

CHANNEL ESTIMATION
AND PREDICTION
FOR MIMO OFDM SYSTEMS
Key Design and Performance Aspects of
Kalman-based Algorithms

Daniel Aronsson

March 2011



UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Polhemssalen, Ångströmlaboratoriet, Lägerhyddsvägen 1, Uppsala, Friday, March 25, 2011 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in Swedish.

Abstract

Aronsson, D. 2011. Channel Estimation and Prediction for MIMO OFDM Systems. Key Design and Performance Aspects of Kalman-based Algorithms. Institutionen för teknikvetenskap. 245 pp. Uppsala. ISBN 978-91-506-2194-5.

Wireless broadband systems based on Orthogonal Frequency Division Multiplexing (OFDM) are being introduced to meet demands for high data transfer rates. In multiple users systems, the available bandwidth has to be shared efficiently by several users. The radio channel quality will fluctuate, or fade, as users move. Fading complicates the resource allocation, but channel prediction may alleviate this problem. A flexible and computationally inexpensive state space representation of fading channels is here used in conjunction with a Kalman filter, operating on special-purpose reference signals, to track and predict fading OFDM channels.

The thesis investigates key design and performance aspects of such estimators. Taking a probabilistic approach, we interpret the output of the Kalman filter as a full representation of a state of knowledge about the fading channels, given whatever information is at hand. For systems analysis, this permits conclusions to be drawn about channel estimation and prediction performance based on only vague information about the fading characteristics of the channel rather than on actual channel measurements. This is an alternative to conducting classic simulation studies. Various reference signal designs are studied and good design choices are recommended. Superimposed reference signal schemes are also proposed for and evaluated in cases where multiple signals are received, e.g. in multi-user (MU), multi-input multioutput (MIMO), or coordinated multi-point (CoMP) settings. By using time-varying reference signals, channel estimation and prediction performance is shown to be improved considerably in crowded frequency bands. The variation of prediction performance with prediction range and Doppler spectrum characteristics is investigated. For link adaptation, we derive the appropriate metric on which adaptation decisions should be used. The probability density function for this metric is derived for general MIMO channels. Link adaptation is studied for a single link system when channel prediction and estimation errors are present, both for uncoded systems and systems using large block codes with soft decoders. Various aspects of channel model acquisition are addressed by conducting studies on measured channels. Owing to the use of special matrix structures and fast convergence to time-invariant or periodic solutions, we find the Kalman filter complexity to be reasonable for future implementation. Finally, expressions for the impact of modelling errors are derived and used to study the impact of modelling errors on channel prediction performance in some example cases.

Keywords: channel estimation, channel prediction, OFDM, Kalman filtering

Daniel Aronsson, Department of Engineering Sciences, Signals and Systems Group, Box 528, Uppsala University, SE-75120 Uppsala, Sweden.

© Daniel Aronsson 2011

ISBN 978-91-506-2194-5

urn:nbn:se:uu:diva-146528 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-146528>)

Sammandrag

Antalet enheter som ansluter till Internet via mobila datanät växer snabbt och efterfrågan på högre dataöverföringshastigheter ökar. Trådlösa bredbands-system baserade på överföringstekniken Orthogonal Frequency Division Multiplexing (OFDM) införs nu för att uppfylla dessa krav. I trådlösa fleranvändarsystem måste den tillgängliga bandbredden delas effektivt mellan flera användare. Dessutom kommer mobila användare som rör sig genom det stående vågmönster som skapas av den utsända radiovågen från basstationer att erfara att den mottagna signalstyrkan varierar, *fäddar*, över tid. Frågan om hur resurser ska fördelas mellan användare kompliceras av fädning, men problemet kan till viss del förenklas genom att låta användarna förutsäga eller *prediktera* den fädande radiokanalen. Kanalprediktion är det huvudsakliga temat för denna avhandling.

En flexibel och beräkningsmässigt relativt snål s.k. tillståndsrepresentation av fädande radiokanaler används här i kombination med ett kalmanfilter, vilket med hjälp av speciella referenssignaler används både för att prediktera och följa fädande OFDM-kanaler. Avhandlingen undersöker centrala design- och prestandaaspekter för denna typ av kanalföljare och -prediktorer. Resultaten från kalmanfiltret tolkas här som en fullständig representation av den kunskap om de fädande kanalerna som brusiga mätningar av referenssignalerna ger. Som ett alternativ till att utföra klassiska simuleringsstudier tillåter detta oss att beräkna kanalföljarens och kanalprediktorns prestanda baserat endast på vag information om fädningens statistiska egenskaper.

Olika konfigurationer för referenssignalerna studeras och rekommendationer för goda designval presenteras. Det ramverk för kanalmodellering som föreslås tillåter att referenssignaler från olika sändarantennor överlagras på varandra i den mottagna signalen. Detta medför att andelen signalbandbredd som upptas av referenssignaler kan hållas på en rimlig nivå, trots att antalet signaler som samsas om frekvensutrymmet är stort. Vi undersöker typer av system där flera signaler tas emot, såsom t.ex. fleranvändarsystem, flerantennsystem (MIMO) eller koordinerade flerpunktsändningar (CoMP), och studerar hur olika val av referenssignaler påverkar prediktionsprestanda. Genom att variera referenssignalerna över tid kan kanalföljnings- och prediktionsprestanda förbättras avsevärt i fall då många signaler måste samsas om samma frekvensband. Vi undersöker även hur prediktionsprestanda beror av kanalens dopplerspektrum samt prediktionshorisont.

Vi beaktar länkadaptation och definierar det mått som bör ligga till grund för länkadaptionsbeslut. Sannolikhetsfördelningen för detta mått härleds för generella MIMO-kanaler. Realistiska kanalföljare och kanalprediktorer ger alltid upphov till ett visst mått av felgissningar. Länkadaptionsprestanda

utvärderas i ett enanvändarsystem där både skattningsfel och prediktionsfel föreligger. Både okodade system och kodade system som använder sig av mjuka avkodare studeras.

Vi studerar mätningar av färdande radiokanaler uppmätta vid låg fordonshastighet i stadsmiljö. Studien görs med avseende på kanalernas tidsdynamik och frekvensegenskaper, liksom på deras förmåga att följas samtidigt då överlagrade referenssignaler används. Baserat på dessa resultat ges rekommendationer för hur modellparametrar för kanalmodeller bör skattas.

Den numeriska komplexitet för de föreslagna kalman-baserade algoritmerna utvärderas. Goda konvergensgenskaper i kombination med att speciella matrisstrukturer används ger en komplexitet som ligger på en nivå som är rimlig för framtida implementering av de föreslagna algoritmerna.

Om de kanalermodeller som används inte beskriver de verkliga kanalernas färdningsegenskaper på ett tillfredsställande sätt så kommer kanalprediktionsprestanda att avta. Vi härleder teoretiska uttryck för effekten av sådana modellfel och studerar några konkreta fall.

At dotrom mine

Contents

1	Introduction	1
1.1	Approaches to channel prediction	2
1.2	Outline	8
2	Wireless communications	11
2.1	A brief historical overview	11
2.2	The usefulness of prediction	15
2.3	Channel models	17
2.3.1	Statistical characterization	18
2.3.2	Two channel models	20
2.4	OFDM	22
2.4.1	Dimensioning and design of an OFDM system	24
2.4.2	Equalization	25
2.4.3	Training	27
3	Linear filtering and inference theory	29
3.1	A channel model	29
3.2	Bayesianism	31
3.2.1	Probability	31
3.2.2	Tools in probability theory	36
3.2.3	Priors	38
3.2.4	How to make a decision	38
3.2.5	Model selection	39
3.2.6	The methods of frequentism and Bayesianism	41
3.2.7	Why Bayesianism?	43
3.3	The MMSE solution	44
3.4	Inference based on state space models	46
3.4.1	Optimal estimation	47
3.4.2	A geometric formulation	49

3.4.3	Prediction and smoothing	54
3.5	Summary	55
3.A	Some matrix results	57
3.A.1	Block triangular factorization	57
3.A.2	The matrix inversion lemma and variants	57
3.B	Properties of the Gaussian distribution	58
3.B.1	Definition	58
3.B.2	Marginalization	58
4	Modelling MIMO-OFDMA systems	61
4.1	Introduction	61
4.2	System model	63
4.2.1	Autoregressive modelling	67
4.2.2	State space modelling	69
4.2.3	Multipath channel model	72
4.2.4	Multi-input channel model	78
4.2.5	A few remarks	80
4.3	Channel inference	81
4.3.1	The Kalman filter recursions	81
4.3.2	The periodic filter	83
4.3.3	Modelling channel estimates	84
4.3.4	Optimal channel gain prediction	85
4.3.5	A special case: The GCG algorithm	86
4.4	Numerical complexity	86
4.5	Summary and discussion	88
4.A	Some useful distributions	89
4.A.1	Definitions	89
4.A.2	Change of variables	90
4.B	Posterior distribution for the channel power	90
4.C	Alternative formulations of the non-central χ^2 -distribution	91
4.D	Iterative channel estimation	92
4.E	A simplified prediction model	93
4.F	Alternative KF formulations	94
4.F.1	The CKMS recursions	94
4.F.2	Array algorithms	96
4.G	Convergence of the periodic filter	96
4.H	Model matrix structures	100
5	A channel estimation case study	101
5.1	Introduction	101
5.2	System model	102
5.3	Pilot-aided channel estimation	106

5.3.1	Non-smoothed Kalman filter	107
5.3.2	Smoothed Kalman filter	108
5.3.3	B-MMSE	108
5.3.4	Criterion	109
5.4	Experimental results	110
5.4.1	Impact of pilot pattern	112
5.4.2	Impact of UE velocity and fading statistics	112
5.4.3	Impact of block allocation choice	112
5.5	Conclusions	113
6	The OFDMA uplink design	121
6.1	Introduction	121
6.2	System model	124
6.3	Channel prediction for multiple users	128
6.3.1	Filter width and Doppler frequency	128
6.4	Pilot strategy	130
6.4.1	Overlapping versus dedicated pilots	131
6.4.2	Time-varying pilots	134
6.5	The impact of Doppler spectra on channel prediction	137
6.5.1	Criterion	138
6.5.2	Studies on different Doppler spectra	138
6.6	Numerical complexity	141
6.7	Discussion and conclusions	142
6.A	Further studies	145
6.A.1	Channel estimate interpolation	145
6.A.2	Pilot design	147
7	Link adaptation for uncertain channel state information	155
7.1	Introduction	155
7.1.1	Decision metric	157
7.2	Decision metrics using channel prediction	159
7.2.1	Channel model	159
7.2.2	The distribution of a future filtered estimate	162
7.2.3	The effective noise at detection	165
7.2.4	Predicted error rate	167
7.2.5	Mutual information for predicted channels	172
7.3	Adaptation of local CM formats	174
7.4	Joint adaptation of local CM formats and outer code rate	176
7.4.1	Perfect channel state information	179
7.4.2	Imperfect channel state information	183
7.5	Summary	185
7.6	Acknowledgements	186

7.A	Signal/noise correlation for decision feedback channel estimation	187
7.B	Rate limit optimization	188
8	Studies on measured channels	191
8.1	Introduction	191
8.2	Measurements	192
8.3	Time dynamics	193
8.3.1	Parameter estimation method	197
8.3.2	Model order	200
8.3.3	Subsampling	201
8.3.4	Parameter estimation in noise	204
8.3.5	Estimation on multiple subchannels	205
8.3.6	Segmentation	206
8.4	MIMO channels	207
8.4.1	Frequency correlation	208
8.4.2	SNR	210
8.4.3	Experiment	211
8.5	Filter convergence	212
8.6	CoMP complexity case study	215
8.7	Concluding remarks	217
8.8	Acknowledgements	218
8.A	Subsampling	219
9	Modelling errors	221
9.1	Introduction	221
9.2	Theory	222
9.2.1	Model	222
9.2.2	Modelling errors with structural consistency	223
9.2.3	Modelling errors with structural inconsistency	225
9.3	A few case studies	226
9.3.1	Post-smoothing of channel estimates	227
9.3.2	Model order sensitivity	229
9.3.3	Tap drift	231
9.3.4	Impulse response length	232
9.4	Summary	233
9.A	Filtered estimates	235
	Bibliography	237

Acknowledgements

First of all, I would like to express my deepest gratitude towards my supervisors Mikael Sternad and Anders Ahlén, for giving me the opportunity and for providing an inspiring work atmosphere. Thank you for always taking the time to discuss ideas and problems. I have enjoyed working with you.

Over the years, I have enjoyed the good company of all my friends and colleagues at the Signals&Systems group. Two of these friends have been especially important to me. Erik Björnemo and Mathias Johansson, thank you for being such good friends. I miss the Bayesian years, when we were young and angry and wanted to change the world. I hope that someday we will find the time and energy to rebel again.

Thank you, all members of the CoRe group, for fun conversations at the coffee table, and for once a week handing out free mobile phones. It is much appreciated.

Together with my supervisors, Rikke Abildgaard Olesen was very helpful in reading and commenting on the manuscript. Thank you.

The financial support, for which I am grateful, has in part been provided by the Swedish Foundation for Strategic Research (SSF) via the Wireless IP project, VINNOVA, and the EU FP6 WINNER and WINNER II integrated projects. I would also like to mention the current EU FP7 ARTIST4G project, which provides a source for ideas and inspiration.

Magnus Rahm and Johan Lundin, my friends at Ellära, thank you for being such fun and genuinely caring friends.

Mom and dad, Ruth and Kurt Aronsson, I am very lucky to have parents like you. Thank you for always supporting me in my decisions and for all the help we get from you.

Friends and family, I have neglected you of late, but that is about to change. I have missed you.

Finally, thank you, Sara, Linn, and Ylva, for everything. Sara, love of my life, I know these last months have been very tough on you too, but you always greet me with a smile when I come home from a late evening at the office. Thank you for all your love and support. I love you.

*Daniel Aronsson
Uppsala, February 2011*

Notation and symbols

$\mathbf{A}, \mathcal{A}, \Lambda$	Boldface capital letters, caligraphic capital letters, and Greek capital letters denote matrices.
$\mathbf{a}, \boldsymbol{\alpha}$	Boldface letters denote column vectors.
A, a, α	Letters in normal font denote scalars or sets.
\mathbf{A}^*	The conjugate transpose of \mathbf{A} .
\mathbf{A}^T	The transpose of \mathbf{A} .
\mathbf{A}^{-*}	The conjugate transpose of the inverse of \mathbf{A} .
$\mathbf{A} > 0$	The matrix \mathbf{A} is positive definite.
$ \mathbf{A} < 1$	All eigenvalues of \mathbf{A} are strictly inside the unit circle.
$\text{diag}(a_1, a_2, \dots)$	A diagonal matrix with diagonal elements a_1, a_2, \dots
$\text{diag}(\mathbf{a})$	A diagonal matrix whose diagonal is given by the vector \mathbf{a} .
$\text{diag}(\mathbf{A}_1, \mathbf{A}_2, \dots)$	A block-diagonal matrix with diagonal blocks $\mathbf{A}_1, \mathbf{A}_2, \dots$
$\text{circ}(\mathbf{a}^T)$	A circulant matrix, defined in (2.4.2), whose first row is \mathbf{a}^T .
\mathbf{I}_α	The $\alpha \times \alpha$ identity matrix.
$ A $	The cardinality of a set A .
$\mathbf{1}_{\alpha \times \beta}$	A $\alpha \times \beta$ matrix containing only ones.
$\mathbf{0}_{\alpha \times \beta}$	A $\alpha \times \beta$ matrix containing only zeros.
$\mathbf{A} \odot \mathbf{B}$	Element-wise multiplication.
$\mathbf{A} \oslash \mathbf{B}$	Element-wise division.
$\mathbf{A}[i, j]$	Element $\{i, j\}$ of the matrix \mathbf{A} .
$\mathcal{CN}(\mathbf{a}; \hat{\mathbf{a}}, \mathbf{R}_a)$	The multivariate circular symmetric complex Gaussian distribution with free variable \mathbf{a} , mean value $\hat{\mathbf{a}}$, and covariance matrix \mathbf{R}_a , defined in Appendix (3.B.1).
$J_0(a)$	The zeroth order Bessel function of the first kind.
$I_0(a)$	The zeroth order modified Bessel function of the first kind.
$\chi^2(z; \hat{a} ^2, \sigma_a^2)$	The non-central χ^2 -distribution with two degrees of freedom, defined in Appendix (4.A).

$p(\mathbf{a} D, I)$	The probability density function of \mathbf{a} given data D and information I .
$E\{\mathbf{a} D, I\}$	The expected value of \mathbf{a} given data D and information I . $E\{\mathbf{a} D, I\} = \int \mathbf{a}p(\mathbf{a} D, I)d\mathbf{a}$, where the integration is taken over the entire domain of \mathbf{a} .
δ_{ij}	The Kronecker delta function. $\delta_{ij} = 1$ if $i = j$ and zero otherwise.
$\delta(a)$	The Dirac delta distribution. $\int_{\mathbb{R}} \delta(a)da = 1$ and $\delta(a) = 0$ if $a \neq 0$.
$\mathcal{L}\{\mathbf{a}_0, \dots, \mathbf{a}_t\}$	The linear vector space spanned by $\mathbf{a}_0, \dots, \mathbf{a}_t$.
$a \in A$	a is a member of the set A .
j	The imaginary unit $\sqrt{-1}$.

Abbreviations

3GPP	Third-Generation Partnership Project
AE	Antenna Element
AR	Auto-Regressive
ARMA	Auto-Regressive Moving Average
AWGN	Additive White Gaussian Noise
BA	Block Allocation
BER	Bit Error Rate
B-EFDMA	Block Equidistant Frequency Division Multiple Access
B-IFDMA	Block Interleaved Frequency Division Multiple Access
BICM	Bit-Interleaved Coded Modulation
BPSK	Binary Phase Shift Keying
BS	Base Station
CKMS	Chandrasekhar Kailath Morf Sidhu
CM	Coded Modulation
CoMP	Coordinated Multi-Point
CP	Cyclic Prefix
CRC	Cyclic Redundancy Check
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CWER	Codeword Error Rate
DARE	Discrete Algebraic Ricatti Equation
DFT	Discrete Fourier Transform
DL	Downlink
EDGE	Enhanced Data rates for GSM Evolution
EM	Expectation Maximization
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FER	Frame Error Rate
FFT	Fast Fourier Transform
GCG	General Constant Gain
GPRS	General Packet Radio Services
GSM	Groupe Spécial Mobile <i>or</i> Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat reQuest
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
ICE	Iterative Channel Estimation
i.i.d	independent and identically distributed

IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
KF	Kalman Filter
LDPC	Low-Density Parity-Check
LMS	Least Mean Squares
LOS	Line-Of-Sight
LS	Least Squares
LTE	Long Term Evolution
MAP	Maximum A Posteriori
MEM	Maximum Entropy Method
MI	Multiple-Input
MI	Mutual Information
MI-ACM	Mutual Information based Adaptive Modulation and Coding
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MMSE	Minimum Mean Square Error
MQAM	M-ary Quadrature Amplitude Modulation
MSE	Mean Square Error
MUSIC	Multiple Signal Classification
NLOS	Non-Line-Of-Sight
NMSE	Normalized Mean Square Error
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
pdf	probability density function
PDP	Power Delay Profile
PSAM	Pilot Symbol Assisted Modulation
QPSK	Quadrature Phase Shift Keying
RB	Resource Block
RCP-BLDPC	Rate-Compatible Punctured Block-circulant Low Density Parity Check
RLS	Recursive Least Squares
SISO	Single-Input Single-Output
SNR	Signal-to-Noise ¹ Ratio
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
UE	User Equipment
UTRA	Universal Terrestrial Radio Access
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
WINNER	Wireless World Initiative New Radio

¹Interference is assumed to be included in the noise.

Introduction

Consider the *channel prediction* situation illustrated in Figure 1.1. Two mobile radio receivers, e.g. pedestrians carrying mobile phones or vehicles equipped with radio receivers, receive a signal transmitted by a central antenna. The transmitted wave is assumed to occupy a very narrow frequency band, which means that it is practically a single tone. The wave bounces off the ground, trees, buildings, and so forth, and interferes with itself so that a standing wave pattern is formed in space. Due to the generally complex geometry of the surroundings, the standing wave will feature a lot of irregularities. The peaks in the figure indicate points where the standing wave pattern interferes constructively, and troughs mark destructive interference. As the receivers move about, they move through the peaks and the troughs and will therefore experience varying quality of reception. Some radio transmission strategies strive to counteract the fluctuations of the radio channel, others are designed to exploit them. For example, the centralized antenna may use *opportunistic scheduling*, always transmitting to the user that experiences the best channel quality. That way, the average quality of reception for each receiver will be higher than if the receivers would blindly share the resource, e.g. by using a time slotted schedule and taking turns in using the time slots.

To allow for opportunistic scheduling, the mobile receivers need to signal their respective received signal strengths to the system. A central problem is that there is an inevitable delay between the time when this information reaches the central system, and the time when data is actually transmitted to one of the receivers. During this delay, a receiver may move from a good spot (peak) to a bad spot (trough), or vice versa, as indicated in the figure. This issue motivates the idea of *prediction*; it is not the *present* channel quality that needs to be reported, but rather the channel quality a short time period into the future.

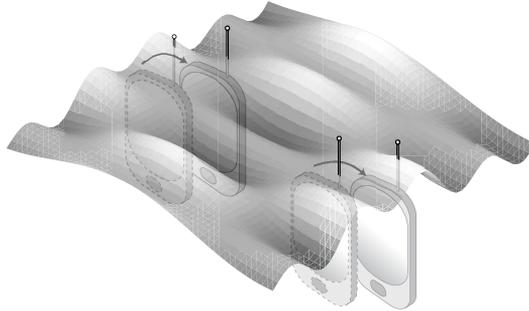


Figure 1.1: Two mobile receivers (equipped with out-of-date antennas for illustration) move through the interference pattern created by a wave transmitted from a centralized antenna.

