that there are K parallel SISO Gaussian channels and that the power allocation coefficient for the kth SISO channel is p_i, under the constraint of

$$\sum_{k=1}^{K} \vartheta_k p_k = K. \tag{B.1}$$

Furthermore, we assume that each of the K SISO Gaussian channels has an SNR of $(\lambda_k p_k)/\sigma^2$, when the system's capacity is given by [147]

$$I = \sum_{k=1}^{K} \left(\log_2 \left[1 + \frac{\lambda_k p_k}{\sigma^2} \right] \right). \tag{B.2}$$

In order to achieve the maximum attainable system capacity, based on the Lagrange mutiplier, we have [147]

$$Z = \sum_{k=1}^{K} \left(\log_2 \left[1 + \frac{\lambda_k p_k}{\sigma^2} \right] \right) + L \left(K - \sum_{k=1}^{K} \vartheta_k p_k \right), \tag{B.3}$$

where L is the Lagrange multiplier. The derivative of Z with respect to p_k is given by [147]

$$\frac{dZ}{dp_k} = \frac{1}{ln2} \frac{\lambda_k/\sigma^2}{1 + p_k \lambda_k/\sigma^2} - L\vartheta_k. \tag{B.4}$$

By setting $\frac{dZ}{dp_k}$ to zero, we arrive at [147]

$$\begin{aligned} p_k &= \frac{1}{L\vartheta_k ln2} - \frac{\sigma^2}{\lambda_k} \\ &= v - \frac{\sigma^2}{\lambda_k}, \end{aligned} \tag{B.5}$$

where we have $v = 1/(L\vartheta_k ln2)$, which can be determined from (B.1)

List of Symbols

- f_n: Doppler frequency shift;
- f_{dm}: maximum Doppler frequency shift;
- $c(\tau, t)$: complex low-pass channel impulse response;
- T_m: maximum delay spread;
- B_c: coherence bandwidth of wireless channel;
- B_d: Doppler spread;
- $(\Delta t)_c$: coherence time of wireless channel
- $S(f)$: power sprectrum of the received signal;
- $c(t)$: complex channel coefficient;
- T_b: data symbol duration;
- c_k: complex fading signal $c(t)$ sampled at the time instant of $t = kT_b$,
- $\hat{c}_n$: prediction of c_n;
- $\mathbf{I}$: identity matrix;
- SNR: signal to noise ratio;
- $\mathbf{c}_n$: $p \times 1$ state vector;
- $\hat{\mathbf{c}}_{[n|n-1]}$: MMSE prediction of $\mathbf{c}_n$;
- $\boldsymbol{\varepsilon}_n$: error between $\hat{\mathbf{c}}_{[n|n-1]}$ and $\mathbf{c}_n$;
- $\mathbf{R}_n$: autocorrelation matrix of $\boldsymbol{\alpha}_n$;
- $\mathbf{K}_n$: predicted state-error correlation matrix of $\boldsymbol{\varepsilon}_n$;
- $\mathbf{G}_n$: Kalman gain;

- $\mathbf{M}_n$: filtered state-error correlation matrix;
- $\mathbf{e}_n$: estimate error of $\mathbf{c}_n$;
- $c_{(n;k)}$: combined channel impulse response at time instant nT_b with dalay kT_b;
- $b_{k;n}$: the nth useful data in the kth data block;
- T_B: duration of data block;
- u: length of the cyclic prefix;
- $r_{(k;n)}$: the nth received signal in the kth data block;
- $c_{(k;n)}$: the combined channel impulse response for the nth received signal in the kth data block;
- $z_{(k;n)}$: the AWGN for the nth received signal in the kth data block;
- $\breve{c}_{(k;n)}$: DFT of $c_{(k;n)}$;
- $\tilde{c}_{(k;n)}$: rough estimate to $\breve{c}_{(k;n)}$;
- $\hat{c}_{(k;n)o}$: optimum frequency domain estimate to $\breve{c}_{(k;n)}$ in the MMSE sense;
- $\acute{c}_{(k;n)o}$: optimum time domain estimate to $\breve{c}_{(k;n)}$ in the MMSE sense;
- $\grave{\mathbf{c}}_{[(k+1;n)|(k;n)]}$: MMSE prediction to $\mathbf{c}_{(k+1;n)}$ for a fixed frequency component;
- $\mathbf{H}$: channel matrix for single user MIMO system;
- $\mathbf{U}$: left singular vector matrix of SVD of channel matrix for single user MIMO system;
- $\mathbf{V}$: right singular vector matrix of SVD of channel matrix for single user MIMO system;
- $\mathbf{\Lambda}$: singular value matrix of SVD of channel matrix for single user MIMO system;
- $\mathbf{U}_s$: basis of column space of channel matrix for single user MIMO system;
- $\mathbf{V}_s$: basis of row space of channel matrix for single user MIMO system;
- $\mathbf{U}_n$: basis of null space of channel matrix for single user MIMO system;
- $\mathbf{V}_n$: basis of left null space of channel matrix for single user MIMO system;
- λ_p: singular values of channel matrix for single user MIMO system;