1.1 Approaches to channel prediction

Assuming that the central antenna is transmitting a single tone of constant unit amplitude, a narrowband signal y_t received in additive noise v_t by a mobile user equipment (UE) may be modelled in the baseband by

$$y_t = h_t + v_t, \quad (1.1.1)$$

where the magnitude of the envelope of the fluctuating radio channel h_t varies with position (and thereby with time), as illustrated by the wave pattern in Figure 1.1. For reasons mentioned above, we want to *predict* the channel, say L time steps into the future. A way to model the predicted channel is to say that it is adequately described by a weighted sum of the n most recent channel measurements [1], [2], [3], [4]:

$$\hat{h}_{t+L} = a_0 h_t + a_1 h_{t-1} + \dots + a_{n-1} h_{t-n+1}, \quad (1.1.2)$$

where the hat symbol ($\hat{\cdot}$) indicates that we refer to a prediction of the channel and not the channel itself. If \hat{h}_{t+L} is identical to the true channel h_{t+L} , then (1.1.2) is autoregressive (AR) in the channel coefficients. One therefore refers to prediction methods based on the model (1.1.2) as *AR methods*. The parameters $\{a_i\}$ in (1.1.2) are time varying if the receiver is mobile, and they must be appropriately tuned and tracked to produce a good channel prediction. Assuming that the true channel $\{h_t\}$ is available, it has been shown [5],

[6] that the AR model parameters can be estimated to high precision, so that the difference between the predicted channel \hat{h}_{t+L} and the true channel h_{t+L} is small for moderate values of L . To determine the time-varying parameters $\{a_i\}$ in (1.1.2), block-wise *Minimum Mean Squares Error* (MMSE) estimation may be used [3],[4], or adaptive methods such as *Least Mean Squares* (LMS) [2],[6], or Wiener LMS [7], or *Recursive Least Squares* (RLS) [6] may be employed to track the AR coefficients.

By the z-transform, the model (1.1.2) can be written on polynomial form [8], as a finite impulse response filter

$$z^L \hat{h}_t = \left(\sum_{i=0}^{n-1} a_i z^{-i} \right) h_t. \quad (1.1.3)$$

A special case, that is commonly separated from AR methods, arises when the polynomial $1 - z^{-L} \sum_{i=0}^{n-1} a_i z^{-i}$ has all its zeros on the unit circle. The fading channel is then modelled as a *sum of sinusoids* [5], [9], [10], [11], [12], [13]. Sinusoid methods usually differ from AR methods in that the AR parameters (the complex sinusoids in the case of sinusoidal modelling) are estimated through subspace methods. A common subspace method used in this context is *Estimation of Signal Parameters via Rotational Invariance Techniques* (ESPRIT) or slight modifications thereof [14], [15], [9], [11], [12]¹. Among other algorithms for estimating the parameters for complex sinusoids, we may mention the Multiple Signal Classification (MUSIC) algorithm [11], [12], and root-MUSIC [17]. Eycoez *et al* [1] formulates the prediction problem on AR form but uses the Maximum Entropy Method (MEM) to estimate the parameters. A summary of performances for both AR and sinusoid methods can be found in [5]. That investigation is undertaken both on synthetic data and on real measurements.

Once the AR model parameters have been obtained, they are used in the channel predictor. An obvious issue with the model (1.1.2) is that the true fading channel coefficients $\{h_t\}$ are not available to the predictor. In practice, noise reduction of noisy measurements $\{y_t\}$ has to be carried out to produce a sequence of estimates that will hopefully resemble the true channel. Some works [1],[2] consider a simplified case where the true channel coefficients are assumed known. Others use low-pass filtering [18] or Wiener filtering [4] to clean measurements from noise.

The noise reduction is a pragmatic intermediate step between the actual available data $\{y_t\}$ and the idealized model (1.1.2). The $\{h_t\}$ in (1.1.2) are

¹Because of a somewhat high complexity of the original ESPRIT formulation, some works look at complexity-reduced versions of ESPRIT [10], [16]. A special utilization of the ESPRIT is one in which the algorithm is employed in two tiers; in a first pass the time-delays and the corresponding complex amplitudes of the sinusoids are estimated, followed by Doppler frequency estimation in a second pass [10], [16].

in practice channel *estimates* and are therefore functions of measurements $\{y_i\}$. An approach that is more straightforward than (1.1.2) to formulate the predictor would be to say that the predictor should be formed by a weighted sum of *all* past measurements:

$$\hat{h}_{t+L} = \sum_{i=0}^t c_i y_i. \quad (1.1.4)$$

More generally, if multiple channels are to be taken into consideration, we may consider predicting a vector-valued channel and expressing it as a weighted sum of past vector-valued measurements $\{\mathbf{y}_i\}$:

$$\hat{\mathbf{h}}_{t+L} = \sum_{i=0}^t \mathbf{M}_i \mathbf{y}_i. \quad (1.1.5)$$

This is the approach that we will take in the present thesis. By appropriately tuning the matrix coefficients $\{\mathbf{M}_i\}$, we may derive the best possible linear predictor given all available measurements. But how do we find the weights $\{\mathbf{M}_i\}$?

Before addressing this question, we should consider another issue; the channel coefficients, contained in the vector \mathbf{h}_t , are never themselves of immediate interest. Rather, it is some parameter *relating* to the channel coefficients that we ultimately wish to infer, such as the maximum data transfer rate that can be used under the current channel conditions, without the level of distortion of the signal exceeding some predefined level.

A channel predictor that only produces *point-wise* predictions of the channel, as (1.1.2) does, will be inadequate under these circumstances, because it does not provide any measure of uncertainty about the predicted parameter. For example, let us put ourselves in the position of a point estimator that is to infer a real-valued parameter x . The measurements, the *evidence*, give us reason to believe equally strongly in the proposition that x is less than zero, as in the proposition that x is larger than zero, and so we report the value 0 to be an arguably reasonable point estimate of x . But now imagine that the person receiving this point estimate is interested in the value of x^2 rather than the value of x . Unless we are certain that x is very close to zero, the point estimate 0 is of little use to him because 0^2 is not a good estimate of x^2 . On the other hand, had we reported an accurate and full *representation of our knowledge* of x given the evidence at hand, then the recipient of that information could transform that knowledge into an equally accurate representation of a state of knowledge about any *function* of x .

Therefore, instead of directly calculating a specific estimate of \mathbf{h}_{t+L} , we want to acquire a complete representation about our state of knowledge about

\mathbf{h}_{t+L} , i.e. we want to obtain the probability density function (pdf), $p(\mathbf{h}_{t+L})$, of \mathbf{h}_{t+L} :

$$\begin{array}{l} \text{full state of knowledge about } \mathbf{h}_{t+L}, \\ \text{given measurements to time } t \end{array} = p(\mathbf{h}_{t+L}|\mathbf{y}_0, \dots, \mathbf{y}_t).$$

In principle, this pdf can then be used to derive the pdf for any function $f(\mathbf{h}_{t+L})$ of \mathbf{h}_{t+L} :

$$p(\mathbf{h}_{t+L}|\mathbf{y}_0, \dots, \mathbf{y}_t) \xrightarrow{\text{change of variables}} p(f(\mathbf{h}_{t+L})|\mathbf{y}_0, \dots, \mathbf{y}_t). \quad (1.1.6)$$

In this thesis we take an information theoretic approach to probability theory. Specifically, we adhere to the so called *Bayesian* school, that considers probability theory as an extension to logic. Bayesianism offers a single method for deriving probability density functions, regardless of the nature of involved parameters. In contrast, classical probability theory constitutes a large set of methods for calculating estimates. These methods may produce mutually inconsistent results and may not provide the whole pdf for a parameter.

To derive a pdf is however in many cases an arduous task of assigning prior distributions, applying Bayes theorem, integrating over irrelevant parameters, and changing variables. We will simplify this problem by using models that restrict the pdf:s of \mathbf{h}_t and \mathbf{y}_t to belong to a small class of functions.

Specifically, we will let the measurement \mathbf{y}_t be a linear mapping of the channel \mathbf{h}_t with added white Gaussian noise. In this thesis, we consider Orthogonal Frequency Division Multiplexing (OFDM) wireless transmission systems. In OFDM, data is transmitted over many parallel *subchannels*, each subchannel carrying data at such a low symbol rate that no intersymbol interference occurs. Channel prediction performance can be improved by taking several parallel subchannels into account at once. Instead of considering only a scalar channel, we use a vector-valued measurement signal

$$\mathbf{y}_t = \Phi_t \mathbf{h}_t + \mathbf{v}_t, \quad (1.1.7)$$

that has, say, w elements. The vector \mathbf{h}_t may hold w parallel subchannels from a single OFDM channel, but could also comprise fading channel coefficients from u different OFDM channels, each constituting w parallel subchannels, so that \mathbf{h}_t has length uw . By an appropriate design of the w -by- uw matrix Φ_t , the u channels are summed, together with the noise \mathbf{v}_t , in the vector \mathbf{y}_t in a way that they can be separated and predicted individually by studying the present and past w -vectors $\{\mathbf{y}_t\}$.

This flexibility is useful in modern multiuser systems for two reasons: first, in a multiuser setting, it allows the receiver (here, the base station)

to separately track the respective channels of multiple users whose signals overlap so that a superposition of all incoming signals is received. Second, in a so called Coordinated Multi-Point (CoMP) transmission, it allows the receiver (here, the mobile user equipment) to simultaneously track overlapping signals from separate base stations.

It is also necessary to model the fading behaviour of the channels. We will assume that the dynamics of the respective channel coefficients contained in \mathbf{h}_t can be modelled as AR processes of finite order. As we shall see, this allows us to set up a discrete-time state-space representation of the channels.

Assume that measurements up to and including time $t = T$ are available. When the total set of measurements $\mathbf{y} = [\mathbf{y}_0^T \dots \mathbf{y}_T^T]^T$ is a linear mapping of the corresponding set of channel coefficient vectors $\mathbf{h} = [\mathbf{h}_0^T, \dots, \mathbf{h}_T^T]^T$ plus additive Gaussian noise and when the statistical properties of \mathbf{h} and the noise are known, then it is straightforward to calculate an expression for the pdf of \mathbf{h} given \mathbf{y} , and prediction is equally straightforward. However, this computation is burdensome and the number of arithmetic operations required for each added measurement vector grows as the time T increases.

In 1960, Rudolf Kalman published a paper [19] showing that the complexity of determining the pdf of e.g. a channel vector \mathbf{h}_T given old measurements \mathbf{h} , can be kept at a constant level, if the process $\{\mathbf{h}_t\}$ can be modelled on linear state space form and the measurements are a linear regression in \mathbf{h}_t .²

Due to the linearity of the model and the Gaussian distribution of the noise processes, the conditional pdf:s of \mathbf{h}_t will be Gaussian. Hence, a pdf $p(\mathbf{h}_t | \mathbf{y}_0, \dots, \mathbf{y}_t)$ is uniquely determined by a mean value $\hat{\mathbf{h}}_{t|t}$ and a covariance matrix $\mathbf{R}_{h,t|t}$. The estimator, denoted the Kalman filter, operates in a recursive fashion, updating the pair $(\hat{\mathbf{h}}_{t|t}, \mathbf{R}_{h,t|t})$ as new evidence arrives:

$$(\hat{\mathbf{h}}_{t|t}, \mathbf{R}_{h,t|t}) \xrightarrow{\text{new data } \mathbf{y}_{t+1}} (\hat{\mathbf{h}}_{t+1|t+1}, \mathbf{R}_{h,t+1|t+1}). \quad (1.1.8)$$

This recursion weighs together old measurements in an optimal manner, to produce the parameters $(\hat{\mathbf{h}}_{t|t}, \mathbf{R}_{h,t|t})$ that represent the pdf for \mathbf{h}_t given all available measurements at time t . The Kalman filter hence provides both the weights $\{\mathbf{M}_i\}$ in (1.1.5), and the full state of knowledge of the channel vector, given some set of data.

The Kalman filter has been used extensively in the OFDM channel estimation and prediction literature. Most works use only the point-wise channel estimates provided by the filter. Although the numerical complexity of the Kalman is only linear in the number of measurements, the complexity per filter update can still be extensive. Many authors therefore use models of low order for the respective subchannels, such as random walk models [20] or

²Actually, the condition is that \mathbf{y}_t is a linear regression in the so called state vector from the state space model of \mathbf{h}_t , as shall be evident in Chapter 4.

AR1 models [21], [22], [23], [24], [25], although others use models of higher order [26], [27]³. The increased interest in recent years in multiple-input multiple-output (MIMO) OFDM systems has led several authors to consider MIMO OFDM channel tracking with Kalman filters [23], [24], [25], [28].

The numerical complexity of the Kalman filter is generally considered to be a problem. However, if constant model parameters are used, the Kalman filter will quickly settle to a stationary state. Once this condition has been reached, many computationally demanding components in the Kalman updating procedure can be turned off until the model parameters change, and this happens, as we shall see, rather seldom.

The focus so far has been on channel prediction, but in the Kalman filter, estimation and prediction are closely related. In fact, channel estimation will fall out as a by-product of channel prediction. Common methods for OFDM channel estimation is to use Wiener filters or 2-dimensional block filters spanning a fixed time-frequency region [29], [30]. Once the Kalman filter has converged, it will be identical to a Wiener filter. Further, if the coherence time of the channels is short, then the Kalman/Wiener filter will practically have a finite memory, making it essentially equal to a block filter. In settings with constant model parameters, the Kalman filter can therefore be seen as an adaptive filter that automatically calculates optimal filter weights from given channel models.

In this thesis, we will use the Kalman filter as an inference engine. Given prior information about the time dispersion and the fading statistics of the channel, and possibly also noisy measurements, the Kalman filter produces complete representations of the knowledge about a present and/or a future channel that can be deduced from the given information. Our motivation for doing so can be subdivided into four topics:

Inference from vague information

Bayesian inference transforms information at hand into a logically equivalent statement about parameters of interest (“logic” is here meant in the sense of the extended logic to be presented in Chapter 3). In the present context, this means that we do not need any channel data, measured or synthetic, in order to infer how well a system will perform in terms of channel estimation and prediction. From only vague information about e.g. channel fading statistics, we can therefore directly calculate channel estimation and prediction performance metrics.

³The filter implementation may also be simplified by modelling the parallel subchannels as uncorrelated [20], [26]. The lack of correlation between the estimated subchannels can then be compensated for by combining the filter outputs in an optimal manner [26].

Superimposed received signals

The quality of the output of a Kalman filter, i.e. the level of uncertainty associated with the output, will depend on the statistical properties of the channel that is observed, but also on how the measurement equation (1.1.7) is constructed. By an appropriate selection of the regressor matrix Φ_t , several channels may be superimposed on one another and tracked simultaneously. It is however up to the system designer to choose the regressor matrix in such a way that the inferences made about the respective channels are of as high quality as possible.

Soft information in link adaptation

Transmitted information may be encoded to make the information more robust to distortions introduced by the fading channel and by noise. The pdf:s produced by the Kalman filter can be used by an efficient channel decoder to retrieve the transmitted message from the noisy received signal with fewer bit errors than if the decoder had only point-wise channel estimates to work with.

Modelling errors

In general, any type of information can be processed using Bayesian inference. For computational reasons, we here restrict the type of information that we feed to the inference engine (the Kalman filter), to statistical descriptions of the channel and possibly also noisy measurements of the channel. This framework does not allow us to express *uncertainty* about the channel's fading statistics. Whatever information about fading statistics we pass to the Kalman filter, it will consider this information to be absolutely accurate and draw conclusions about channel estimation and prediction accordingly. If the characteristics that we provide about the channel is incorrect, then so will the performance conclusions reported by the Kalman filter be. However, it is possible to calculate how a *misinformed* filter performs in terms of channel estimation and prediction. The misinformed filter believes that it has a fully accurate description of the channel's statistical properties, while in reality, the actual channel is described by a different model.

1.2 Outline

The thesis is outlined as follows:

Chapter 2 presents a short history of wireless technologies and puts channel prediction into context. A short survey of fading radio channel modelling and the basics of Orthogonal Frequency Division (OFDM) are also given as a prerequisite for later chapters.

Chapter 3 gives an overview of linear filter theory. We take an information theoretic approach in which results from calculations are interpreted as an agent's state of knowledge, given some specified set of data and background information. To make sense of this interpretation, we present an overview of an approach to probability theory known as Bayesianism. The Kalman filter is then derived using a geometric perspective, and formulas that will be useful in later chapters are given.

The material presented in this chapter has in part been presented in “*Channel Estimation and Prediction from a Bayesian Perspective*”,
D. Aronsson, Licentiate thesis, Uppsala University, 2007

Chapter 4. This chapter lays the foundation for coming chapters. In a step-by-step fashion, a state space representation of multiple-input fading channels is constructed. We use special matrix structures that permits easy scaling of the channels. The numerical complexity for Kalman filters operating on these state-space models is investigated and we find that the complexity is considerably alleviated by the use of the special matrix structures, as compared to a general Kalman filter. We also investigate some less conventional Kalman filter formulations but find that they are not useful in the present context.

The material presented in this chapter has in part been presented in “*Channel Estimation and Prediction from a Bayesian Perspective*”,
D. Aronsson, Licentiate thesis, Uppsala University, 2007

Chapter 5. Here we consider channel estimation in a special type of transmission scheme proposed within the WINNER project [31]. The objective has been to see how well channel estimation performs for different block sizes within the transmission scheme referred to as *Block Interleaved Frequency Division Multiple Access* (B-IFDMA). By exploring a large number of designs for the regressor matrix in the measurement equation, we find that blocks as small as $22 \text{ Hz} \cdot \text{s}$ can be used with adequate estimation performance.

The material presented in this chapter has in part been presented in “*Performance Evaluation of Memory-less and Kalman-based Channel Estimation for OFDMA*”,
D. Aronsson, T. Svensson and M. Sternad, IEEE Vehicular Technology Conference VTC-Spring 2009, and “Block Interleaved Frequency Division Multiple Access for Power Efficiency, Robustness, Flexibility and Scalability”, T. Svensson, T. Frank, T. Eriksson, D. Aronsson, M. Sternad and A. Klein, EURASIP Journal on Wireless Communications and Networking, Special Issue on 3GPP LTE and LTE Advanced, vol. 2009

Chapter 6. Channel prediction when multiple signals are received simultaneously is considered here. We study how to design special-purpose reference signals so as to facilitate the prediction of multiple channels. A special type, “time-varying pilots”, turns out to be important in cases where the number of estimated channel coefficients is larger than the number w of utilized subchannels. We also study the importance of the Doppler spectrum, the signal-to-noise ratio, and the channel frequency selectivity on channel prediction performance.

The material presented in this chapter has in part been presented in “*Kalman Predictor Design for Frequency-adaptive Scheduling of FDD OFDMA Uplinks*”, D. Aronsson and M. Sternad, *IEEE Conference on Personal, Indoor and Mobile Radio Communications (PIMRC) 2007*, and “*OFDMA Uplink Channel Prediction to Enable Frequency-adaptive Multiuser Scheduling*”, D. Aronsson and M. Sternad, *European Signal Processing Conference (EUSIPCO) 2007*

Chapter 7. We consider link adaptation for unknown predicted fading channels. We show how to incorporate uncertainties about the channel, both regarding the present channel (channel estimation) and the true future channel (channel prediction). It is emphasized that the degree of uncertainty that should appropriately be considered is that of the future channel *estimate*, and not the future channel itself. We consider uncoded as well as coded systems and address the problem of choosing the code rate for large code blocks spanning a large portion of the total frequency bandwidth. Since the optimal link adaptation strategy is then difficult to implement, a suboptimal scheme working for pedestrian velocities is suggested.

Chapter 8. When a Kalman filter is implemented as a channel estimator/predictor, a channel model needs to be constructed from measurements of the channel. In this chapter, we investigate various aspects of channel model parameter acquisition. These include the method to use for parameter acquisition, model order selection, block-size selection, noise suppression, and model acquisition for multiple channels. We also study the convergence rate of the Kalman filter in some typical scenarios.

Chapter 9. Here we derive formulas for one-step prediction performance when the Kalman filter is misinformed. A few case studies illustrate the utility of the formulas.

Wireless communications

2.1 A brief historical overview

The steadily increasing demand for mobile services has transformed early analog radio systems into the packet-based digital wireless systems we see today. Early radio systems used a single central antenna, but due to limited spectrum availability, the number of users such a systems could support was low [32]. A crucial step towards increasing capacity of large-coverage wireless multiuser systems was taken in the 1950's and 60's when the cellular system concept was developed; instead of using only a single central antenna, multiple centralized antennas connected through a *core network* can be deployed. Each such antenna is called a *base station* (BS). Taking advantage of the fact that received signal strength falls off with increasing distance, different user equipments (UEs) can communicate with separate base stations and thereby utilize the same radio resource (frequency, time slot, spreading code, etc.). The location of a BS in a cellular system is sometimes referred to as a *site*, and the coverage area served by each BS is called a *cell*. Having small cells means that many users within a given geographical area can share the same resource. On the other hand, small cells also require many sites, making the system expensive to deploy.

The concept of resource reuse generates interference among different UEs and BSs using the same resource. The *reuse factor* is defined as the ratio of the total number of cells to the number of cells providing a specific resource. The reuse factor needs to be chosen based on an optimization of total system capacity and performance that takes inter-cell interference into account. Figure 2.1 illustrates a few reuse patterns on a hexagonal grid. (It is common to model a site as a hexagon, since hexagons tile the plane. Alternatively, because each site is often equipped with three base station, each BS serving a cell in the form of a 120° sector, the three cells covered by a site can be

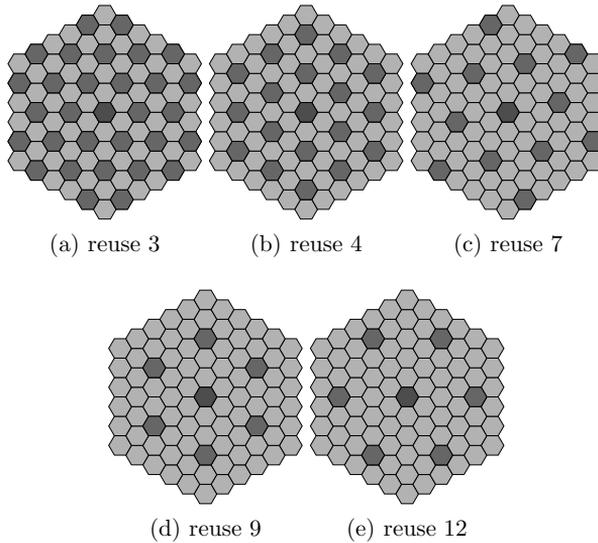


Figure 2.1: Regular reuse patterns

modelled as three connected hexagons.) The reuse factor is chosen as low as possible. This implies that the noise and interference experienced by a UE in a cellular system is normally dominated by the interference from other UEs, rather than by thermal noise. We then say that the system is *interference limited* rather than *noise limited*.

New technologies would soon be necessary to meet the demand for a further increase of capacity and support for more services. While first generation systems were analog, enabling voice traffic only, second generation systems introduced digital communication, which enabled, apart from simple voice traffic, data services such as e-mail and sms. In contrast to analog systems, where each user needs to be allocated an exclusive frequency band, digital systems allow more flexibility in the form of a number of different *multiple access* techniques, allowing multiple users to share the same channel. The most wide-spread of the second generation systems is GSM, which is based on *Time Division Multiple Access* (TDMA) in combination with *Frequency Division Multiple Access* (FDMA). GSM uses a number of separate radio channels. These channels are divided into time frames, each frame consisting of several time slots. A voice user is assigned a prespecified part of a frame, regardless of how many other users are sharing the channel. This ensures quality of service, since a user will always be guaranteed a constant throughput if the signal is received with sufficient power. On the other hand, the channel is only partially utilized in a system with few UEs.

Data networking paradigms can be divided into *circuit switching* and *packet switching*. Circuit switching provides an unbroken and dedicated link between sender and receiver, making it ideal for traditional voice traffic. Packet switching on the other hand, groups information bits, regardless of the type of information they represent, into packets which are then transferred over non-dedicated links. Packet switching is therefore suitable for systems that should support general data transfer. During the second half of the 1990's, GSM was enhanced with *General Packet Radio Services* (GPRS). In this standard, several time slots can be aggregated so that a user can use more than one time slot in a frame. Exploiting this simple technique of utilizing unused resources, GPRS increased the peak data rate from 9.6 kbps to about 140 kbps. Similar techniques were used in the American IS-136 standard.

The data rate was later increased further up to 348 kbps with the introduction of *Enhanced Data rates for GSM Evolution* (EDGE). EDGE uses higher order modulation and link adaptation to increase capacity. Measurements of the received *Signal-to-Noise Ratio* (SNR) is fed back to the transmitter and used for choosing coding and modulation (CM) formats [32],[33].

The third generation (3G) cellular systems, based on wideband *Code Division Multiple Access* (WCDMA), were introduced at the turn of the millennium. Target data rates for 3G were initially set to up to 2 Mbps for indoor users, up to 144 kbps for pedestrians, and up to 64 kbps for vehicular users, but today, these number are often widely exceeded in deployed systems.

A major step that boosted initial 3G performance was taken with *High Speed Packet Access* (HSPA), which is the joint name for *High Speed Downlink Packet Access* (HSDPA), introduced in Release 5 of the 3GPP¹/WCDMA specifications, and the *Enhanced Uplink*, introduced in Release 6. HSPA uses higher order modulation, channel-dependent scheduling and rate control, and fast *Hybrid Automatic Repeat reQuest* (HARQ) with soft combining. This provides data rates of 5.7 Mbps in the *uplink* (the link from the UE to the BS) and data rates of 14 Mbps in the *downlink* (the link from the BS to the UE). It also provides reduced delays and delay variations over the wireless links, which is important for high-speed data traffic.

3G evolution has since 2005 forked into two parallel tracks: HSPA Evolution and *Long Term Evolution* (LTE). The LTE standard [34] has fewer restrictions on backwards compatibility than HSPA and also addresses more complex spectrum situations. LTE is optimized for broadband data traffic, and is based on packet switching. The multiple access technique used in the LTE downlink is *Orthogonal Frequency Division Multiple Access* (OFDMA) which builds on *Orthogonal Frequency Division Multiplexing* (OFDM). OFDM

¹The *Third-Generation Partnership Project* (3GPP) develops the specifications for UTRA (the official name for the 3G standard described here) and GSM systems.

divides the radio resource into small time-frequency units. In OFDMA, these units can be distributed among users in a flexible way, which takes their channel variability over frequency and time into account. We return to OFDM in Section 2.4. LTE is sometimes identified as “almost” a fourth generation (3.9G) system. Among many improvements over earlier systems, LTE is specified to support up to 50 Mbps and 100 Mbps in the uplink and downlink, respectively, over a bandwidth of 20 MHz. It also provides further reductions of latencies and delays in the packet transmission.

Among other techniques, HSPA Evolution and LTE introduce multi-antenna support, which is the principal technology for increasing capacity in these systems. Multi-antenna techniques can be divided into three categories:

- **Spatial multiplexing** makes it possible to transmit several parallel data streams over the *Multiple Input Multiple Output* (MIMO) link, hence possibly increasing spectral efficiency several times. By appropriate signal processing at the transmitter and the receiver, the channel can be used to serve at most $\min(n_T, n_R)$ independent streams, where n_T and n_R are the number of transmitting and receiving antennas, respectively.
- **Diversity techniques** suppress fading by constructively adding base-band signals from multiple antenna elements. It is common to assume that the channels are Rayleigh fading [32]. The perceived channel after reception will feature more favourable statistical properties of the fading than Rayleigh fading, in the sense that the fading dips will be fewer and not as deep.
- **Beamforming** is a means for steering radiated energy in a prescribed range of directions (the *antenna beam*), or for making a receiving antenna array more sensitive to radiated energy coming from a certain range of directions.

Spatial multiplexing uses antenna arrays on both the transmitter and receiver sides, while diversity techniques and beamforming can be used on either or on both sides. All these techniques require some amount of information about the current radio channel quality (the *Channel State Information* (CSI)). Techniques employed on the transmitter side require *Channel State Information at the Transmitter* (CSIT), which in a *Frequency Division Duplex* (FDD) system, i.e. a system using different frequency bands for the uplink and the downlink, requires the receiver to signal CSI back to the transmitter. *Time Division Duplex* (TDD) system, on the other hand, alternately transmit and receive on the same frequency band. In a TDD system, the transmitter

may therefore use channel estimates obtained from recent transmissions over the reverse direction (the reverse link). This assumes channel reciprocity, i.e. that the channel is the same regardless of direction.

Many MIMO techniques, as well as link adaptation and scheduling, require CSI feedback. When a UE is mobile, the reported quality indicator will sometimes be outdated when the transmission takes place, owing to the transmission frame structure, the feedback delay, and the computational delays of the system. This is true for TDD as well as for FDD systems. Although link adaptation is employed in 3G and in GSM, current systems do not attempt to compensate for this effect. The alternative is to use channel prediction, so that the reported channel quality matches the quality experienced at the time of transmission as well as possible. In this thesis, we investigate channel prediction and estimation when optimal observers of the time varying channel is used.

2.2 The usefulness of prediction

In a cellular system with mobile users, base stations as well as UEs will normally experience rapid changes of their received channel quality. The idea with prediction is then for the receiver to, well in time for transmission, signal its future expected channel quality to the transmitter. The transmitter may then opportunistically schedule UEs and/or opportunistically choose transmission method, so as to increase link performance. However, channel predictions are always associated with a degree of uncertainty. How to take this uncertainty into account when making scheduling and link adaptation decisions is very difficult to say in general, because it depends on the scheduling algorithm and on the available transmission methods.

In a MIMO system with flat fading channels, i.e. channels that fade equally for all frequencies, the total n_R -by- n_T MIMO channel can be described by a matrix \mathcal{H} . The channel estimator calculates an estimate $\hat{\mathcal{H}}_{est}$ based on as recent noisy measurements as possible, while the predictor produces a prediction $\hat{\mathcal{H}}_{pred}$ which is based on somewhat older measurements. Different transmission techniques will place different requirements on $\hat{\mathcal{H}}_{est}$ and $\hat{\mathcal{H}}_{pred}$.

To use spatial multiplexing, the channel estimation $\hat{\mathcal{H}}_{est}$ has to be very accurate. Although $\hat{\mathcal{H}}_{est}$ is not available at the time of prediction, we will see that the *accuracy* of the estimate can be calculated beforehand. It is in this way possible to decide whether to use spatial multiplexing. The expected capacity and performance of spatial multiplexing is given by the eigenvalues of the matrix $\hat{\mathcal{H}}_{est}\hat{\mathcal{H}}_{est}^*$, whose distribution can, in principle, be calculated from the predictions [35] [36].

When the channel quality is too low for spatial multiplexing to be used, beamforming may still be employed. If the channel estimate is of high quality, then the link performance when combining beamforming at the transmitter and the receiver is dictated by the largest eigenvalue of $\hat{\mathcal{H}}_{est}\hat{\mathcal{H}}_{est}^*$, whose distribution can be derived from the predictions. Beamforming on the transmitter side requires some degree of CSIT. Optimal beamforming from n_T transmit antennas to a receiving antenna, in the sense that the received signal power is maximized under a constraint on transmitted power, is given by scaling the signal transmitted from the respective antennas with weights equal to \mathbf{h}^* , where \mathbf{h} is the n_T channel coefficients and $(\cdot)^*$ indicates complex conjugation. This requires the full feedback of (an accurate estimate of) the channel. To enable the use of a more modest amount of feedback information, the transmitter may have a predefined set of antenna weight configurations (beams) from which one can be selected, based on the channel estimation performed by the receiver.

The SNR for the beamformed link is approximately a factor n_T higher than for the corresponding SISO link. A similar array gain can be achieved on the receiver side without the need of channel prediction, if the receiver has multiple antennas.

The choice of MIMO technique offers a tradeoff between on the one hand, increasing the data rate by multiplexing many streams in parallel, and on the other hand, multiplexing fewer (or only one) streams, which are transmitted over more “stable” channels, with better fading statistics. For a given multiple antenna transmission technique, link adaptation can then be used to further increase performance. Channel prediction will then need to be employed for making appropriate link adaptation decisions.

In modern systems and systems proposals such as LTE, WINNER, and WiMAX, link adaptation is now being used increasingly and over shorter and shorter time scales. Furthermore, as MIMO techniques are gaining a wider interest, the need for higher quality channel estimates and predictions is becoming evident. Also, recent research investigates the prospect of using *Coordinated Multi-Point transmission* (CoMP), i.e. letting multiple base stations cooperate when communicating with UEs. This potentially increases the system delays so that longer prediction ranges have to be used. These issues make channel prediction increasingly important.

As previously noted, cellular systems are interference limited, and all decisions regarding beamforming and spatial multiplexing come with an interference penalty for other UEs. A scheduling strategy that provides close to optimal performance is therefore very complicated to design. The main focus in this thesis will not be on answering how to take decisions regarding link adaptation and transmission techniques, but rather on how to produce the predictions that necessarily underlie such decisions.

2.3 Channel models

Efficient channel estimation and prediction in wideband systems necessitate the modelling of broadband frequency selective radio channels. The channel impulse response in the passband is time-varying for a mobile receiver and can be written as a sum of components that have different propagation delays $\{\tau_n\}$, due to reflections or scattering in the transmission environment [32]:

$$c(\tau, t) = \sum_{n=0}^N \alpha_n e^{-j\phi_n(t)} \delta(\tau - \tau_n), \quad (2.3.1)$$

where the factor $\alpha_n e^{-j\phi_n(t)}$ is the *complex fading amplitude* of path n . Each path is characterized by a real-valued path gain α_n , a phase angle $\phi_n(t)$, and a propagation delay τ_n . We have here assumed that the model is valid over a short period of time, and that the number of paths, N , the path gains, and the propagation delays are independent of time over this period. The phase angle for a path will however generally be time varying and can be expressed as

$$\phi_n(t) = 2\pi f_c \tau_n(t) - \phi_{D_n}(t), \quad (2.3.2)$$

where f_c is the carrier frequency and the contribution from the Doppler shift for path n , with Doppler frequency $f_{D_n}(t)$, is

$$\phi_{D_n}(t) = \int 2\pi f_{D_n}(t) dt. \quad (2.3.3)$$

Each path is associated with a distinct reflector, scatterer, or cluster of scatterers and hence a distinct incident angle θ_n . We therefore have

$$f_{D_n}(t) = v \cos \theta_n(t) / \lambda = f_D \cos \theta_n(t), \quad (2.3.4)$$

where v is the velocity, λ is the wavelength, and f_D is the maximum Doppler frequency (obtained for a path aligned with the direction of movement). If we would assume that the incident angle is constant so that $\theta_n(t) = \theta_n$, then

$$\phi_{D_n}(t) = 2\pi f_D t \cos \theta_n, \quad (2.3.5)$$

but this relation will not hold in general over long intervals because nearby point scatterers will cause θ_n to vary over time which means that (2.3.5) will contain higher order terms in t [4]. Also, clusters of scatterers may cause the Doppler frequency to be time varying.

The parameters N , $\{\alpha_n\}$, and $\{\tau_n\}$ in the model (2.3.1) vary over a considerably longer time scale than the phase angles $\{\phi_n\}$. As the basis of the design of a channel estimator/predictor, we shall therefore assume that the

passband channel can be modelled locally in time as a finite number of paths of known propagation delays $\{\tau_n\}$ but unknown complex fading amplitudes $\{\alpha_n e^{-j\phi_n(t)}\}$. The set of fading frequencies (the *Doppler spectrum*) may feature strong spectral components.

Multiple paths may have approximately the same delays. The sum of all paths having delay $\approx \tau_i$ is denoted a *cluster* and will then constitute a single term in (2.3.1). If the number of paths, N_i , in a cluster goes to infinity and the phase angles of these paths are independent, then the frequency distribution over time for the channel coefficient $c_i \triangleq c(\tau_i, t)$ will be a circular symmetric complex Gaussian distribution in the complex plane. Its squared magnitude $|c_i|^2$ will then be frequency distributed according to an exponential distribution, while the magnitude $|c_i|$ will be distributed according to a Rayleigh distribution [32], see also Appendix 4.A. However, the Gaussian distribution will also be used in the form of a *probability density* function to express our ignorance as to the value of the channel coefficient c_i . Accordingly, the probability density of the magnitude will be Rayleigh distributed. This interpretation does not mean that the *frequency* distribution has to be a Rayleigh distribution, or that the number of paths in the cluster needs to be large.

It is often convenient to use baseband representations of both the transmitted signal and the channel, so that mathematical manipulations become independent of the carrier frequency f_c . The channel model (2.3.1) can however not be used as a baseband channel model straight off since its frequency response in a region centred at f_c is not necessarily the same as that centred at 0 Hz. A proper translation in the frequency domain, that is a multiplication with $e^{-2\pi j f_c t}$, would therefore be necessary. However, we will here be studying OFDM systems exclusively, in which time domain channels and signals are transformed to the frequency domain via a discrete Fourier transform. There is therefore no direct need for a baseband representation. A factor $e^{-2\pi j f_c t}$ translating the baseband to the passband could be included in the Fourier matrix \mathcal{F} , but this is unnecessary since we shall primarily be interested in studying covariance matrices, in which every occurrence of a multiplication with \mathcal{F} will always be complemented with a multiplication with \mathcal{F}^* , effectively cancelling out the factor $e^{-2\pi j f_c t}$.

2.3.1 Statistical characterization

The channel can be statistically characterized by studying the autocorrelation function [32]

$$A_c(\tau_1, \tau_2; t, t + \Delta t) \triangleq E\{c^*(\tau_1; t)c(\tau_2; t + \Delta t)\}. \quad (2.3.6)$$

Table 2.1: Various channel characteristics in terms of carrier frequency f_c , UE velocity v , speed of light c_0 , and sampling period t_p . The expressions are valid in time-dispersive environments, where the maximum Doppler frequency dominates the Doppler spectrum. In line-of-sight scenarios, or when considering individual beams, the velocity v should be replaced with $v \cos \theta$, where θ is the incident angle.

Wavelength	$\lambda = c_0/f_c$
Maximum Doppler frequency	$f_D = v f_c/c_0$
Number of samples per period (based on f_D)	$c_0/(v f_c t_p)$
Time period t measured in wavelengths	$t f_c v/c_0$

We shall assume that $A_c(\cdot)$ is time-invariant over short periods of time and that the scattering is uncorrelated between paths so that $A_c(\tau_1, \tau_2; t_1, t_2) = 0$ when $\tau_1 \neq \tau_2$. $A_c(\cdot)$ may then be defined as a function of τ and t :

$$A_c(\tau, \Delta t) \triangleq A_c(\tau, \tau; 0, \Delta t). \quad (2.3.7)$$

With the channel model (2.3.1), $A_c(\tau, \Delta t)$ is non-zero only for discrete values on the τ -axis but features generally continuous-time functions depending on the time dynamics of the paths, e.g. Bessel functions, along the Δt -axis. See Figure 2.2 for an illustration. The range over which the *multipath intensity profile* or *delay power spectrum* or *power delay profile* $A_c(\tau, 0)$ is essentially non-zero is called the *multipath spread* T_m . Its reciprocal $(\Delta f)_c \triangleq T_m^{-1}$, which is the range over which the *spaced frequency correlation function* $\int_{-\infty}^{\infty} A_c(\tau, 0) e^{-2\pi j f \tau} d\tau$ has essential support, is called the *coherence bandwidth*. Similar measures exist for the Δt direction. Integrating over τ , we define the *spaced time correlation function*

$$A_c(\Delta t) \triangleq \int_0^{\infty} A_c(\tau, \Delta t) d\tau. \quad (2.3.8)$$

The range over which $A_c(\Delta t)$ has essential support is called the *coherence time* $(\Delta t)_c$. Its reciprocal $B_d \triangleq (\Delta t)_c^{-1}$, which is the range over which the *Doppler power spectrum* $\int_{-\infty}^{\infty} A_c(\Delta t) e^{-2\pi j f \Delta t} d\Delta t$ has support, is called the *Doppler spread*. The Doppler spread is dictated by the maximum Doppler frequency of the system, which hence is an important parameter. It can be expressed in terms of UE velocity and carrier frequency. Table 2.1 lists the maximum Doppler frequency as well as some other useful characteristics for fading channels.

Studying OFDM systems, it is convenient to characterize a channel in terms of the multipath spread and the Doppler spread. The value of the former indicates the delay difference of the significant transmission paths that contribute to the channel. Each μs of relative delay corresponds to a

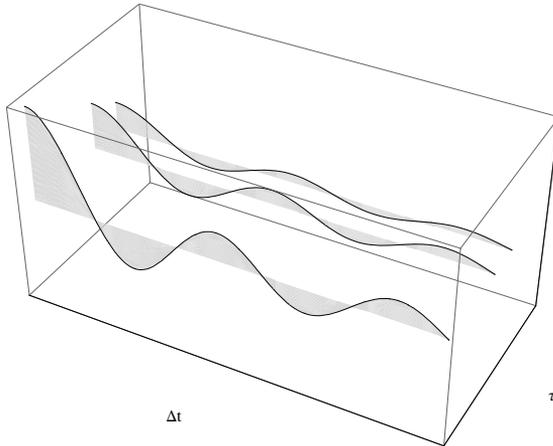


Figure 2.2: The autocorrelation function $A_c(\tau, \Delta t)$ for a channel described by the specular model (2.3.1) with three multipath components (clusters). Each multipath component here experiences fading according to Jakes' model [37],[32]. Their spaced time autocorrelation functions are therefore Bessel functions.

path difference of 300 m, so typical values for T_m is up to $10 \mu\text{s}$ in urban and suburban areas, and up to $30 \mu\text{s}$ in rural hilly environments.

A worst case value for the Doppler spread is the maximum Doppler frequency which scales linearly with the carrier frequency and the receiver's velocity. As a simple rule of thumb, with a 3 GHz carrier, the worst case value of B_d (measured in Hz) is ten times the velocity (measured in m/s).

An important characteristic used in the design of an OFDM system is the size of the time-frequency region over which the channel is essentially static. It is approximately given by $(\Delta t)_c(\Delta f)_c = (T_m B_d)^{-1}$, and is typically a few thousand Hz·s.

2.3.2 Two channel models

Here we present two channel models that will be used throughout the thesis.

The WINNER II C2 NLOS channel

The WINNER II C2 non-line-of-sight channel[38], used within the WINNER II project [39], is designed to represent a suburban macro-cell environment where the BS antenna is clearly above surrounding buildings. It represents a situation with non-line-of-sight (NLOS) propagation paths only. Each individual path is assumed to be Rayleigh fading, and the fading of all paths

Table 2.2: Power delay profile of the WINNER II C2 NLOS channel model. To each of the 24 clusters correspond a propagation delay and a total power for the paths that constitute the cluster.

cluster no.	0	1	2	3	4	5	6	7	8	9	10	11
delay[ns]	0	60	75	145	150	155	150	190	220	225	230	335
power[dB]	-6.4	-3.4	-2.0	-3.0	-5.2	-7.0	-1.9	-3.4	-3.4	-5.6	-7.4	-4.6
cluster no.	12	13	14	15	16	17	18	19	20	21	22	23
delay[ns]	370	430	510	685	725	735	800	960	1020	1100	1210	1845
power[dB]	-7.8	-7.8	-9.3	-12.0	-8.5	-13.2	-11.2	-20.8	-14.5	-11.7	-17.2	-16.7

are uncorrelated. Its *power delay profile* (PDP) is listed in Table 2.2, and the PDP is plotted along with the spaced frequency correlation spectrum in Figure 2.3.

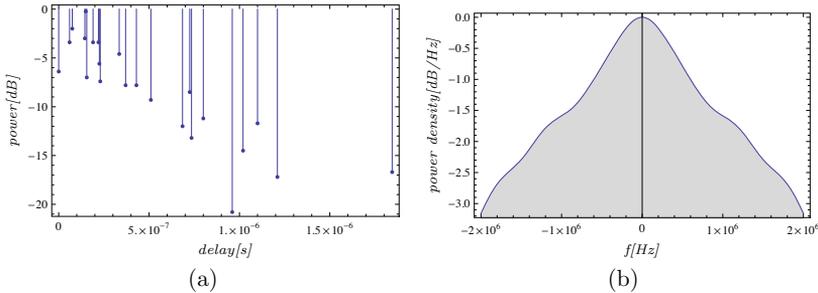


Figure 2.3: Power delay profile and spaced frequency correlation function for the WINNER II C2 NLOS channel model [38].

The -3 dB coherence bandwidth is about 4 MHz for this model, as can be seen from Figure 2.3. From Table 2.2 we see that 90 % of the total power is collected within about 450 ns, which by the relation $(\Delta f)_c = T_m^{-1}$ would suggest a coherence bandwidth of a little less than 2.5 MHz. The exact value of the coherence bandwidth is a matter of definition.