- $\mathbf{x}_{(dl,k)}$: symbol vector for downlink transmission from BS to the kth MS;
- $\mathbf{P}_k$: preprocessing matrix for symbol vector $\mathbf{x}_{(dl,k)}$;
- $\mathbf{d}_{(dl,k)}$: preprocessed symbol vector for $\mathbf{x}_{(dl,k)}$;
- $\mathbf{d}_{dl}$: overall preprocessed symbol vector for all K MSs;
- $\mathbf{P}$: overall preprocessed matrix for all K MSs;
- $\mathbf{x}_{dl}$: overall symbol vector for downlink transmission from BS to all K MSs;
- $\mathbf{H}$: overall channel matrix for all K MSs for downlink transmission;
- $\mathbf{H}^{+}$: pseudo-inverse of the matrix $\mathbf{H}$;
- $\mathbf{x}_{(ul,k)}$: symbol vector for uplink transmission from the kth MS to BS;
- $\mathbf{y}$: overall received singal at BS from all K MSs for uplink transmission;
- $\mathbf{x}_{ul}$: overall symbol vector for uplink transmission from all K MSs to BS;
- $\mathbf{x}_k$: transmitted symbol vector from the kth MS to BS for multiple user MIMO systems;
- $\mathbf{Q}_k$: preprocessing matrix at the kth MS for multiple user MIMO systems;
- $\mathbf{d}_k$: effective transmitted symbol vector from the kth MS to BS for multiple user MIMO systems;
- $\mathbf{H}_k$: channel matrix between the kth MS and the BS for multiple user MIMO systems;
- $\mathbf{y}$: received signal vector at the BS for multiple user MIMO systems;
- $\mathbf{x}$: overall transmitted symbol vector from all K MSs to BS for multiple user MIMO systems;
- $\mathbf{Q}$: overall preprocessing matrix for all K MSs for multiple user MIMO systems;
- $\mathbf{d}$: overall effective transmitted symbol vector from all K MSs for multiple user MIMO systems;
- $\mathbf{H}$: overall channel matrix between the K MSs and the BS for multiple user MIMO systems;
- $\mathbf{T}_k$: postprocessing matrix for the kth MS at BS for multiple user MIMO systems;

- $\mathbf{T}$: overall postprocessing matrix for all K MSs at BS for multiple user MIMO systems;
- $\mathbf{U}_k$: left singular vector matrix of channel matrix between the kth MS and the BS for multiple user MIMO systems;
- $\mathbf{V}_k$: right singular vector matrix of channel matrix between the kth MS and the BS for multiple user MIMO systems;
- $\mathbf{\Lambda}_k$: eigenvalue matrix of channel matrix between the kth MS and the BS for multiple user MIMO systems;
- $\bar{\mathbf{x}}_k$: transmitted symbol vector from BS to the kth MS for multiple user MIMO systems;
- $\mathbf{P}_k$: preprocessing matrix at the BS for the kth MS for multiple user MIMO systems;
- $\bar{\mathbf{d}}_k$: effective transmitted symbol vector at the BS for the kth MS for multiple user MIMO systems;
- $\bar{\mathbf{H}}_k$: channel matrix from the BS to the kth MS for multiple user MIMO systems;
- $\bar{\mathbf{y}}_k$: received signal vector from the BS to the kth MS for multiple user MIMO systems;
- $\bar{\mathbf{x}}$: overall transmitted symbol vector from the BS to all K MSs for multiple user MIMO systems;
- $\mathbf{P}$: overall preprocessing matrix at the BS for multiple user MIMO systems;
- $\bar{\mathbf{d}}$: overall effective transmitted symbol vector from the BS to all K MSs for multiple user MIMO systems;
- $\bar{\mathbf{U}}_k$: left singular vector matrix of channel matrix from the BS to the kth MS for multiple user MIMO systems;
- $\bar{\mathbf{V}}_k$: right singular vector matrix of channel matrix from the BS to the kth MS for multiple user MIMO systems;
- $\bar{\mathbf{\Lambda}}_k$: eigenvalue matrix of channel matrix from the BS to the kth MS for multiple user MIMO systems;

Index

Author Index

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

W

X

Y

Z

Glossary

3G	third-generation
ACF	autocorrelation function
AOA	angles of arrival
AR	autoregessive
BS	base station
CDMA	Code Division Multiple Access
CE	Channel estimation
CIR	channel impulse response
CRTRL	complex-valued RTRL
CSI	channel state information
CTF	channel transfer fucntion
DDCE	decision-directed channel estimator
DEKF	decoupled extended Kalman filter
DL	downlink
DPSK	differential phase shift keying
EKF	extended Kalman filter
FCRNN	fully connected recurrent neural network
FCRTRL	full CRTRL

FDD	frequency division duplexing
FD	frequency-domain
GEKF	global extended Kalman filter
GSNR	geometric signal-to-noise ratio
IBI	interblock interference
ISI	intersymbol interference
LMS	least mean square
LRP	long-range prediction
MIMO	multiple input multiple output
MMSE	minimum mean square error
MS	mobile station
MUI	Multiple User Interference
OFDM	Orthogonal Frequency-Division Multiplexing
PAPR	peak-to-average power ratio
PASTD	projection approximation tracking combined with deflation
RLS	recursive least square
RNN	Recurrent neural network
RTRL	real time recurrent learning
SC-FDE	single carrier frequency domain equalization
SCRTRL	split CRTRL
SDMA	Space Division Multiple Access
SISO	single input single output
ST-CIR	spatio-temporal CIR
SVD	singular value decomposition
TD-SCDMA	Time Division-Synchronous Code Division Multiple Access
TDD	time division duplexing

TDMA	Time Division Multiple Access
TD	time-domain
UL	uplink
UTRA	UMTS Terrestrial Radio Access
WSSUS	wide sense stationary uncorrelated scattering
WSS	wide sense stationary