The WINNER I B1 NLOS channel

This channel model was defined within the WINNER I project[40]. It is defined for outdoor environments where both the BS antennas and the UE antennas are below surrounding buildings. We use this channel model to be consistent with reference [41], whose results we will build upon in Chapter 7. The power delay profile for this channel model is described in Table 2.3. Although the original model attributes non-zero-mean values to some of the paths, i.e. some paths have *Rice* components, we will not use any Rice components in our investigations. The PDP is plotted along with the spaced frequency correlation spectrum in Figure 2.4. The -3 dB coherence

Table 2.3: Power delay profile of the WINNER I B1 NLOS channel model. To each of the 7 clusters correspond a propagation delay and a total power for the paths that constitute the cluster.

cluster no.	0	1	2	3	4	5	6
delay[ns]	0	10	40	60	85	110	135
power[dB]	-1.25	0	-0.38	-0.10	-0.73	-0.63	-1.78

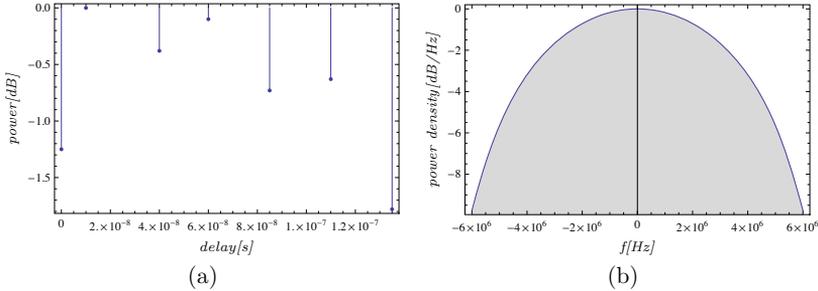


Figure 2.4: Power delay profile and spaced frequency correlation spectrum for the WINNER I B1 NLOS channel model.

bandwidth is about 8 MHz for this channel model, as can be seen from Figure 2.4. This is in agreement with the multipath spread; 90 % of the total power is collected within about 120 ns.

2.4 OFDM

Orthogonal frequency division multiplexing (OFDM) is a technique in which a large number of narrowband *subchannels* are transmitted in parallel over a large bandwidth. The technique allows for maximally dense packing of these subchannels. OFDM has been chosen as radio access technique for the LTE downlink, and for both uplink and downlink in the WiMAX standard [42]. OFDM was also used in the WINNER radio concept (uplink and downlink). OFDM is therefore the transmission technique to be studied in this thesis. A main advantage of OFDM is that equalization of frequency-selective channels becomes almost trivial, since the channel from one transmit antenna to one receiver antenna within each narrowband subchannel can be expressed as a scalar complex gain, see Section 2.4.2. Furthermore, in multiuser systems that use link adaptation, the radio resource may be divided into small time-frequency resource blocks that can be allocated to different users, using different link adaptation parameters. This allows the variations of fading channels to be exploited, making OFDM an attractive technique in such scenarios. On the downside, OFDM is sensitive to frequency offsets and requires

a comparably complex receiver architecture.

In *Single Input Single Output* (SISO) OFDM, the data stream to be transmitted is mapped onto symbols from a finite symbol alphabet (in general in combination with e.g. encoding and bit-interleaving to make the signal robust to errors). The symbols are then divided into groups of length N . These N -vectors are regarded to lie in the frequency domain, so that each element in a vector \mathbf{s}_f corresponds to one subchannel. Each vector \mathbf{s}_f is then processed by an IFFT to a vector $\mathcal{F}^*\mathbf{s}_f$ of length N , where \mathcal{F} is the Fourier matrix (see below). A ‘‘cyclic prefix’’ of length N_{CP} is added. A digital-to-analog converter transforms the sequence of length $(N + N_{CP})$ into an analog signal, which modulates the carrier. On the receiver side, an analog-to-digital converter produces the corresponding sequence of length $(N + N_{CP})$, now noisy and distorted by the channel, after which the cyclic prefix is removed. Studying one of the received sequences after cyclic prefix removal, omitting time indices, we can write the received signal \mathbf{y}_{time} of length N as

$$\mathbf{y}_{\text{time}} = \text{circ}([\mathbf{g}^T \quad \mathbf{0}_{1 \times (N-l)}])\mathcal{F}^*\mathbf{s}_f + \mathbf{w}, \quad (2.4.1)$$

where \mathbf{g} is the baseband channel impulse response of length l , \mathbf{s}_f is the frequency-domain symbol vector of length N , and \mathbf{w} is a white Gaussian noise vector of length N with covariance matrix $\sigma_w^2 \mathbf{I}_N$. The subscript $(\cdot)_f$ indicates that we are here considering the full bandwidth of N subchannels. The operator $\text{circ}(\cdot)$ is defined as

$$\text{circ}([c_0 \ c_1 \ c_2 \ \dots \ c_{n-1}]) = \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \dots & c_{n-2} \\ & & \vdots & & \\ c_1 & c_1 & c_3 & \dots & c_0 \end{pmatrix}, \quad (2.4.2)$$

so that $\text{circ}(\mathbf{a}^T)$ is a square circulant matrix whose first row is \mathbf{a}^T , and the N -by- N Fourier matrix \mathcal{F} is defined as

$$\mathcal{F}[i, j] = N^{-1/2} e^{-2\pi j i j / N}, \quad i = 0, \dots, N-1, \quad j = 0, \dots, N-1, \quad (2.4.3)$$

which implies that $\mathcal{F}\mathcal{F}^* = \mathbf{I}$.

Applying an FFT to the received signal \mathbf{y}_{time} , we use the eigenvalue decomposition of a circulant matrix,

$$\text{circ}(\mathbf{a}) = \mathcal{F}^* \cdot \text{diag}(N^{1/2} \mathcal{F} \mathbf{a}) \cdot \mathcal{F} \quad (2.4.4)$$

to get the frequency-domain signal

$$\begin{aligned}
 \mathbf{y}_f &= \mathcal{F}\mathbf{y}_{\text{time}} \\
 &= \mathcal{F}(\text{circ}([\mathbf{g}^T \quad \mathbf{0}_{1 \times (N-l)}]))\mathcal{F}^*\mathbf{s}_f + \mathbf{w}) \\
 &= \mathcal{F}\mathcal{F}^* \cdot \text{diag}(\mathbf{h}_f) \cdot \mathcal{F}\mathcal{F}^*\mathbf{s}_f + \mathcal{F}\mathbf{w} \\
 &= \text{diag}(\mathbf{h}_f)\mathbf{s}_f + \mathbf{v}_f \\
 &= \text{diag}(\mathbf{s}_f)\mathbf{h}_f + \mathbf{v}_f,
 \end{aligned} \tag{2.4.5}$$

where $\mathbf{h}_f = N^{1/2}\mathcal{F}[\mathbf{g}^T \quad \mathbf{0}_{1 \times (N-l)}]^T$ is the channel frequency response for the N subchannels, and $\mathbf{v}_f = \mathcal{F}\mathbf{w}$ has covariance matrix $\sigma_v^2\mathbf{I}_N$, where $\sigma_v^2 = \sigma_w^2$.

Note that (2.4.5) holds for a single link, i.e. one transmit antenna sending to one receiving antenna. In Chapter 4 we will extend the discussion and consider MIMO OFDM systems. This is useful for modelling channels between BSs and UEs equipped with antenna arrays. In OFDM uplinks, multi-input modelling is furthermore useful for modelling multiple UEs sharing the same radio resource. Also, in the downlink, multi-input modelling can be used to model *Coordinated Multipoint Transmission* (CoMP), where several BSs cooperate in transmitting to a UE.

2.4.1 Dimensioning and design of an OFDM system

The realistic worst case scenario of the channel's characteristics dictates how an OFDM system is normally dimensioned. First, the cyclic prefix needs to be longer than the largest expected multipath spread of the channel in order for the system to capture the energy from all propagation paths within one OFDM symbol. Hence we would set the length of the cyclic prefix to the maximum realistically expected multipath spread in the types of environments for which the transmission system is designed. The energy contained in the cyclic prefix will however be discarded at reception (unless some algorithm to collect part of this energy is used), so the length of the OFDM symbol needs to be considerably longer, say about ten times longer, than the length of the cyclic prefix, in order to limit the relative power loss (or overhead) due to the cyclic prefix. Finally, the bandwidth of a single subchannel is given by the reciprocal of the OFDM symbol duration (without cyclic prefix). The fact that the OFDM symbol duration is selected to be much longer than the multipath spread means that the coherence bandwidth will greatly exceed the bandwidth occupied by a subchannel. This means that a symbol stream transmitted over a subchannel experiences flat fading. Table 2.4 lists the OFDM symbol duration and subchannel bandwidth for a few example OFDM systems.

The symbol carried by a single subchannel of an OFDM symbol is sometimes referred to as a *channel symbol* or a *time-frequency symbol*. In this

thesis we will use the notion *sub-symbol*. The location for a sub-symbol, identified by its subchannel and OFDM symbol indices, will be referred to as a *sub-location*.

Note that, since the bandwidth of a subchannel is the reciprocal of the OFDM symbol duration, the “size” of a sub-symbol, i.e. its duration times its bandwidth, is precisely equal to 1 (without physical unit). In fact, a sub-symbol constitutes exactly one complex degree of freedom of the channel.

2.4.2 Equalization

The transmitted symbols need to be retrieved from noisy received data. Since each subchannel is subjected to flat fading, the only part of the channel affecting the received sub-symbol on sub-location $\{f, t\}$ is the complex-valued channel coefficient $h_{f,t}$ for that location:

$$y_{f,t} = h_{f,t}s_{f,t} + v_{f,t}, \quad (2.4.6)$$

where $s_{f,t}$ is the unknown transmitted sub-symbol that we want to retrieve, and $v_{f,t}$ is noise. The most common and direct approach to equalizing the channel is simply to invert the received signal with some estimate $\hat{h}_{f,t}$ of the channel:

$$\tilde{s} = \hat{h}^{-1}y = \hat{h}^{-1}hs + \hat{h}^{-1}v = \hat{h}^{-1}(\hat{h} + \tilde{h})s + \hat{h}^{-1}v = s + \hat{h}^{-1}\tilde{h}s + \hat{h}^{-1}v, \quad (2.4.7)$$

where, for brevity, we have excluded the $\{f, t\}$ subindexing, and \tilde{h} is the estimation error $h - \hat{h}$. Here, \tilde{s} is a “soft” estimate of s and forms the input to the detector. Clearly, the total noise contribution has two sources: one from the noise v and one from the estimation error \tilde{h} . The estimate (2.4.7) is called the *least squares* (LS) estimate of s .

The LS equalization is suboptimal in the sense that it does not take into consideration the variances of v and s . An alternative to LS equalization is to form the *minimum mean squares error* (MMSE) estimate (*cf.* Section 3.1):

$$\tilde{s}_{MMSE} = \frac{\hat{h}^*}{|\hat{h}|^2 + \sigma_v^2/\sigma_s^2}y. \quad (2.4.8)$$

Here, σ_s^2 and σ_v^2 are the variances of s and v , respectively. Inserting (2.4.6) into (2.4.8) and using the simplified notation, we see, after some algebra, that

$$\tilde{s}_{MMSE} = s + \frac{\tilde{h}\hat{h}^* - \sigma^2}{|\hat{h}|^2 + \sigma^2}s + \frac{\hat{h}^*}{|\hat{h}|^2 + \sigma^2}v, \quad (2.4.9)$$

where $\sigma^2 = \sigma_v^2/\sigma_s^2$. The noise terms are now slightly different from those in (2.4.7).

We will assume LS equalization in this thesis because of its simplicity compared to MMSE equalization.

Table 2.4: OFDM symbol duration and subchannel bandwidth for the WINNER II radio concept FDD and TDD modes (*WIN II FDD* and *WIN II TDD*, respectively), and the downlink for the LTE standard, using a short preamble and a long preamble (*LTE short preamble*, and *LTE long preamble*, respectively).

	WIN II FDD	WIN II TDD	LTE short preamble.	LTE long preamble.
OFDM symbol duration [μ s]	25.6	20.48	\sim 66.7	\sim 66.7
subchannel spacing [kHz]	\sim 39	\sim 49	15	15
cyclic prefix duration [μ s]	3.2	2.0 ^a	\sim 4.7 ^b	\sim 16.7

^a A switching time interval of 8.4 μ s is added at the end of each time slot of 15 OFDM symbols.

^b In a time slot of 7 OFDM symbols, the first CP is slightly longer (\sim 5.2 μ s)

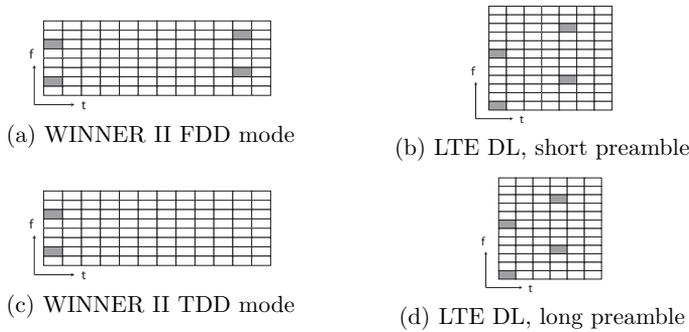


Figure 2.5: Pilot patterns for a few systems. The grids illustrate the sub-locations for pilots(gray) and payload data(white). The LTE patterns refer to the pilot layout used *per transmitting antenna*, so that when multiple transmitting antennas are used, they use similar but non-overlapping patterns. Frequency runs vertically and time runs horizontally. One square symbolizes one OFDM symbol times one subchannel. The OFDM symbol duration and subchannel bandwidth for the respective systems are listed in Table 2.4.

2.4.3 Training

To facilitate channel estimation and prediction, known reference sub-symbols are distributed across the time-frequency grid. A minimum requirement on the placement of these sub-symbols, that we will refer to as *pilots*, is that the spacing in frequency is narrower than the coherence bandwidth, and that the spacing in time is shorter than the coherence time. Figure 2.5 shows the pilot patterns used in a few example systems.

Linear filtering and inference theory

This chapter presents fundamental results in linear estimation theory needed for conducting channel estimation and prediction in later chapters. We will stress the usefulness of interpreting the outputs from channel estimators and predictors as a complete state of the knowledge that is acquired from noisy channel measurements. The foundations for interpreting probabilities as a state of knowledge is here presented, in Section 3.2. Readers that are already acquainted with this interpretation of probability may just skim over this section. In Section 3.4, we present fundamental results for Kalman filter theory that will be useful later in the thesis. Again, readers familiar with the subject may skip this Section at this moment and return to specific results as they are needed.

3.1 A channel model

In this thesis we shall be concerned with estimation and prediction of unknown radio channels. The two are closely related, so to simplify the exposition we will initially restrict the discussion to channel estimation. To begin with, we must describe the relationship between some kind of measurement \mathbf{y} and the unknown channel(s) \mathbf{h} . As a first illustration, we will take the model

$$\mathbf{y} = \Phi\mathbf{h} + \mathbf{v} \tag{3.1.1}$$

to be an appropriate description of reality. The term \mathbf{v} stands for unknown *additive white Gaussian noise* (AWGN), however with known covariance matrix \mathbf{R}_v . We assume that the \mathbf{v} and \mathbf{h} are uncorrelated. The regressor matrix Φ is also assumed known, as well as the covariance matrix \mathbf{R}_h for the prior distribution of \mathbf{h} .

Channel estimation and prediction theory has a somewhat backwards

approach to modelling; it is the channel that is the desired “signal”. The transmitted signals, which are included in Φ , merely serve as an aid for estimation. With this approach, known “reference” signals have to be transmitted at the time-frequency sub-locations where the channel estimation takes place. Once an estimate has been obtained, it can be inter- or extrapolated to sub-locations where unknown payload data reside.

The objective is now to calculate an estimate $\hat{\mathbf{h}}$ of \mathbf{h} of highest possible accuracy, based on a measurement \mathbf{y} . The probably oldest approach, going back to Legendre and Gauss, is to form the estimate that produces the least possible power for the noise \mathbf{v} , i.e. to choose $\hat{\mathbf{h}}$ such that

$$\|\mathbf{y} - \Phi\hat{\mathbf{h}}\|^2 \leq \|\mathbf{y} - \Phi\check{\mathbf{h}}\|^2 \quad (3.1.2)$$

for all $\check{\mathbf{h}} \in \mathbb{C}^n$, where n is the dimensionality for \mathbf{h} . Given in most textbooks on linear estimation, the solution follows from differentiating $\|\mathbf{y} - \Phi\check{\mathbf{h}}\|^2$ with respect to $\check{\mathbf{h}}$ and finding the root. It is usually referred to as the *least squares* (LS) solution. Assuming Φ to be a matrix with full rank, it is given by

$$\hat{\mathbf{h}}_{LS} = (\Phi^*\Phi)^{-1}\Phi^*\mathbf{y}, \quad (3.1.3)$$

where $(\cdot)^*$ denotes Hermitian transposition.

The LS estimate can be improved upon by taking into account the prior knowledge that we have about \mathbf{h} and \mathbf{v} . Based on our above stated prior information of the properties of \mathbf{h} and \mathbf{v} (the matrices \mathbf{R}_h and \mathbf{R}_v), we below attempt to find a *linear* estimator $\hat{\mathbf{h}} = \mathbf{K}_0\mathbf{y}$, where \mathbf{K}_0 is in general a matrix, that minimizes the error variance. Let $\mathbf{R}_h \triangleq E\{\mathbf{h}\mathbf{h}^*\}$, $\mathbf{R}_y \triangleq E\{\mathbf{y}\mathbf{y}^*\}$, $\mathbf{R}_v \triangleq E\{\mathbf{v}\mathbf{v}^*\}$, and $\mathbf{R}_{hy} = \mathbf{R}_{yh}^* = E\{\mathbf{h}\mathbf{y}^*\}$. We also assume $E\{\mathbf{h}\} = \mathbf{0}$ and $E\{\mathbf{v}\} = \mathbf{0}$. Then we obtain

$$\begin{aligned} E\{(\mathbf{h} - \mathbf{K}_0\mathbf{y})(\mathbf{h} - \mathbf{K}_0\mathbf{y})^*\} &= \\ \mathbf{R}_h - \mathbf{R}_{hy}\mathbf{K}_0^* - \mathbf{K}_0\mathbf{R}_{yh} + \mathbf{K}_0\mathbf{R}_y\mathbf{K}_0^* &= [\mathbf{I} \quad -\mathbf{K}_0] \begin{bmatrix} \mathbf{R}_h & \mathbf{R}_{hy} \\ \mathbf{R}_{yh} & \mathbf{R}_y \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{K}_0^* \end{bmatrix} = \\ [\mathbf{I} \quad -\mathbf{K}_0] \begin{bmatrix} \mathbf{I} & \mathbf{R}_{hy}\mathbf{R}_y^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{R}_h - \mathbf{R}_{hy}\mathbf{R}_y^{-1}\mathbf{R}_{yh} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_y \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_y^{-1}\mathbf{R}_{yh} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{K}_0^* \end{bmatrix} = \\ [\mathbf{I} \quad \mathbf{R}_{hy}\mathbf{R}_y^{-1} - \mathbf{K}_0] \begin{bmatrix} \mathbf{R}_h - \mathbf{R}_{hy}\mathbf{R}_y^{-1}\mathbf{R}_{yh} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_y \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{R}_y^{-1}\mathbf{R}_{yh} - \mathbf{K}_0^* \end{bmatrix} = \\ \mathbf{R}_h - \mathbf{R}_{hy}\mathbf{R}_y^{-1}\mathbf{R}_{yh} + (\mathbf{R}_{hy}\mathbf{R}_y^{-1} - \mathbf{K}_0)\mathbf{R}_y(\mathbf{R}_y^{-1}\mathbf{R}_{yh} - \mathbf{K}_0^*). \end{aligned} \quad (3.1.4)$$

In the above, we have used block triangular matrix factorization (see Appendix 3.A.1). Assuming that $\mathbf{R}_y > 0$, it follows that (3.1.4) is minimized when $\mathbf{K}_0 = \mathbf{R}_{hy}\mathbf{R}_y^{-1}$. The corresponding channel estimate, denoted the *minimum mean square error* (MMSE) solution is $\hat{\mathbf{h}} = \mathbf{R}_{hy}\mathbf{R}_y^{-1}\mathbf{y}$. From the model

(3.1.1) we have that $\mathbf{R}_{hy} = \mathbf{R}_h \Phi^*$ and $\mathbf{R}_y = \Phi \mathbf{R}_h \Phi^* + \mathbf{R}_v$, assuming that \mathbf{h} and \mathbf{v} are uncorrelated. Inserting these, we find the MMSE solution to be

$$\hat{\mathbf{h}} = \mathbf{R}_h \Phi^* (\Phi \mathbf{R}_h \Phi^* + \mathbf{R}_v)^{-1} \mathbf{y} \quad (3.1.5)$$

with mean square error (MSE)

$$\begin{aligned} E\{(\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^*\} &= \mathbf{R}_h - \mathbf{R}_{hy} \mathbf{R}_y^{-1} \mathbf{R}_{yh} \\ &= \mathbf{R}_h - \mathbf{R}_h \Phi^* (\Phi \mathbf{R}_h \Phi^* + \mathbf{R}_v)^{-1} \Phi \mathbf{R}_h. \end{aligned} \quad (3.1.6)$$

The above derivation (3.1.4) does not reveal the full power of the results (3.1.5) and (3.1.6) and how they go beyond the estimate (3.1.3). What the MMSE estimator gives us is actually a full description of our *state of knowledge* about the quantity \mathbf{h} . Common statistical theory, which talks about unknown quantities as *random variables* and sees probabilities as imagined frequencies, does not allow us to clearly see this. In the following, we will refer to this classical interpretation of statistics as *orthodox statistics* or *frequentism*. An alternative formulation of statistical theory, which regards probability theory to be a direct extension to logic, however offers an interpretation of the probability concept that bears a direct correspondence to available information. This theory is commonly referred to as *Bayesianism*. It shows us how to calculate a representation of our state of knowledge about a parameter of interest (\mathbf{h}), given some information (\mathbf{y}) and background information (the model (3.1.1) plus covariance matrices). In the next section, we give a short account on *Bayesian probability theory*.

3.2 Bayesianism

In this section we will give a brief account on Bayesian probability theory. The two main tools in Bayesianism – Bayes’ theorem and marginalisation – are presented, as well as the general method for conducting inference according to the Bayesian school. We also take the opportunity to point out a few differences in method between orthodoxy and Bayesianism. The ambition is however not to give a comprehensive review of Bayesianism. We have previously given the subject more attention in [43], and we encourage interested readers to read the book by Jaynes [44].

3.2.1 Probability

Probability theory operates on *propositions*, which we denote by capital letters. A proposition is described by a statement, for example $A = \text{‘the Earth is orbiting the Sun’}$, and by declaring A we declare that that statement is true. Propositions conform to the axioms of logic:

not not A	is equivalent to	A
A or B	is equivalent to	B or A
A or A	is equivalent to	A
A or (B or C)	is equivalent to	(A or B) or C
not (A and B)	is equivalent to	not A or not B
A or (A and B)	is equivalent to	A

Using these axioms, a complicated chain of statements can be simplified to a shorter and simpler form. The axioms of logic are however only helpful in conducting deductive reasoning in which no element of uncertainty exists. In real-world situations, statements of any sort can rarely be declared as being absolutely true or absolutely false. Rather, we may know that a proposition A is *likely* to be false, or that B is *probably* true. To be useful in practice, it would therefore be highly desirable to be able to extend the axioms of logic to hold also for such vague information.

The first steps to pursue this goal were taken more than two centuries ago by Bayes and Laplace. In the 1930s, the theory was refined by Harold Jeffreys, and then, in 1946, Richard Cox published a paper which proved that probability theory can be seen as an extension to logic. Cox [45] introduced the concept of *plausibility*¹. Plausibility is a general measure of degree of belief. The plausibility of a specific proposition will vary depending on what other propositions we know to be true. For example, it is more plausible that it is freezing outside if we know that it is winter, than if we are ignorant to the time of year. In accordance with a notation introduced by John Maynard Keynes in 1921, we will denote plausibilities by

$$A|B, \tag{3.2.1}$$

meaning the plausibility of proposition A given that proposition B is true.

Bayesian theory defines a probability as a state of knowledge, whereas in orthodox probability theory, probabilities represent observed or imagined limiting frequencies. Cox aimed to show that rules for probability theory interpreted in the Bayesian sense – rules that by the time of Cox already had been employed as axioms by several generations of workers in the field – could be derived from the axioms of classical logic by only adding a few ‘common sense’ requirements.

Cox stated these requirements as functional relationships between plausibilities. Later, Edwin Jaynes [44] chose to reformulate these as verbal statements, making the exposition easier to follow. We will adopt Jaynes’ view here, although we leave out the actual derivation.

¹To be precise, Cox really used the term *likelihood*, but in modern statistical theory the meaning of that notion has shifted.

Jaynes uses three *desiderata*, starting with

- (I) *Degrees of plausibility are represented by real numbers.*

We will adopt the convention that higher numbers correspond to higher plausibilities, without further specifying the exact relationship. Jaynes discusses other possibilities for constructing a theory for plausible reasoning, where this desideratum is not needed. Instead he replaces desideratum (I) with two more elementary ones where only comparisons between degrees of plausibilities are needed,

- (Ia) If $(A|X) \geq (B|X)$ and $(B|X) \geq (C|X)$ then $(A|X) \geq (C|X)$, and
 (Ib) given A, B, C , one of $(A|C) > (B|C)$, $(A|C) = (B|C)$, $(A|C) < (B|C)$ must hold,

and argues that these are equivalent to desideratum (I), see [44, Appendix A.3].

The second desideratum is concerned with how plausibilities change when new data are obtained. If old information C is updated to new information C' so that the plausibility for A is increased,

$$(A|C') > (A|C), \quad (3.2.2)$$

while the plausibility for B stays the same,

$$(B|A, C') = (B|A, C), \quad (3.2.3)$$

then common sense says that

$$(A, B|C') \geq (A, B|C), \quad (3.2.4)$$

and that

$$(\overline{A}|C') < (\overline{A}|C), \quad (3.2.5)$$

where \overline{A} denotes the logical complement of A , that is the proposition that is always true when A is false and vice versa, and A, B means that A and B are both true. The above ‘common sense’ requirements are expressed by desideratum II :

- (II) *Qualitative correspondence with common sense.*

The third desideratum is divided into three statements, all having to do with the consistency of the theory :

- (IIIa) *If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.*
- (IIIb) *We must always take into account all of the evidence available that is relevant to the problem.*
- (IIIc) *Equivalent states of knowledge must always be represented in the same way.*

Somewhat surprisingly, these three desiderata are all that is needed to derive a consistent theory for plausible reasoning. Although the scale of the plausibility measure is completely arbitrary, Cox' derivation revealed that one can define *functions* operating on plausibilities, such that these functions need to conform to quantitative rules. The most convenient is a function $P(\cdot)$ that fulfills

$$P(A, B|C) = P(A|C)P(B|A, C) = P(B|C)P(A|B, C) \quad (\text{The product rule})$$

$$P(A|B) + P(\bar{A}|B) = 1 \quad (\text{The sum rule})$$

The function $P(A|B)$ is termed *the probability of A given B*. It has the additional property that $P(\text{'true statement'}) = 1$ and $P(\text{'false statement'}) = 0$.

Summarizing, we have the following definitions:

Plausibility is a measure of belief isomorphic to the real numbers, so that the plausibility can be either increased, decreased or unaltered by new information.

Probability is a monotonic increasing function of plausibility and obeys the product and sum rules, necessitated by Cox's desiderata.

An important point often neglected in the literature is that the function $P(\cdot)$ is not in any way a more correct probability definition than some monotonously increasing mapping $q(\cdot) = f \circ P(\cdot)$. However, P has qualities that make it preferable to other functions apart from the fact that the sum and product rules look attractively simple. Knowing that there are, say, seven red and three white balls in an urn, then our choice of function P gives a probability 3/10 of a white ball being drawn. This certainly seems a sound property for a definition of probability. In fact, Laplace used this property as the definition of probability:

The probability for an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally probable.

We shall therefore adopt the same definition and choose the function P as the definition for probability. One should however keep in mind that specific probability values only possess *relative* meaning when connections to frequencies cannot be made. Probability values can only tell whether a proposition is more or less plausible than some other proposition.

For it is important to note that the present theory makes no references to repeated experiments, observed frequencies, or hypothetically observed frequencies. The definition of probability used here applies to all kinds of propositions. Since the rules of probability follows directly from logic and the three desiderata, probability according to Bayesianism is therefore an *extension of deductive reasoning*.

Bayesianism also differs from orthodox probability theory in that there is no such thing as an *unconditional* probability in Bayesianism. Whereas one according to the orthodox school can talk about $P(A)$, Bayesianism will always require us to specify what information is available. Even if no cogent evidence for A is present, there will always be *some* background information at hand, such as a mathematical model and/or some vague, uninformative prior probability distribution. We shall generally denote such prior information by the symbol I .

Continuous variables

When conducting inferences about a continuous parameter θ , as most scientific problem formulations would require us to do, we may define a mutually exclusive set of propositions

$$H_k = 'k\Delta\theta \leq \theta < (k+1)\Delta\theta', \quad (3.2.6)$$

where $\Delta\theta$ is the size of the intervals represented by the propositions. We can then define a function $p(\theta|I)$ so that

$$P(H_k|I) = p(\theta|I)\Delta\theta, \quad \theta = k\Delta\theta. \quad (3.2.7)$$

Letting the number of propositions go to infinity, so that $\Delta\theta$ goes to an infinitesimally small interval $d\theta$, $p(\theta|I)$ becomes the *probability density function* (pdf) of θ . From recursive application of the sum rule we see that a pdf must

have the property that

$$p(\theta|I) \geq 0, \text{ and that}$$

$$\int_{-\infty}^{\infty} p(\theta|I)d\theta = 1.$$

In the above, we have assumed that θ is real-valued. If θ is an element of some other set, we would take the integral over the entire domain of θ .

3.2.2 Tools in probability theory

From the sum and product rules we may derive more useful tools. Essentially only two tools are needed to conduct general inferential calculus. Working from a pdf with a set of parameters to the left of the conditioning bar and another set of parameters to the right, *Bayes' theorem* is used to switch places between parameters on the respective sides of the conditioning bar, while *marginalisation* is used to remove parameters from the left side of the conditioning bar. Appropriately combining these two tools allows us to produce a pdf that has parameters of interest on the left side of the conditioning bar, and available data on the right.

Bayes' theorem

By a simple rearrangement of the product rule, we get Bayes' theorem:

$$P(X|D, I) = P(X|I) \frac{P(D|X, I)}{P(D|I)} \quad (3.2.8)$$

Bayes' theorem describes how to update our state of knowledge about a certain proposition X when we get some data D that is related to X in some way. That is to say, starting with a representation of our prior knowledge $P(X|I)$, reception of information D lets us update this to $P(X|D, I)$. Usually, we will have some physical relationship between X and D which makes it easy to say what we know about D given that X is true ($P(D|X, I)$). Bayes' theorem is then used to switch places between these propositions.

Bayes' theorem applies equally well to probability density functions as it does to probabilities:

$$p(x|y, I) = p(x|I) \frac{p(y|x, I)}{p(y|I)}. \quad (3.2.9)$$

Marginalisation

For a set of mutually exclusive and exhaustive propositions H_1, H_2, \dots, H_K we can write

$$P(A, H_1|I) + P(A, H_2|I) + \dots + P(A, H_K|I) = P(A|I). \quad (3.2.10)$$

In the limit it holds that

$$p(x|I) = \int p(x, \theta|I) d\theta. \quad (3.2.11)$$

The parameter θ that was eliminated by this so called *marginalization* is called a *nuisance parameter*. Nearly every problem in probability theory requires the use of marginalisation in this way to remove parameters that are useful as intermediate parameters but are not of interest in the final result.

Change of variables

It is common that a pdf $p(\mathbf{x}|I)$ is known but that we want to know a pdf $p(\mathbf{y}|I)$ where \mathbf{x} and \mathbf{y} are related through some known functional relationship $\mathbf{y} = f(\mathbf{x})$. We then need to perform a change of variables. Of course, changing variables is not a method exclusive to bayesian probability theory, but we mention it here because of its common usage in the field. It holds that^[46]

$$p(\mathbf{y}|I) = p(\mathbf{x}|I) \left| \frac{\partial(x_1, \dots, x_m)}{\partial(y_1, \dots, y_n)} \right|_{\mathbf{x}=f^{-1}(\mathbf{y})}, \quad (3.2.12)$$

where $\left| \frac{\partial(x_1, \dots, x_m)}{\partial(y_1, \dots, y_n)} \right|$ is the Jacobian and m and n are the dimensions for \mathbf{x} and \mathbf{y} , respectively. To take an example, let

$$p(x|I) = 1, \quad 0 < x < 1, \quad (3.2.13)$$

and let $y = x^2$. Then

$$p(y|I) = p(x|I) \left| \frac{dx}{dy} \right|_{x=\sqrt{y}} = \frac{dx}{2x dx} \Big|_{x=\sqrt{y}} = (4y)^{-1/2}, \quad 0 < y < 1. \quad (3.2.14)$$

Changing variables can become very complicated if the functions involved are multidimensional and/or the mapping from \mathbf{x} to \mathbf{y} is not one-to-one.

The type of notation “ $p(x|I)$ ” is unorthodox from a mathematical point of view in that x is not only the free variabel in the function $p(x|I)$, but also a part of the function *name*. What if we want to evaluate $p(x|I)$ at the point $x = y$? How do we express it? $p(y|I)$ will not do, because that would

refer to a totally different function, namely the distribution for the variable y . $p('x = y'|I)$ could work but it would not mean the same thing as $p('y = x'|I)$, which would seem strange. To avoid confusion in such situations, one should use a temporary function

$$fnc(x) = p(x|I), \quad (3.2.15)$$

thus eliminating the risk for ambiguous meaning.

3.2.3 Priors

Looking at the product and sum rules on page 34, it is evident that a pdf can only be deduced from other pdf:s. Hence there is the need for principles that can produce the initial distributions required to “get started”. These initial pdf:s must be based on – and only on – whatever information we have beforehand. Such information is sometimes very vague and difficult to describe in quantitative terms. So how do we calculate *prior probability densities* (or simply *priors*) when seemingly there is almost no prior information available? There are a number of principles available, see e.g. [43] for a few examples. In the present thesis, the issue of prior assignment will be restricted to priors for fading radio channels. Assuming that a good estimate of the covariance matrix \mathbf{R}_h for a multidimensional channel \mathbf{h} is available, we invoke the *maximum entropy principle* [44] which says that the least informative prior distribution that we can assign given this particular piece of information is zero-mean Gaussian:

$$p(\mathbf{h}|I) = \mathcal{CN}(\mathbf{h}; \mathbf{0}, \mathbf{R}_h). \quad (3.2.16)$$

This is in fact the only prior we will need. One may object to the prior (3.2.16) in that it assigns non-zero probability to any value of \mathbf{h} , no matter how high. In practice there must be an upper bound to \mathbf{h} . However, truncating the Gaussian would yield a non-Gaussian prior distribution, and this would lead to calculatory problems, as we shall see later. The prior (3.2.16) is therefore partly motivated on practical grounds.

3.2.4 How to make a decision

Priors define a ‘starting point’ of an inference problem. Then, through the use of Bayes’ theorem and marginalisation, we may produce a ‘post-data’ (posterior) pdf for the parameter that we are interested in. But what do we do with this pdf? The answer to this question depends on what type of answer we ultimately seek. The perhaps most common type of inference problem is the point estimation, in which we want to extract one single

value from the posterior pdf. So, given the posterior pdf, which value do we extract? Value judgment inevitably enters the theory at this point; our estimation strategy will depend on what we are prepared to lose if we happen to make the wrong guess. The theory behind all this is called *decision theory* and it is a large field on its own. Looking for the estimate $\hat{\theta}$ of a parameter θ , we construct a *loss function* $L(\theta, \hat{\theta})$ that quantifies the loss associated with making the wrong guess. One possible and common choice is then to choose the estimator $\hat{\theta}$ that minimizes the expected loss

$$\int L(\theta, \hat{\theta})p(\theta|D, I)d\theta. \quad (3.2.17)$$

The *quadratic* loss function

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 \quad (3.2.18)$$

is by far the most commonly employed criterion and is reasonably the optimal criterion to use in repeated scenarios, which typically arise in digital communications problems. Minimizing the quadratic error conforms to choosing the mean value of a pdf as estimate for that parameter,

$$\hat{\theta} = \int \theta p(\theta|D, I)d\theta, \quad (3.2.19)$$

which minimizes the expected value of the error $(\theta - \hat{\theta})^2$.

Note however that using the quadratic loss function is not always appropriate. For example, in estimating the amount of rocket fuel that is required to take an astronaut to the Moon and back, we want the estimate to be as small as possible but not too small. A slight overestimation is alright, but an underestimation would be disastrous. Also, θ may not be an element in a metric space, in which case the error $(\theta - \hat{\theta})^2$ cannot even be defined. To illustrate, let θ be an element in a discrete set of models. Based on evidence D , we wish to decide which model is the correct one. The *difference* between two models however cannot be appropriately defined, and so (3.2.18) cannot be used.

3.2.5 Model selection

So far we have considered assignment of prior pdf:s, the manipulation of pdf:s by means of marginalisation and Bayes' theorem, and taking a decision from a posterior pdf. In most problems of inference however, we have a set of data from which we wish to draw conclusions. It is then necessary to have a model which relates the measured data D to the parameter(s) of interest θ :

$$D = f(\theta, \xi). \quad (3.2.20)$$

Additional nuisance parameters ξ are also in general included in the model. The posterior $p(\theta|D, I)$ can now be constructed by using the tools of probability theory and the principles for assigning priors. A general scheme for Bayesian inference can be summarized as follows:

1. Construct a model (3.2.20),
2. assign priors,
3. use probability theory to derive the posterior pdf for θ given whatever information is available, and
4. invoke decision theory to produce an answer to whatever question was asked.

But then, how do we know how to set up the model in the first place? Imagine that we have a set $\{M_k\}$ of candidate models, and that we want to evaluate them against one another based on evidence D . Using Bayes' theorem, we have

$$P(M_k|D, I) = P(M_k|I) \frac{p(D|M_k, I)}{p(D|I)}. \quad (3.2.21)$$

The set of models is discrete, so we will choose the model that gives the highest value of the posterior $P(M_k|D, I)$. This corresponds to using a loss function that attributes the same penalization to every wrong estimate. Since we only want to compare the probabilities of the different models we do not need to calculate the denominator $p(D|I)$. Also, in most cases we would assign equal values to all prior probabilities $P(M_k|I)$. Thus we have

$$P(M_k|D, I) \propto p(D|M_k, I). \quad (3.2.22)$$

The right-hand-side is usually called the *likelihood* of M_k and is sometimes denoted $L(M_k)$. Although marginalisation over a few nuisance parameters may be required, $L(M_k)$ is usually much easier to evaluate than the probability $P(M_k|D, I)$.

Somewhat dishearteningly, there exists no formal procedure for model selection. That is to say, once we have established a particular *set* of models, then probability theory provides us with the means of telling which one is best suited given a set of measurements, but we are still left in the dark when it comes to choosing the original set $\{M_k\}$.

The model selection process therefore consists of first choosing a set of models, and then evaluating the likelihood for each model given a set of data. Taken that all models are assigned the same prior probability, we choose the model that gives the highest likelihood.

3.2.6 The methods of frequentism and Bayesianism

Problems of inference² come in many different forms of which the most common ones are sampling theory, hypothesis testing, and parameter estimation.

Sampling theory is the theory of determining the probabilities for outcomes (samples) in data series, discrete or continuous. Sampling theory is seldom relevant to Bayesianism, except in the sense that sampling distributions often occur inside calculations. It is however central to orthodox statistics; in order to find the most likely values of some model parameters given some data, the orthodox statistician uses sampling theory to calculate the probability for the data *that was actually received*, as a function of the model parameters, and then searches for the values of those parameters that maximize the probability. This principle is called *the maximum likelihood principle*. When no cogent prior information is available, and when the size of the error in the final guess is irrelevant, the maximum likelihood solution coincides with the Bayesian solution.

Hypothesis testing is the procedure of deciding which model that best describes a given set of data. Model selection is an alternative name for the same thing, although model selection often refers to estimation of the model *structure*, while hypothesis testing usually concerns values of fixed parameters in the model.

Parameter estimation is, as the name strongly indicates, the estimation of one or many parameters from given data. Its output can either be specific values, in which case one talks about *point estimation*, or intervals, which is called *interval estimation*.

All problems of inference begin with a model relating the parameters of interest with the data and possibly additional nuisance parameters,

$$D = f(\theta, \xi), \quad (3.2.23)$$

where the measured data D , the parameters of interest θ , and the nuisance parameters ξ are generally vector valued.

The general method for obtaining information about θ varies depending on whether one confesses to the frequentist or the Bayesian school. In frequentist theory, which method to use depends on the type of problem, whereas in Bayesianism the same procedure for inference is applied regardless of the type of problem. Bayesianism therefore has one single approach to

²'Inference theory' is normally distinguished from 'decision theory'; In Bayesianism, the inference part of a problem is to produce the posterior of the parameter under consideration, whereas the decision part amounts to determining the course of action from the posterior. For an engineer it is of little interest to produce a pdf without any concrete suggestion to a course of action coming out of the calculations, and so here we will be careless about the terminology and talk about 'inference' when we mean the joint process of inference and decision making.

inference, while frequentism consists of a large number of methods. Below follows a short summary of methods used in the respective camps. *Bayesianism*:

1. Determine the model and the prior distributions.
2. Use the tools of probability theory to derive the posterior distribution. In sampling theory the posterior is often given directly by the model and the prior. Estimation and hypothesis testing often requires application of Bayes' theorem and marginalisation.
3. Investigate the posterior to produce the sought-after result. *Point estimation* is performed through the use of a loss function integrated over the posterior (see Section 3.2.4). *Interval estimation* amounts to finding an interval of the posterior distribution (usually the shortest interval possible) having a certain area. *Hypothesis testing* can be seen as a special case of parameter estimation. In hypothesis testing, it is mostly common to choose the hypothesis having the highest probability. It is equivalent to using a loss function that ascribes the same loss to errors of any size. This estimate is called the *Maximum A Posteriori* (MAP) estimate.

Frequentism:

- **Point estimation** of a parameter θ is commonly performed by using intuition to invent an *unbiased estimator* $\hat{\theta}(D)$ which is a function of data D . It is then adjusted so that its mean value *over the sampling distribution*, $\int \hat{\theta}(D)p(D|\theta) dD$, equals the parameter value θ . Note that this may be quite different from the Bayesian least mean squares estimate, $\int \theta p(\theta|D, I) d\theta$. Other orthodox parameter estimation methods include the maximum likelihood method.
- **Interval estimation** also starts by inventing an estimator $\hat{\theta}(D)$. The sampling distribution $p(\hat{\theta})$ is then calculated. Technically, this is done by a change of variables, $p(\hat{\theta})d\hat{\theta} = p(D|\theta)dD$. Lastly, one finds the least interval over which this distribution has an area of 0.9 or so.
- **Hypothesis testing** relies on a number of *significance tests*, for example the χ^2 -test. The maximum likelihood method is another method, which – if no cogent prior information is available and the loss function prescribes the same value to any error, regardless of its size – produces the same answer as the Bayesian approach.

Frequentist theory requires the division of model parameters into *random variables* and *deterministic but unknown parameters*. Only random variables

are allowed to have pdf:s. There exists no formal procedure for categorizing a parameter as either random or deterministic. Conventionally, measured data is taken to be random variables while all other parameters are considered deterministic. A rule of thumb is therefore to let that which is known (the data) be random variables and let all other parameters – those that are the subject of the inference in most cases – be deterministic. Conversely, the concept of random/deterministic parameters bears no meaning in Bayesianism.

3.2.7 Why Bayesianism?

After this somewhat lengthy excursion into Bayesianism, one may enquire as to what the significant difference between Bayesianism and classical statistics actually is. After all, the rules of probability are the same in both schools. Is not then the confession to one or the other merely a matter of ideology with little practical relevance?

Actually, conclusions reached by the Bayesian method can be totally different from those attained by using conventional methods. Whereas inferences drawn with aid of the Bayesian method is guaranteed to be consistent with logic, orthodox statistics may leave you with unreasonable results. A few examples were presented in [43]. The differences tend to be especially large when cogent prior information is available. In the present thesis we will study inferences of radio channels where new measurements are made available on a regular basis, so that prior information quickly becomes obsolete, but there are still a number of reasons for choosing Bayesianism over frequentism:

- Bayesianism always produces a full representation of one's state of knowledge (the pdf). This provides a quality measure of the inference in terms of how uncertain it is. The pdf is also necessary when the parameter of interest has some complicated functional relationship to parameters whose pdf has been found, so that a change of variables need to be performed. With conventional statistical tools, an estimator of the parameter of interest may have to be invented from intuition, which may be very difficult to do, and the quality of such an estimator cannot be guaranteed.
- Bayesianism offers one single method for inference. Orthodoxy offers many, and they generally yield different answers.
- Bayesianism does not equate probabilities with frequencies of any sort. A frequency distribution is a measurable quantity just as any physical property, so that a Bayesian can talk about pdf:s for frequency distributions. This is useful in a context such as digital communications, where frequency distributions abound.

- We will later be interested in examining the behaviour of channel estimates of a special kind. These estimates are based on a certain amount of data, but we will have to infer them before all of this data is available. One could say that we are interested in a parameter θ given evidence D_1 , given evidence D_2 . A framework like Bayesianism that is always specific about background information will serve us well when conducting inference in involved problems like these.
- A central element in the Kalman filter, that will be studied shortly, is a matrix that we will denote by the symbol $\mathbf{P}_{t|t}$. In Bayesianism, this matrix is straightforwardly interpreted as the covariance matrix for the posterior distribution of the so called *state vector*, allowing it to be used as an uncertainty measure or for changing variables. In orthodox statistics however, the interpretation of $\mathbf{P}_{t|t}$ depends on the nature ascribed to the state vector. If the state vector can be said to be of a random nature, then $\mathbf{P}_{t|t}$ is interpreted in the same way as in Bayesianism, but if the state vector is said to be deterministic (e.g. if it is known to be constant), then $\mathbf{P}_{t|t}$ is not allowed this interpretation. This is a pity, because it may withhold from us the opportunity to conduct inferences that are consistent with logic.

3.3 The MMSE solution

Returning now to the problem of finding an estimate $\hat{\mathbf{h}}$ of \mathbf{h} in (3.1.1), we apply Bayes' theorem and find

$$\begin{aligned}
 p(\mathbf{h}|\mathbf{y}, I) &\propto p(\mathbf{y}|\mathbf{h}, I)p(\mathbf{h}|I) \\
 &= \mathcal{CN}(\mathbf{y}; \Phi\mathbf{h}, \mathbf{R}_v) \times \mathcal{CN}(\mathbf{h}; \mathbf{0}, \mathbf{R}_h) \\
 &\propto \exp\left(-\frac{1}{2}((\mathbf{y} - \Phi\mathbf{h})^* \mathbf{R}_v^{-1}(\mathbf{y} - \Phi\mathbf{h}) + \mathbf{h}^* \mathbf{R}_h^{-1} \mathbf{h})\right),
 \end{aligned} \tag{3.3.1}$$

where $\mathcal{CN}(\cdot)$ is the multivariate circular symmetric complex Gaussian distribution defined in Appendix 3.B.1. Note that we have saved ourselves the trouble of calculating the denominator $p(\mathbf{y}|I)$ in Bayes' theorem since it is independent of \mathbf{h} and therefore just constitutes a normalization factor. If needed, the proper normalization factor can at any time be calculated from the exponent of the Gaussian distribution. By completing the squares with

respect to \mathbf{h} , we obtain

$$\begin{aligned}
p(\mathbf{h}|\mathbf{y}, I) &\propto \\
&\exp\left(-\frac{1}{2}\left(\mathbf{h}^*\Phi^*\mathbf{R}_v^{-1}\Phi\mathbf{h} + \mathbf{h}^*\mathbf{R}_h^{-1}\mathbf{h} - \mathbf{y}^*\mathbf{R}_v^{-1}\Phi\mathbf{h} - \mathbf{h}^*\Phi^*\mathbf{R}_v^{-1}\mathbf{y} + \mathbf{y}^*\mathbf{R}_v^{-1}\mathbf{y}\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\mathbf{h}^*(\Phi^*\mathbf{R}_v^{-1}\Phi + \mathbf{R}_h^{-1})\mathbf{h} - \mathbf{y}^*\mathbf{R}_v^{-1}\Phi\mathbf{h} - \mathbf{h}^*\Phi^*\mathbf{R}_v^{-1}\mathbf{y} + \mathbf{y}^*\mathbf{R}_v^{-1}\mathbf{y}\right)\right) \\
&= \exp\left(-\frac{1}{2}\left((\mathbf{h} - \hat{\mathbf{h}})^*\mathbf{R}_{\hat{\mathbf{h}}}^{-1}(\mathbf{h} - \hat{\mathbf{h}}) - \hat{\mathbf{h}}^*\mathbf{R}_{\hat{\mathbf{h}}}^{-1}\hat{\mathbf{h}} + \mathbf{y}^*\mathbf{R}_v^{-1}\mathbf{y}\right)\right) \quad (3.3.2)
\end{aligned}$$

where $\hat{\mathbf{h}} = (\Phi^*\mathbf{R}_v^{-1}\Phi + \mathbf{R}_h^{-1})^{-1}\Phi^*\mathbf{R}_v^{-1}\mathbf{y}$ and $\mathbf{R}_{\hat{\mathbf{h}}}^{-1} = (\Phi^*\mathbf{R}_v^{-1}\Phi + \mathbf{R}_h^{-1})$. Hence, the posterior of \mathbf{h} is

$$\begin{aligned}
p(\mathbf{h}|\mathbf{y}, I) &= \mathcal{CN}(\mathbf{h}; (\mathbf{R}_h^{-1} + \Phi^*\mathbf{R}_v^{-1}\Phi)^{-1}\Phi^*\mathbf{R}_v^{-1}\mathbf{y}, (\mathbf{R}_h^{-1} + \Phi^*\mathbf{R}_v^{-1}\Phi)^{-1}) \quad (3.3.3) \\
&= \mathcal{CN}(\mathbf{h}; \mathbf{R}_h\Phi^*(\Phi\mathbf{R}_h\Phi^* + \mathbf{R}_v)^{-1}\mathbf{y}, \mathbf{R}_h - \mathbf{R}_h\Phi^*(\Phi\mathbf{R}_h\Phi^* + \mathbf{R}_v)^{-1}\Phi\mathbf{R}_h). \quad (3.3.4)
\end{aligned}$$

In the last equality, we have used (3.A.5) and (3.A.7) in Appendix 3.A.2. The MMSE solution is given by the mean value of the pdf (3.3.3) or (3.3.4). The latter is the most common formulation of the MMSE solution, consistent with (3.1.5) and (3.1.6).

The formulation (3.3.3) is sometimes referred to as an *information* form of the MMSE filter. When $\mathbf{R}_h^{-1} = 0$ in (3.3.3), which means that no relevant prior info is available, and $\mathbf{R}_v = \sigma_v^2\mathbf{I}$ for some variance σ_v^2 , so that the noise is white with equal variance in all of its components, the MMSE solution coincides with the LS solution (3.1.3), as is easily seen from the information form.

The MMSE solution is a very powerful result in that it is a fairly simple expression and uses all available data in an optimal manner. However, since it requires multiplications and inversions of matrices of size N , where N is the length of the measurement vector \mathbf{y} , its complexity is $\mathcal{O}(N^3)$, which may be extremely high. Moreover, if new data arrives so that \mathbf{y} grows, then the MMSE solution as expressed by (3.3.4) and (3.3.3) does not immediately provide us with a way to incorporate this new data without carrying out the calculation all from the beginning. Furthermore, the vector of channel coefficients \mathbf{h} may also be growing with time. In channel estimation and prediction, incorporation of new data is essential, so we will require a framework that allows a periodic addition of new evidence and parameter space without the complexity of the algorithm growing out of proportions.

3.4 Inference based on state space models

Throughout this thesis we shall let a linear discrete-time state space model describe a *time varying* multidimensional channel \mathbf{h}_t :

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t, \quad (3.4.1)$$

$$\mathbf{h}_t = \mathbf{H}\mathbf{x}_t, \quad (3.4.2)$$

with the measurements described by

$$\mathbf{y}_t = \Phi_t \mathbf{h}_t + \mathbf{v}_t = \mathbf{J}_t \mathbf{x}_t + \mathbf{v}_t, \quad (3.4.3)$$

where $\mathbf{J}_t = \Phi_t \mathbf{H}$. The process noise \mathbf{u}_t , the measurement noise and interference \mathbf{v}_t , and the initial state \mathbf{x}_0 are multivariate circular symmetric complex Gaussian, and

$$E\left\{ \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \\ 1 \end{bmatrix}^* \right\} = \begin{pmatrix} \mathbf{Q}\delta_{ij} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}\delta_{ij} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Pi} & \mathbf{0} \end{pmatrix}, \quad (3.4.4)$$

where $\mathbf{Q} > 0$, $\mathbf{R} > 0$, and $\mathbf{\Pi} > 0$.

Comparing the state-space model (3.4.1), (3.4.2), (3.4.3) with the static model (3.1.1), we may consider the case where \mathbf{h} and \mathbf{y} in (3.1.1) correspond to a single time step in the state-space model. The advantage of the state-space model is then that it takes into account earlier measurements that aid in the estimation of \mathbf{h} .

More interestingly, we may consider the case where \mathbf{h} and \mathbf{y} comprise several time steps in the state-space formulation, so that e.g. $\mathbf{h} = [\mathbf{h}_0^T, \dots, \mathbf{h}_t^T]^T$ and $\mathbf{y} = [\mathbf{y}_0^T, \dots, \mathbf{y}_t^T]^T$. The optimal estimate of \mathbf{h} is given by the mean value of (3.3.4) with the covariance matrices \mathbf{R}_h and \mathbf{R}_v and the regressor matrix Φ being given implicitly by the state-space model, but this comes at a considerable computational complexity (cubic in the length of \mathbf{y}).

It has been shown that fading radio channels can be well represented by linear dynamic models [47],[4]. The reason for letting a discrete-time state space model represent the fading radio channel is practical; the state-space model expresses the channel in a way so that the next state depends only on the current state. In the next section we will see how this property allows optimal inferences about the channel state to be drawn in a recursive way, so that the numerical complexity of optimal channel estimation and prediction is linear in the number of measurements (to be compared with the cubic complexity of the MMSE estimator in Section 3.3).

As the state space model is motivated on practical grounds, so is the choice of Gaussian priors with known covariance matrices for the processes

$\{\mathbf{u}_t\}$, $\{\mathbf{v}_t\}$, and \mathbf{x}_0 , because these properties have to be assumed to allow for an optimal inference formulation. One could certainly argue that the pdfs for some or all of these processes should have upper bounds over which they should have no support, but that would lead to non-Gaussian distributions, which, as we will see presently, would prevent recursive inference to be carried out.

In the following discussion, the equations (3.4.1)–(3.4.4) constitute our prior information I .

3.4.1 Optimal estimation

The state vector \mathbf{x}_t represent the full dynamics of the channel \mathbf{h}_t . We therefore wish to infer the state vector from available measurements, and this should be done in an iterative manner, so that $p(\mathbf{x}_t|\mathbf{y}_0, \dots, \mathbf{y}_t, I)$ can be updated to $p(\mathbf{x}_{t+1}|\mathbf{y}_0, \dots, \mathbf{y}_{t+1}, I)$ when a new measurement arrives. Following [48] and denoting $\{\mathbf{y}_0, \dots, \mathbf{y}_t\} \triangleq \mathbf{Y}_t$, we find by the use of Bayes' theorem and marginalization that

$$\begin{aligned} p(\mathbf{x}_{t+1}|\mathbf{Y}_{t+1}, I) &= p(\mathbf{x}_{t+1}|\mathbf{y}_{t+1}, \mathbf{Y}_t, I) = \\ &= p(\mathbf{x}_{t+1}|\mathbf{Y}_t, I) \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \mathbf{Y}_t, I)}{p(\mathbf{y}_{t+1}|\mathbf{Y}_t, I)} = \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, I) p(\mathbf{x}_{t+1}|\mathbf{Y}_t, I)}{\int p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, I) p(\mathbf{x}_{t+1}|\mathbf{Y}_t, I) d\mathbf{x}_{t+1}} = \\ &= \frac{\int p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, I) p(\mathbf{x}_{t+1}|\mathbf{x}_t, I) p(\mathbf{x}_t|\mathbf{Y}_t, I) d\mathbf{x}_t}{\iint p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, I) p(\mathbf{x}_{t+1}|\mathbf{x}_t, I) p(\mathbf{x}_t|\mathbf{Y}_t, I) d\mathbf{x}_t d\mathbf{x}_{t+1}}. \end{aligned} \quad (3.4.5)$$

We have expressed the original pdf $p(\mathbf{x}_{t+1}|\mathbf{Y}_{t+1}, I)$ in terms of $p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, I)$, which is given by the measurement equation (3.4.3); $p(\mathbf{x}_{t+1}|\mathbf{x}_t, I)$, which is given by the state equation (3.4.1); and $p(\mathbf{x}_t|\mathbf{Y}_t, I)$, which we assume is known, since we seek a recursive update of the distribution for the state vector. By the model (3.4.1)–(3.4.4), all these pdfs are Gaussian. The multiplications, integrations, and division in (3.4.5) will preserve Gaussianity so that also $p(\mathbf{x}_{t+1}|\mathbf{Y}_{t+1}, I)$ is Gaussian and can be written on the form

$$p(\mathbf{x}_{t+1}|\mathbf{Y}_{t+1}, I) = \mathcal{CN}(\mathbf{x}_{t+1}; \hat{\mathbf{x}}_{t+1|t+1}, \mathbf{P}_{t+1|t+1}). \quad (3.4.6)$$

We therefore seek to find the mean value $\hat{\mathbf{x}}_{t+1|t+1}$ and the covariance matrix $\mathbf{P}_{t+1|t+1}$.

In the above, observe the notation that we will use throughout this thesis: in subscripts of the type $t|t_0$, t indicates the current time while t_0 indicates that measurements up to and including time t_0 is available.

Carrying out the marginalizations in (3.4.5) using the result of Appendix 3.B.2, we find

$$\begin{aligned}
p(\mathbf{x}_{t+1}|\mathbf{Y}_{t+1}, I) &= \\
&\left(\int \mathcal{CN}(\mathbf{y}_{t+1}; \mathbf{J}_{t+1}\mathbf{x}_{t+1}, \mathbf{R}) \times \mathcal{CN}(\mathbf{x}_{t+1}; \mathbf{F}\mathbf{x}_t, \mathbf{G}\mathbf{Q}\mathbf{G}^*) \right. \\
&\quad \left. \times \mathcal{CN}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) d\mathbf{x}_t \right) / \left(\iint \mathcal{CN}(\mathbf{y}_{t+1}; \mathbf{J}_{t+1}\mathbf{x}_{t+1}, \mathbf{R}) \right. \\
&\quad \left. \times \mathcal{CN}(\mathbf{x}_{t+1}; \mathbf{F}\mathbf{x}_t, \mathbf{G}\mathbf{Q}\mathbf{G}^*) \times \mathcal{CN}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) d\mathbf{x}_t d\mathbf{x}_{t+1} \right) = \\
&\frac{\mathcal{CN}(\mathbf{y}_{t+1}; \mathbf{J}_{t+1}\mathbf{x}_{t+1}, \mathbf{R}) \times \mathcal{CN}(\mathbf{x}_{t+1}; \mathbf{F}\hat{\mathbf{x}}_{t|t}, \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*)}{\mathcal{CN}(\mathbf{y}_{t+1}; \mathbf{J}_{t+1}\mathbf{F}\hat{\mathbf{x}}_{t|t}, \mathbf{J}_{t+1}(\mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*)\mathbf{J}_{t+1}^* + \mathbf{R})}. \quad (3.4.7)
\end{aligned}$$

It should be noted that the fact that the above marginalization could be carried out symbolically leaned on the assumption of linear models (3.4.1), (3.4.2) and (3.4.3), as well as the assignment of Gaussian distributions to $\{\mathbf{u}_t\}$, $\{\mathbf{v}_t\}$, and \mathbf{x}_0 . Non-linear models and/or non-Gaussian distributions for the processes $\{\mathbf{u}_t\}$, $\{\mathbf{v}_t\}$, and \mathbf{x}_0 would generally permit a symbolic calculation of $p(\mathbf{x}_{t+1}|\mathbf{Y}_{t+1}, I)$.

By completing the squares with respect to \mathbf{x}_{t+1} in (3.4.7) along the lines of [48], yields, after quite some algebra,

$$\begin{aligned}
\hat{\mathbf{x}}_{t+1|t+1} &= \mathbf{F}\hat{\mathbf{x}}_{t|t} + \mathbf{A}_t\mathbf{J}_{t+1}^*(\mathbf{J}_{t+1}\mathbf{A}_t\mathbf{J}_{t+1}^* + \mathbf{R})^{-1}(\mathbf{y}_{t+1} - \mathbf{J}_{t+1}\mathbf{F}\hat{\mathbf{x}}_{t|t}), \text{ and} \\
\mathbf{P}_{t+1|t+1} &= \mathbf{A}_t - \mathbf{A}_t\mathbf{J}_{t+1}^*(\mathbf{J}_{t+1}\mathbf{A}_t\mathbf{J}_{t+1}^* + \mathbf{R})^{-1}\mathbf{J}_{t+1}\mathbf{A}_t,
\end{aligned} \quad (3.4.8)$$

where $\mathbf{A}_t = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*$. By (3.4.6), the pair $(\hat{\mathbf{x}}_{t+1|t+1}, \mathbf{P}_{t+1|t+1})$ define the probability density function for the state vector at time $t + 1$ given measurements up to and including time $t + 1$.

Again, since we have used only the rules of probability we know that (3.4.8) is in every way consistent with logic. The solution illuminates a couple of facts. Firstly, if the process $\{\mathbf{h}_t\}$ can be modelled on state space form, we see that there does exist an expression that lets us carry out the update

$$p(\mathbf{x}_t|\mathbf{Y}_t, I) \rightarrow p(\mathbf{x}_{t+1}|\mathbf{Y}_{t+1}, I). \quad (3.4.9)$$

Secondly, the model matrices $\{\mathbf{F}, \mathbf{G}, \mathbf{J}_t, \mathbf{Q}, \mathbf{R}\}$ are allowed to be time varying without impacting the optimality of the algorithm. In our case, we shall only require \mathbf{J}_t to be time varying.

The solution (3.4.8) is given by the Kalman filter (KF) which R. Kalman derived in his seminal 1960 paper, although he used a somewhat different approach.

3.4.2 A geometric formulation

The calculations leading to (3.4.8) is tailored to the specific problem of finding the pdf for \mathbf{x}_t given measurements \mathbf{Y}_t . If we would be interested in a slightly different formulation, say to find the pdf for \mathbf{x}_{t+10} , then we would have to carry out a corresponding derivation, using Bayes' theorem all over again. In later chapters, we will require a flexible framework that will let us derive pdf:s for channel estimates and predictions for arbitrary prediction ranges. Instead of using Bayes' theorem and marginalization, pdf:s may be derived using a geometric approach [49], which is more appealing to intuition and will prove to be very flexible. Recall that an inner product obeys the following rules:

1. Linearity: $\langle \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2, \mathbf{y} \rangle = \alpha_1 \langle \mathbf{x}_1, \mathbf{y} \rangle + \alpha_2 \langle \mathbf{x}_2, \mathbf{y} \rangle$
for any $\alpha_1, \alpha_2 \in \mathbb{C}$.
2. Reflexivity: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*$.
3. Nondegeneracy: $\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{0} \Leftrightarrow \mathbf{x} = \mathbf{0}$.

Noting that the cross-correlation operation $E\{\mathbf{x}\mathbf{y}^*\}$ conforms to these rules, we define the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle \triangleq E\{\mathbf{x}\mathbf{y}^*\}. \quad (3.4.10)$$

This inner product lets us regard *unknown vectors* as elements in an inner product space. A couple of facts should be noted regarding the inner product (3.4.10):

- $\langle \mathbf{x}, \mathbf{y} \rangle$ is generally a rectangular matrix.
- \mathbf{x} and \mathbf{y} need not have the same length.

That is, all vectors, regardless of length, live in the same inner product space with inner product defined according to (3.4.10). We stress the rather unusual fact that this inner product is not scalar, which is otherwise a property that is assumed for inner products by most textbooks and papers. Among the sparse literature that do consider matrix-valued inner products, see e.g. [50]. Inner product spaces are normed spaces and we define

$$\|\mathbf{x}\|^2 \triangleq \langle \mathbf{x}, \mathbf{x} \rangle. \quad (3.4.11)$$

A norm can then be defined as a Cholesky factor of $\|\mathbf{x}\|^2$, e.g. the unique lower triangular Cholesky factor. However, we shall only find use for the squared norm and so we do not need to uniquely define a norm.

With this formalism, covariance matrices can be handled algebraically as inner products, which provides convenient tools for manipulation of signals and optimization of estimates.

In the above, we are actually allowing ourselves a small degree of carelessness regarding notation. In Section 3.2.1 we were eager to point out that it is central to Bayesianism to be specific about what information underlies an inference process. Yet, when defining the inner product (3.4.10), no such specification was made. In most uses of the inner product notation it will however be clear what the background information is. When we speak of vectors that are not estimates of other vectors we will be referring to prior distributions, so that for example $\langle \mathbf{x}_t, \mathbf{u}_t \rangle = E\{\mathbf{x}_t \mathbf{u}_t^* | I\}$. When we operate on vectors referring to estimates, the background information is defined by the variable names themselves, e.g. $\langle \hat{\mathbf{x}}_{t+2|t-1}, \hat{\mathbf{x}}_{t+1|t-1} \rangle = E\{\hat{\mathbf{x}}_{t+2|t-1} \hat{\mathbf{x}}_{t+1|t-1}^* | \mathbf{Y}_{t-1}, I\}$. The careless notation does however illuminate another fact: the inner product space is not actually home to vectors such as \mathbf{x} and \mathbf{y} , but rather to probability entities such as $(\mathbf{x}|I)$ and $(\mathbf{y}|I)$, i.e. vectors in combination with background information. Whereas a “normal” (Euclidean) vector space can house vectors all having the same length, an inner product space of the kind presented here can contain vectors (probability entities) of different lengths, *but only if they all have the same background information I*.

In accordance with the MSE minimization (3.1.4) we will seek an estimate $\hat{\mathbf{x}}$ of a vector \mathbf{x} , expressed as a linear combination of the measurements, that minimizes the mean value of the error

$$\tilde{\mathbf{x}} \triangleq \mathbf{x} - \hat{\mathbf{x}}. \quad (3.4.12)$$

With the geometric formulation of the problem, the optimal estimate is then the projection of the parameter of interest onto the subspace spanned by given data:

$$\hat{\mathbf{x}} = \text{projection of } \mathbf{x} \text{ onto } \mathcal{L}\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t\}, \quad (3.4.13)$$

where $\mathcal{L}\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t\}$ is the linear subspace spanned by $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_t\}$. By construction, this means that $\langle \hat{\mathbf{x}}, \tilde{\mathbf{x}} \rangle = 0$.

The concept of projection suggests the use of the Gram-Schmidt procedure. The Gram-Schmidt procedure finds an orthogonal basis $\{\mathbf{e}_t\}$ that spans the same subspace as a generally non-orthogonal set of vectors $\{\mathbf{y}_t\}$. Once this basis has been found, finding a projection is a simple matter of summing the individual projections of the vector onto the respective base vectors:

$$\hat{\mathbf{x}} = \sum_{i=0}^t \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i, \quad (3.4.14)$$

where $\mathbf{R}_{e,i} = \|\mathbf{e}_i\|^2$.

In the Gram-Schmidt procedure, the orthogonal basis $\{\mathbf{e}_t\}$ is found by iteratively forming normals to hyperplanes spanned by earlier base vectors:

$$\begin{aligned}\mathbf{e}_0 &= \mathbf{y}_0, \\ \mathbf{e}_1 &= \mathbf{y}_1 - \langle \mathbf{y}_1, \mathbf{e}_0 \rangle \mathbf{R}_{e,0}^{-1} \mathbf{e}_0, \\ \mathbf{e}_2 &= \mathbf{y}_2 - \langle \mathbf{y}_2, \mathbf{e}_1 \rangle \mathbf{R}_{e,1}^{-1} \mathbf{e}_1 - \langle \mathbf{y}_2, \mathbf{e}_0 \rangle \mathbf{R}_{e,0}^{-1} \mathbf{e}_0, \\ \mathbf{e}_3 &= \mathbf{y}_3 - \langle \mathbf{y}_3, \mathbf{e}_2 \rangle \mathbf{R}_{e,2}^{-1} \mathbf{e}_2 - \langle \mathbf{y}_3, \mathbf{e}_1 \rangle \mathbf{R}_{e,1}^{-1} \mathbf{e}_1 - \langle \mathbf{y}_3, \mathbf{e}_0 \rangle \mathbf{R}_{e,0}^{-1} \mathbf{e}_0,\end{aligned}$$

and so on. The number of terms seems to be growing with time, but owing to the state space formulation we can write

$$\mathbf{e}_t = \mathbf{y}_t - \sum_{i=0}^{t-1} \langle \mathbf{y}_t, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i = \mathbf{y}_t - \mathbf{J}_t \sum_{i=0}^{t-1} \langle \mathbf{x}_t, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i, \quad (3.4.15)$$

where we have exploited the measurement equation (3.4.3) which gives that $\langle \mathbf{y}_t, \mathbf{e}_i \rangle = \mathbf{J}_t \langle \mathbf{x}_t, \mathbf{e}_i \rangle$ (note that \mathbf{x}_t and \mathbf{e}_i do not generally have the same length). The sum in (3.4.15) is identified as

$$\sum_{i=0}^{t-1} \langle \mathbf{x}_t, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i = \text{proj}(\mathbf{x}_t, \mathcal{L}\{\mathbf{y}_0, \dots, \mathbf{y}_{t-1}\}) \triangleq \hat{\mathbf{x}}_{t|t-1}, \quad (3.4.16)$$

where $\text{proj}(\mathbf{x}_t, \mathcal{L}\{\mathbf{y}_0, \dots, \mathbf{y}_{t-1}\})$ is the projection of \mathbf{x}_t onto the linear subspace $\mathcal{L}\{\mathbf{y}_0, \dots, \mathbf{y}_{t-1}\}$ so that

$$\mathbf{e}_t = \mathbf{y}_t - \mathbf{J}_t \hat{\mathbf{x}}_{t|t-1}. \quad (3.4.17)$$

The process $\{\mathbf{e}_t\}$ is called the *innovations*. By construction, the innovations are white: $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \mathbf{0}$, $i \neq j$. The innovations form an orthogonal basis spanning the same subspace as the measurements, i.e.

$$\mathcal{L}\{\mathbf{y}_0, \dots, \mathbf{y}_t\} = \mathcal{L}\{\mathbf{e}_0, \dots, \mathbf{e}_t\}. \quad (3.4.18)$$

for any t . Since $\hat{\mathbf{x}}_{t+1|t} \in \mathcal{L}\{\mathbf{e}_0, \dots, \mathbf{e}_t\}$, it holds that $\langle \hat{\mathbf{x}}_{t+1|t}, \mathbf{e}_i \rangle = \mathbf{0}$ when $i > t$.

Letting the parameter of interest, \mathbf{x} , be the state vector one time step into the future, that is \mathbf{x}_{t+1} , we see from (3.4.14) that the optimal estimate $\hat{\mathbf{x}}_{t+1|t}$ is a function of \mathbf{e}_t (and earlier innovations), while \mathbf{e}_t in turn is a function of $\hat{\mathbf{x}}_{t|t-1}$. It is therefore possible to find a recursive update

$$\hat{\mathbf{x}}_{0|-1} \rightarrow \mathbf{e}_0 \rightarrow \hat{\mathbf{x}}_{1|0} \rightarrow \mathbf{e}_1 \rightarrow \hat{\mathbf{x}}_{2|1} \rightarrow \dots$$

Projecting the one-step predictions onto the innovation vectors, we see that we can write

$$\hat{\mathbf{x}}_{t+1|t} = \sum_{i=0}^t \langle \mathbf{x}_{t+1}, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i = \sum_{i=0}^{t-1} \langle \mathbf{x}_{t+1}, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i + \langle \mathbf{x}_{t+1}, \mathbf{e}_t \rangle \mathbf{R}_{e,t}^{-1} \mathbf{e}_t. \quad (3.4.19)$$

From the state equation (3.4.1) we see that $\langle \mathbf{x}_{t+1}, \mathbf{e}_t \rangle = \mathbf{F}\langle \mathbf{x}_t, \mathbf{e}_t \rangle$ so that

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t-1} + \mathbf{F}\langle \mathbf{x}_t, \mathbf{e}_t \rangle \mathbf{R}_{e,t}^{-1} \mathbf{e}_t. \quad (3.4.20)$$

Optimal one-step predictions are therefore found by starting with an initial prediction $\hat{\mathbf{x}}_{0|-1}$ and then alternately compute innovations (3.4.17) and predictions (3.4.20). It remains to calculate $\langle \mathbf{x}_t, \mathbf{e}_t \rangle$ and $\mathbf{R}_{e,t}$, which should also be done in a recursive manner. The key to doing this is to express the two quantities in terms of the covariance matrix for the one-step estimation error, $\|\tilde{\mathbf{x}}_{t|t-1}\|^2$.

First, note that by (3.4.3) and (3.4.17) we can write

$$\mathbf{e}_t = \mathbf{J}_t \tilde{\mathbf{x}}_{t|t-1} + \mathbf{v}_t. \quad (3.4.21)$$

Defining

$$\|\tilde{\mathbf{x}}_{t|t_0}\|^2 \triangleq \mathbf{P}_{t|t_0}, \quad (3.4.22)$$

it then follows that

$$\|\mathbf{e}_t\|^2 = \mathbf{J}_t \mathbf{P}_{t|t-1} \mathbf{J}_t^* + \mathbf{R} \quad (3.4.23)$$

and, by linearity of the inner product,

$$\langle \mathbf{x}_t, \mathbf{e}_t \rangle = \langle \hat{\mathbf{x}}_{t|t-1} + \tilde{\mathbf{x}}_{t|t-1}, \mathbf{J}_t \tilde{\mathbf{x}}_{t|t-1} + \mathbf{v}_t \rangle = \mathbf{P}_{t|t-1} \mathbf{J}_t^*. \quad (3.4.24)$$

We now require a recursion for $\mathbf{P}_{t|t-1}$. To find this recursion, we make the following definitions:

Definition 3.4.1 (Covariance matrices for the state vector and state estimate vector). *The covariance of the state vector is defined as*

$$\Pi_t \triangleq \|\mathbf{x}_t\|^2. \quad (3.4.25)$$

The covariance of the state estimate vector is defined as

$$\Sigma_{t|t_0} \triangleq \|\hat{\mathbf{x}}_{t|t_0}\|^2. \quad (3.4.26)$$

Defining

$$\mathbf{K}_{f,t} \triangleq \langle \mathbf{x}_t, \mathbf{e}_t \rangle \mathbf{R}_{e,t}^{-1} \quad (3.4.27)$$

and

$$\mathbf{K}_{p,t} \triangleq \mathbf{F} \mathbf{K}_{f,t}, \quad (3.4.28)$$

we can then easily prove the following theorem:

Theorem 3.4.1 (Recursions for the state and one-step state prediction covariance matrices). *Consider Definition 3.4.1 and the model (3.4.1)–(3.4.4). Then it holds that*

$$\Pi_{t+1} = \mathbf{F} \Pi_t \mathbf{F}^* + \mathbf{G} \mathbf{Q} \mathbf{G}^*, \quad (3.4.29)$$

and that

$$\Sigma_{t+1|t} = \mathbf{F} \Sigma_{t|t-1} \mathbf{F}^* + \mathbf{K}_{p,t} \mathbf{R}_{e,t} \mathbf{K}_{p,t}^*. \quad (3.4.30)$$

Proof. The recursion (3.4.29) follows from (3.4.1) and the fact that $\langle \mathbf{x}_t, \mathbf{u}_t \rangle = \mathbf{0}$. The recursion (3.4.30) follows from (3.4.20) and the fact that $\langle \hat{\mathbf{x}}_{t|t-1}, \mathbf{e}_t \rangle = \mathbf{0}$. \square

Since $\langle \hat{\mathbf{x}}_{t|t_0}, \tilde{\mathbf{x}}_{t|t_0} \rangle = \mathbf{0}$, it must hold that

$$\mathbf{\Pi}_t = \Sigma_{t|t_0} + \mathbf{P}_{t|t_0}. \quad (3.4.31)$$

Using Theorem 3.4.1 and (3.4.31) for the special case $t_0 = t - 1$, we get a recursion for \mathbf{P}_t :

$$\begin{aligned} \mathbf{P}_{t+1|t} &= \mathbf{\Pi}_{t+1|t} - \Sigma_{t+1|t} \\ &= \mathbf{F}(\mathbf{\Pi}_{t|t-1} - \Sigma_{t|t-1})\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* - \mathbf{K}_{p,t}\mathbf{R}_{e,t}\mathbf{K}_{p,t}^* \\ &= \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* - \mathbf{K}_{p,t}\mathbf{R}_{e,t}\mathbf{K}_{p,t}^*. \end{aligned} \quad (3.4.32)$$

We can now summarize the Gram-Schmidt procedure in the form of the KF recursions (3.4.23), (3.4.28), (3.4.17), (3.4.20), and (3.4.32):

$$\begin{aligned} \mathbf{R}_{e,t} &= \mathbf{J}_t\mathbf{P}_t\mathbf{J}_t^* + \mathbf{R}, \\ \mathbf{K}_{p,t} &= \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{J}_t^*\mathbf{R}_{e,t}^{-1}, \\ \mathbf{e}_t &= \mathbf{y}_t - \mathbf{J}_t\hat{\mathbf{x}}_{t|t-1}, \\ \hat{\mathbf{x}}_{t+1|t} &= \mathbf{F}\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_{p,t}\mathbf{e}_t, \\ \mathbf{P}_{t+1|t} &= \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* - \mathbf{K}_{p,t}\mathbf{R}_{e,t}\mathbf{K}_{p,t}^*, \end{aligned}$$

initialized with known prior mean value $\hat{\mathbf{x}}_{0|-1}$ and prior error covariance matrix $\mathbf{P}_{0|-1}$. As an extra measure of convenience, the procedure can be divided into a *measurement update* step and a *time update* step:

- Kalman gain calculation:

$$\mathbf{R}_{e,t} = \mathbf{J}_t\mathbf{P}_t\mathbf{J}_t^* + \mathbf{R} \quad (3.4.33)$$

$$\mathbf{K}_{f,t} = \mathbf{P}_{t|t-1}\mathbf{J}_t^*\mathbf{R}_{e,t}^{-1} \quad (3.4.34)$$

- Measurement updates:

$$\mathbf{e}_t = \mathbf{y}_t - \mathbf{J}_t\hat{\mathbf{x}}_{t|t-1} \quad (3.4.35)$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_{f,t}\mathbf{e}_t \quad (3.4.36)$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_{f,t}\mathbf{J}_t)\mathbf{P}_{t|t-1} \quad (3.4.37)$$

- Time updates:

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t} \quad (3.4.38)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* \quad (3.4.39)$$

From the Kalman recursion we may also easily derive a recursion for the *filtered* state estimation vector:

$$\hat{\mathbf{x}}_{t+1|t+1} = \mathbf{F}\hat{\mathbf{x}}_{t|t} + \mathbf{K}_{f,t+1}\mathbf{e}_{t+1} \quad (3.4.40)$$

as well as a recursion for the covariance matrix for the filtered state estimation error vector:

$$\mathbf{P}_{t+1|t+1} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* - \mathbf{K}_{f,t+1}\mathbf{R}_{e,t+1}\mathbf{K}_{f,t+1}^* \quad (3.4.41)$$

When studying channel estimation and prediction, it may be of interest to generate channel estimates/predictions from channel measurements. It is then useful to rearrange a few equations in the KF recursions and formulate a state space model for the state estimates:

$$\begin{aligned} \hat{\mathbf{x}}_{t+1|t} &= \mathbf{F}\hat{\mathbf{x}}_{t|t-1} + \mathbf{F}\mathbf{K}_{f,t}\mathbf{e}_t, \\ \mathbf{y}_t &= \mathbf{J}_t\hat{\mathbf{x}}_{t|t-1} + \mathbf{e}_t. \end{aligned} \quad (3.4.42)$$

Note however, that the “process noise” and “measurement noise” in this model is the same process $\{\mathbf{e}_t\}$ with covariance matrix $\mathbf{R}_{e,t}$.

3.4.3 Prediction and smoothing

The state equation (3.4.1) makes KF prediction trivial. Assuming $t \geq t_0$, optimal predictions are found recursively by

$$\hat{\mathbf{x}}_{t+1|t_0} = \sum_{i=0}^{t_0} \langle \mathbf{x}_{t+1|t_0}, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i = \mathbf{F} \sum_{i=0}^{t_0} \langle \mathbf{x}_{t|t_0}, \mathbf{e}_i \rangle \mathbf{R}_{e,i}^{-1} \mathbf{e}_i = \mathbf{F}\hat{\mathbf{x}}_{t|t_0}, \quad (3.4.43)$$

and we assume that the filtered state estimation $\hat{\mathbf{x}}_{t_0|t_0}$ is known. When the state transition matrix \mathbf{F} is time static, we have that

$$\hat{\mathbf{x}}_{t|t_0} = \mathbf{F}^{t-t_0} \hat{\mathbf{x}}_{t_0|t_0}. \quad (3.4.44)$$

The recursion for the many-step state prediction covariance matrix follows directly from (3.4.43):

$$\Sigma_{t+1|t_0} = \mathbf{F}\Sigma_{t|t_0}\mathbf{F}^*, \quad (3.4.45)$$

initialized e.g. with $\Sigma_{t_0+1|t_0}$ which is given by Theorem 3.4.1.

The many-step state prediction error covariance matrix is also calculated recursively:

$$\mathbf{P}_{t+1|t_0} = \|\tilde{\mathbf{x}}_{t+1|t_0}\|^2 = \|\mathbf{F}\tilde{\mathbf{x}}_{t|t_0} + \mathbf{G}\mathbf{u}_t\|^2 = \mathbf{F}\mathbf{P}_{t|t_0}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*, \quad (3.4.46)$$

initialized e.g. with $\mathbf{P}_{t_0+1|t_0}$ obtained with (3.4.32).

In Chapter 7, we will need to know the *prior* pdf for a predicted state vector $\hat{\mathbf{x}}_{t|t_0}$. Since the process noise \mathbf{u}_t , the measurement noise \mathbf{v}_t , and the initial state \mathbf{x}_0 are zero-mean processes, it follows that $E\{\hat{\mathbf{x}}_{t|t_0}|I\} = \mathbf{0}$. The state prediction covariance matrix is given by (3.4.45). The prior pdf for predicted state vector can then be written

$$p(\hat{\mathbf{x}}_{t|t_0}|I) = \mathcal{CN}(\hat{\mathbf{x}}_{t|t_0}; \mathbf{0}, \Sigma_{t|t_0}). \quad (3.4.47)$$

The case $t < t_0$ is called a *smoothing* problem. Kalman smoothing is not as straightforward as prediction and there are many variants to smoothing formulas. In this thesis, we will only be interested in calculating the state covariance matrices, which we find with the Bryson-Frazier formulas [51], presented here without proof:

$$\mathbf{P}_{t|t_0} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\Lambda_{t,t_0}\mathbf{P}_{t|t-1}, \quad (3.4.48)$$

where Λ_{t,t_0} is calculated recursively in reversed time through

$$\Lambda_{\tau,t_0} = \mathbf{F}_{p,\tau}^* \Lambda_{\tau+1,t_0} \mathbf{F}_{p,\tau} + \mathbf{J}_{\tau}^* \mathbf{R}_{e,\tau}^{-1} \mathbf{J}_{\tau}, \quad \Lambda_{t_0+1,t_0} = 0, \quad (3.4.49)$$

with $\mathbf{F}_{p,t}$ defined by

$$\mathbf{F}_{p,t} = \mathbf{F}(\mathbf{I} - \mathbf{K}_{f,t}\mathbf{J}_t). \quad (3.4.50)$$

With the Kalman recursions and the prediction and smoothing formulas, the full pdf of a state vector \mathbf{x}_t given measurements up to and including time t_0 are given by the mean value $\hat{\mathbf{x}}_{t|t_0}$ and the covariance matrix $\mathbf{P}_{t|t_0}$:

$$p(\mathbf{x}_t|\mathbf{y}_0, \dots, \mathbf{y}_{t_0}, I) = \mathcal{CN}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t_0}, \mathbf{P}_{t|t_0}). \quad (3.4.51)$$

3.5 Summary

In this chapter we have shown that inferences derived with the tools of probability theory are consistent with an extended logic. The Kalman filter (KF) was shown to be a special case of such inference. We also took the opportunity to derive the KF from a geometric approach, which hopefully gave the reader an intuitive feeling for the ubiquitous usefulness of the KF. So long as a fading channel or a sum of fading channels can be described on linear discrete-time state space form, the KF will always output a complete representation of our state of knowledge regarding the channel(s). This makes the KF an incredibly flexible instrument that can be used in many sorts of intricate problems. And our needs will be intricate; in Chapter 5, we consider irregular training signal designs (referred to as *pilot patterns* shortly), which necessitates the use of time-variant filters; in Chapter 6, we consider the reception of simultaneous signals coming from multiple sources, meaning that

superpositions of signals need to be modelled and inferred; and in Chapter 7 we study how probability density functions produced by the KF should be used properly.

3.A Some matrix results

3.A.1 Block triangular factorization

Consider a block matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad (3.A.1)$$

where \mathbf{A} and \mathbf{D} are invertible. Then it holds that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} - \mathbf{BD}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix}, \quad (3.A.2)$$

which is easily verified by carrying out the multiplication. The quantity $\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}$ is called the *Schur complement* to \mathbf{D} . Since

$$\begin{bmatrix} \mathbf{I} & \mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix}, \quad (3.A.3)$$

we may also easily derive a block factorization of the inverse of (3.A.1):

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \quad (3.A.4)$$

3.A.2 The matrix inversion lemma and variants

For non-singular matrices \mathbf{A} and \mathbf{C} it holds, given that all matrices have appropriate dimensions, that

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}. \quad (3.A.5)$$

This result, commonly referred to as *the matrix inversion lemma* (or the Woodbury matrix identity), is so commonly used in linear algebra that it is sometimes called just “the lemma”.

A simple rearrangement of (3.A.5) gives the following result:

$$\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{D} = \mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{A} \quad (3.A.6)$$

The following result will be useful:

$$(\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{B} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} - \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{C}^{-1}. \quad (3.A.7)$$

This can be proved as follows:

$$\begin{aligned} \mathbf{B} + \mathbf{BCDA}^{-1}\mathbf{B} &= \mathbf{B} + \mathbf{BCDA}^{-1}\mathbf{B} \\ \mathbf{BC}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}) &= (\mathbf{A} + \mathbf{BCD})\mathbf{A}^{-1}\mathbf{B} \\ (\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{B} &= \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{C}^{-1} \end{aligned}$$

In a similar manner,

$$\mathbf{C}\mathbf{D}(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} = (\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}, \quad (3.A.8)$$

which follows from

$$\begin{aligned} \mathbf{D} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\mathbf{C}\mathbf{D} &= \mathbf{D} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\mathbf{C}\mathbf{D} \\ (\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})\mathbf{C}\mathbf{D} &= \mathbf{D}\mathbf{A}^{-1}(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D}) \\ \mathbf{C}\mathbf{D}(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} &= (\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}. \end{aligned}$$

3.B Properties of the Gaussian distribution

3.B.1 Definition

The multivariate circular symmetric complex Gaussian (sometimes just Gaussian for short) distribution for $\mathbf{x} \in \mathbb{C}^{n \times 1}$ with mean value $\boldsymbol{\mu} \in \mathbb{C}^{n \times 1}$ and positive definite covariance matrix $\Sigma \in \mathbb{C}^{n \times n}$ is defined by

$$\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \Sigma) \triangleq \pi^{-n} |\Sigma|^{-1} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^* \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right). \quad (3.B.1)$$

3.B.2 Marginalization

Consider a multivariate Gaussian parameter vector \mathbf{x} with mean value $\mathbf{A}\boldsymbol{\mu}$ and covariance Σ , and that we are uncertain as to the value of $\boldsymbol{\mu}$. We assign a Gaussian distribution to $\boldsymbol{\mu}$ with mean value $\boldsymbol{\mu}_0$ and covariance Σ_0 . What is then the distribution for \mathbf{x} ? Carrying out the marginalization, we have

$$\begin{aligned} p(\mathbf{x} | \boldsymbol{\mu}_0, \mathbf{A}, \Sigma, \Sigma_0, I) &= \int \mathcal{CN}(\mathbf{x}; \mathbf{A}\boldsymbol{\mu}, \Sigma) \times \mathcal{CN}(\boldsymbol{\mu}; \boldsymbol{\mu}_0, \Sigma_0) d\boldsymbol{\mu} \\ &\propto \int \exp\left(-\frac{1}{2} \underbrace{\left((\mathbf{A}\boldsymbol{\mu} - \mathbf{x})^* \Sigma^{-1} (\mathbf{A}\boldsymbol{\mu} - \mathbf{x}) + (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^* \Sigma_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \right)}_c\right) d\boldsymbol{\mu} \end{aligned} \quad (3.B.2)$$

Examining now the exponent (except the factor $-1/2$) c , we can complete the squares and write it as

$$\begin{aligned} c &= (\boldsymbol{\mu} - \boldsymbol{\alpha})^* (\mathbf{A}^* \Sigma^{-1} \mathbf{A} + \Sigma_0^{-1}) (\boldsymbol{\mu} - \boldsymbol{\alpha}) \\ &\quad \underbrace{-\boldsymbol{\alpha}^* (\mathbf{A}^* \Sigma^{-1} \mathbf{A} + \Sigma_0^{-1}) \boldsymbol{\alpha} + \mathbf{x}^* \Sigma^{-1} \mathbf{x} + \boldsymbol{\mu}_0^* \Sigma_0^{-1} \boldsymbol{\mu}_0}_{\beta}, \end{aligned} \quad (3.B.3)$$

with $\boldsymbol{\alpha} = (\mathbf{A}^* \boldsymbol{\Sigma}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_0^{-1})^{-1} (\mathbf{A}^* \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0)$. Carrying out the integration over $\boldsymbol{\mu}$, the first term in (3.B.3) evaluates to a constant independent of \mathbf{x} and $\boldsymbol{\mu}_0$. The remaining three terms β can be written

$$\begin{aligned} \beta &= \mathbf{x}^* (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{A} (\mathbf{A}^* \boldsymbol{\Sigma}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_0^{-1})^{-1} \mathbf{A}^* \boldsymbol{\Sigma}^{-1}) \mathbf{x} \\ &\quad + \boldsymbol{\mu}_0^* (\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1} (\mathbf{A}^* \boldsymbol{\Sigma}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_0^{-1}) \boldsymbol{\Sigma}_0^{-1})^{-1} \boldsymbol{\mu}_0 \\ &\quad - \mathbf{x}^* \boldsymbol{\Sigma}^{-1} \mathbf{A} (\mathbf{A}^* \boldsymbol{\Sigma}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_0^{-1})^{-1} \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 \\ &\quad - \boldsymbol{\mu}_0^* \boldsymbol{\Sigma}_0^{-1} (\mathbf{A}^* \boldsymbol{\Sigma}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_0^{-1})^{-1} \mathbf{A}^* \boldsymbol{\Sigma}^{-1} \mathbf{x} \end{aligned} \quad (3.B.4)$$

Using (3.A.5), (3.A.6), and (3.A.7) on the respective terms in (3.B.4), we see that

$$\beta = (\mathbf{x} - \mathbf{A} \boldsymbol{\mu}_0)^* (\boldsymbol{\Sigma} + \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}^*)^{-1} (\mathbf{x} - \mathbf{A} \boldsymbol{\mu}_0) \quad (3.B.5)$$

so that

$$p(\mathbf{x} | \boldsymbol{\mu}_0, \mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_0, I) = \mathcal{CN}(\mathbf{x}; \mathbf{A} \boldsymbol{\mu}_0, \boldsymbol{\Sigma} + \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}^*). \quad (3.B.6)$$

Modelling MIMO-OFDMA systems

4.1 Introduction

Modern wireless cellular system concepts should be general enough to support user equipments (UEs) with a wide range of data transfer requirements, ranging from relatively simple devices requiring only a very limited transfer rate, to advanced devices with the need for supporting high speed data transfer. A UE in such a system is characterised by a large number of factors, such as its *Signal-to-Noise Ratio* SNR, its speed, its power availability, the number of antenna element it uses, its channels' frequency selectivity and fading characteristics, its hardware complexity, and so on. This diversity makes system design complex. Using different frequency bands and transmission techniques for different kinds of UEs is an option, but the sheer number of relevant characteristics makes that kind of solution unnecessarily inflexible and wasteful with the bandwidth resource.

Furthermore, with the introduction of MIMO techniques combined with opportunistic scheduling and link adaptation in wireless cellular systems, the need for simultaneously supporting many UEs with widely varying characteristics over the same frequency band is becoming evident. Whereas one UE may be stationary (nomadic) and indoor, featuring a frequency-selective but non-fading channel, another may have a high velocity and line-of-sight. Differences in requirement, complexity, and power availability may cause one UE to prefer a low output power and/or sending few reference symbols (henceforth *pilots*) while another may prefer sending many, possibly with high power.

To support channel equalization, link adaptation, and scheduling, the unknown fading radio channel needs to be inferred. Methods for channel estimation may be blind or pilot assisted. Throughout this thesis we shall assume pilot assisted channel estimation and prediction. To enable separate

estimation of channels from multiple transmit antennas (henceforth *inputs*), which is required in MIMO systems, a straightforward approach is to use disjoint sets of pilots for all inputs and all UEs. A problem is however that each input will then occupy somewhere between 5 and 10 percent of the resources, and this could quickly lead to training data taking over the whole bandwidth.

A different approach is to place pilots from different inputs at the same sub-locations, hence letting them overlap. The drawback of this approach is an increased filter complexity and possibly a reduced estimation and prediction performance. The advantage is that pilot overhead will not be affected by the number of tracked inputs.

A filter that tracks signals from multiple inputs needs to take into account as much information as possible about each and every input. Given the many characteristics describing each input, the number of different MIMO filters that would be needed to address every possible combination would be extremely large, making the use of a lookup table infeasible. Moreover, because of differences in coherence time, UEs should probably use filters of varying temporal extensions; a stationary UE should look far back in time when inferring the channel since it is nearly static, whereas the channel measured by a fast-moving UE will quickly get outdated, meaning that a short filter will suffice.

Thus there is the need for an algorithm that can be fed with all the information describing such things as channel fading characteristics and pilot arrangements, and from this produce a filter of appropriate length suited for channel estimation and prediction for the multiple inputs.

This is precisely what the Kalman filter (KF) recursions can do. Not only does the KF constitute the optimal channel estimator/predictor given input characteristics, but it does so also during the “transient” phase when these characteristics change.

In this chapter we present in detail how to model a multi-input OFDM system and how to use the model in a KF. The purpose is twofold. First, we want to examine the prospects of using the KF as an actual channel estimation/prediction implementation. For this purpose, we analyze the numerical complexity of the KF towards the end of the chapter. We will also study an alternative KF formulation that generally has a lower complexity than the standard KF formulation. We find however that this variant of the KF does not necessarily decrease numerical complexity in the present context. Second, we want to use the KF as an instrument for analyzing a system’s limits of performance.

Along with the optimal channel estimate/prediction, the KF also produces a measure of uncertainty about the estimate. This measure of uncertainty bears a direct correspondence to system performance. Also, the

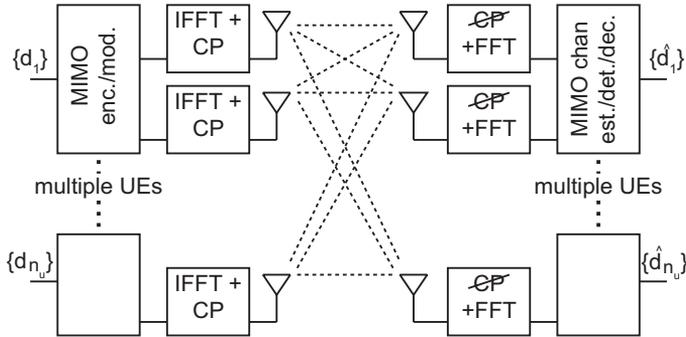


Figure 4.1: Schematic illustration of a multiuser MIMO system.

uncertainty measure is in fact data-independent and is therefore a function of the system parameters only. This means that conclusions about system performance (or rather upper bounds on system performance) can be drawn directly from the system model, without the need for conducting simulations. By setting up a state space system representation from the UEs characteristics and then studying the KF, quantitative conclusions about system performance can be drawn.

We consider a Multiple-Input Multiple-Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM) system. OFDM alleviates the problem of multipath propagation by dividing the system bandwidth into many narrowband flat fading subchannels. Relatively easily extendable to MIMO systems, MIMO-OFDM has therefore become a popular choice of transmission technique in modern wireless multiuser systems where multipath propagation is a major issue.

The present chapter shows, step-by-step, how to construct a state space representation of time-varying MIMO-OFDM channels. Such a representation can be used to represent fading channels for point-to-point transmission when the transmitter and/or the receiver are(is) equipped with array antennas. It can also be used in situations where multiple UEs are communicating with a BS over the same radio resource, so that the inputs are distributed among a set of different UEs. Also, the framework presented in this chapter can be used to model fading channels in a Coordinated MultiPoint (CoMP) setting, where a single UE is receiving signals from multiple BSs.

4.2 System model

The baseband multiuser MIMO-OFDM model for n_u users is shown in Figure 4.1. In this context, “ n_u users” refers to either n_u UEs communicating with

one BS in a multiuser scenario, or to n_u BSs communicating with one UE in a CoMP scenario. Data streams from each user are mapped onto symbols from a finite symbol alphabet using conventional techniques (encoding, bit-interleaving, space-time block coding, etc.) and serial-to-parallel converted to vectors of length N . Since each transmitter may be equipped with more than one transmit antenna, the number u of received data streams is generally larger than n_u . Each N -vector is processed by an IFFT to a vector of length N and a cyclic prefix of length N_{CP} is added. The u inputs (transmit antennas), which may be partitioned arbitrarily among the n_u users, then transmit the sequences of length $N + N_{CP}$ with sample period t_s after which each receiver antenna receives a noisy and channel distorted superposition of the u sequences. Between transmitter and receiver, we assume a timing synchronization well within the duration of a cyclic prefix, as well as a frequency synchronization error that is much smaller than the maximum Doppler frequency. The receiver antennas are in many cases separated such that all channels fade independently. In a channel estimation/prediction context we may then look at only one single receiver antenna without loss of generality or attainable performance. When discussing modelling, the number of receiver antennas are therefore irrelevant and we are in effect studying a multi-input single-output (MISO) system. Note however that the complexity of the receiver scales linearly with the number of receiver antennas, since each of them needs a channel estimator/predictor.

After removing the cyclic prefix, we saw in equation (2.4.5) in Chapter 2 that we can write the received sequence \mathbf{y}_f in the frequency domain as

$$\mathbf{y}_f = \text{diag}(\mathbf{s}_f)\mathbf{h}_f + \mathbf{v}_f, \quad (4.2.1)$$

where we have excluded time indices for brevity and the subscript $(\cdot)_f$ indicates that we are here considering the full system bandwidth, in total N subchannels. The complex-valued column vector \mathbf{h}_f represents the channel frequency response, and the noise¹ \mathbf{v}_f is assumed to be white, with covariance matrix $\sigma_v^2\mathbf{I}_N$. The column vector \mathbf{s}_f holds the transmitted symbols. When receiving signals from u simultaneous inputs, we hence have the received multiple-input (MI) signal

$$\begin{aligned} \mathbf{y}_{f,MI} &= \sum_{i=1}^u \text{diag}(\mathbf{s}_f^i)\mathbf{h}_f^i + \mathbf{v}_f \\ &= \Phi_f\mathbf{h}_{f,MI} + \mathbf{v}_f. \end{aligned} \quad (4.2.2)$$

¹The noise term represents both thermal noise and interference from other transmitters. The latter is appropriately modelled as white noise since the transmitted symbols are unknown.

where \mathbf{y} , \mathbf{s}^i , \mathbf{h}^i , and \mathbf{v} are all column vectors of length w and given by $\mathbf{y} = \mathbf{E}\mathbf{y}_{f,MI}$, $\mathbf{s}^i = \mathbf{E}\mathbf{s}_{f,i}^i$, $\mathbf{h}^i = \mathbf{E}\mathbf{h}_{f,i}^i$, and $\mathbf{v} = \mathbf{E}\mathbf{v}_f$. The matrix Φ , which is called to *pilot* matrix, is constructed as in (4.2.3), but with the vectors $\{\mathbf{s}_f^i\}$ exchanged for $\{\mathbf{s}^i\}$, and the column vector $\mathbf{h} = [(\mathbf{h}^1)^T \dots (\mathbf{h}^u)^T]^T$.

We will assume that the w subchannels hold known training symbols (pilots) so that Φ is known. Alternatively, decoded payload information may be used to construct Φ using iterative channel estimation (ICE), see Appendix 4.D. Hence we assume the use of pilot symbol assisted modulation (PSAM). In PSAM, known reference symbols (pilots) are intermixed with payload symbols. The pilots are typically laid out according to a uniform time-frequency grid and the channel estimation/prediction filter that is to be constructed operates on – observes – this grid only. The channel estimates obtained from these sub-locations are subsequently interpolated/extrapolated to the sub-locations of all payload sub-symbols.

Note that, analogous to (2.4.5), we can write

$$\mathbf{h}^i = N^{1/2}\mathbf{E}\mathcal{F}\mathbf{g}^i = N^{1/2}\mathcal{F}_w\mathbf{g}^i, \quad (4.2.7)$$

where \mathbf{g}^i is the baseband channel impulse response for input i , zero-padded to length N , and the partial Fourier matrix $\mathcal{F}_w = \mathbf{E}\mathcal{F}$ constitutes w rows from \mathcal{F} and is explicitly defined below in (4.2.23). From (4.2.6), it then follows that

$$\mathbf{y} = \sum_{i=1}^u \text{diag}(\mathbf{s}^i)N^{1/2}\mathcal{F}_w\mathbf{g}^i + \mathbf{v}. \quad (4.2.8)$$

Whether we will use (4.2.6) with frequency-domain channels $\{\mathbf{h}^i\}$, or (4.2.8) with time-domain channels $\{\mathbf{g}^i\}$, will depend on the situation. In case the time-domain representation (4.2.8) is used, we will attempt to model the most significant taps in $\{\mathbf{g}^i\}$.

The system considered is a single user or multiuser OFDM system, possibly MIMO, where each receiver antenna experiences a time varying flat fading or frequency selective channel. Each received channel is therefore characterized by one or many time varying channel coefficients. We will here model the channels on linear state space form. The state space model will be constructed hierarchically, from the modelling of one single channel coefficient, via the modelling of a single receiver antenna and a single user, up to the whole multi-input system model.

We shall begin by constructing a set of matrices $\{\mathbf{F}, \mathbf{G}, \mathbf{J}_t, \mathbf{Q}, \mathbf{R}, \Pi_0\}$ that characterises the state space. This model may then be used to construct an optimal observer of the channel coefficients, or we may draw inferences about different aspects of the system directly from the model.

4.2.1 Autoregressive modelling

We begin by considering a single channel coefficient, by which we mean either a tap in an impulse response or the complex-valued scalar channel for a single subchannel. The behaviour of each coefficient is determined by the local scattering environment and the velocity of the UE. The channel is sampled with a sampling rate determined by the period between two consecutive pilot-bearing OFDM symbols, which we denote t_p .

As described in Section 2.3, a mobile UE will experience fading of the received signal. The spaced time correlation function of a channel (denoted $A_c(\Delta t)$ in Chapter 2) describes the correlation properties of this fading. Although the fading in itself may be rapid, the spaced time correlation function will generally change only slowly with time. For any short period of time, a static ARMA model for the fading channel coefficient can be constructed to well represent the correct fading behaviour [4]. For complexity reasons it is important to keep the model order low. Since fading channel coefficients commonly exhibit oscillatory behaviour, autoregressive (AR) models are suitable. A fading tap h_t in the impulse response \mathbf{g}_t is modelled by

$$h_t + a_1 h_{t-1} + \dots + a_k h_{t-k} = u_t, \quad (4.2.9)$$

where k is the model order and u_t is the process noise that excites the process. The AR parameters $\{a_i\}_{i=1}^k$ should be appropriately adjusted according to the local environment. As briefly discussed in Chapter 1, the AR parameters can either be estimated based on blocks of measurements, or continuously tracked with e.g. the LMS method or the RLS method. The noisy measurements on which the AR parameter estimation is based is here considered to be prior information. In Chapter 8, we will consider various aspects of AR parameter estimation on real channel measurements. Among these aspects is the duration over which a channel may be considered to be static, which we in the investigated case find to be at least 0.5 seconds.

If no measurements are available, the AR parameters may be assigned based on less informative prior information. Below, we discuss two types of AR models that are based on different kinds of prior information. These models will be used in subsequent chapters.

The first type of AR model discussed here is *Jakes'* model [52],[37], which is motivated by the prior information that equally distant and evenly distributed scatterers surround the moving UE. Jakes' model attributes the autocorrelation function

$$r_t(\tau) \triangleq E\{h_t h_{t+\tau}^*\} = J_0(\Omega_D \tau) \quad (4.2.10)$$

to a fading tap h_t in \mathbf{g}_t . The symbol $(\cdot)^*$ denotes complex conjugation. Here, J_0 is the zeroth order Bessel function of the first kind, $\Omega_D = 2\pi f_D t_p$

is the normalised maximum angular Doppler frequency, and τ is the time lag expressed in samples. The Doppler frequency f_D is proportional to the UE velocity v (see Table 2.1 in Chapter 2). To the autocorrelation function (4.2.10) corresponds the classic U-shaped Doppler spectrum illustrated in Figure 4.2.

An AR process (4.2.9) of finite order cannot mimic the autocorrelation (4.2.10) exactly. Instead, the AR parameters need to be fitted in some way to the desired model. Multiplying (4.2.9) with $h_{t-l_i}^*$ for some integer l_i and taking the expected value, we get

$$E\{h_t h_{t-l_i}^*\} + a_1 E\{h_{t-1} h_{t-l_i}^*\} + \dots + a_k E\{h_{t-k} h_{t-l_i}^*\} = E\{u_t h_{t-l_i}^*\} = 0, \quad (4.2.11)$$

which for $l_i = 0, \dots, k-1$ are the Yule-Walker equations [53]. By choosing a set of integers $\{l_0, l_1, \dots\}$ larger than the model order k , we obtain an overdetermined set of equations [7]:

$$\begin{pmatrix} J_0(\Omega_D(l_0 - 1)) & \dots & J_0(\Omega_D(l_0 - k)) \\ J_0(\Omega_D(l_1 - 1)) & \dots & J_0(\Omega_D(l_1 - k)) \\ \vdots & & \vdots \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} = - \begin{pmatrix} J_0(\Omega_D l_0) \\ J_0(\Omega_D l_1) \\ \vdots \end{pmatrix}. \quad (4.2.12)$$

The AR parameters $\{a_i\}_{i=1}^k$ can then be found from the least squares solution of (4.2.12). The $\{l_i\}$ are the time lags in the autocorrelation function (4.2.10) for which the best fit is desired. Figure 4.2 illustrates the spectrum of an AR4 model fitted to the Jakes autocorrelation for indices $\{l_i\} = \{1, 51, 101, 151, \dots, 451\}$.

Equation (4.2.9) can be written on operator form as

$$h_t = \frac{1}{1 + a_1 q^{-1} + \dots + a_k q^{-k}} u_t \quad (4.2.13)$$

or by means of the z-transform as

$$h(z) = \frac{z^k}{z^k + a_1 z^{k-1} + \dots + a_k} u(z) = \frac{z^k}{(z - p_1)(z - p_2) \dots (z - p_k)} u(z), \quad (4.2.14)$$

where q^{-1} is the backward unit delay operator. Once the parameters $\{a_i\}$ have been found, the corresponding poles $\{p_i\}$ can be found from (4.2.14).

It should be noted that the matrix in (4.2.12) is in general poorly conditioned which may cause the least squares solution to yield unstable poles. Such poles should be reflected in the unit circle, so that a stable model is constructed while preserving the spectral density.

The second type of AR model is a model that is based on less informative prior information than Jakes' model. Here, we are ignorant to what the

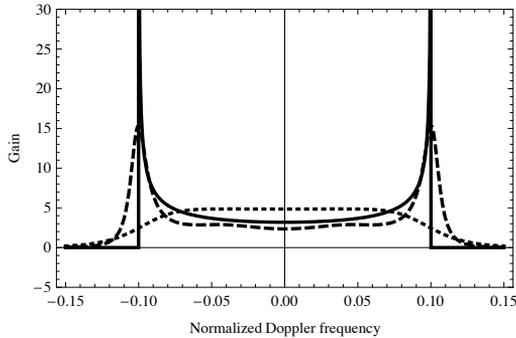


Figure 4.2: Doppler spectra for three different models with normalized Doppler frequency $\Omega_D = 0.1$. The solid line is the standard Jakes model. The dashed line is an AR4 process which has been fitted to ten elements with spacing 50 in the Jakes autocorrelation function. The dotted line is an AR4 model derived from a Butterworth filter, yielding a flat Doppler spectrum.

Doppler frequency of the received channel coefficient is, except that it is upper bounded by the maximum Doppler frequency f_D . Based on this ignorance, we should then assign a Doppler spectrum that attributes the same probability density to any Doppler frequency in the range $[-f_D, f_D]$. AR models with flat Doppler spectra can be constructed from continuous-time Butterworth filters. These have poles evenly distributed over a half-circle with the “prewarped” radius $2 \tan(\Omega_D/2)/t_p$ in the negative half of the complex plane. Each continuous-time pole p_s can then be translated to a discrete-time pole p_z e.g. through the bilinear transform $p_z = (2/t_p + p_s)/(2/t_p - p_s)$. The flat Doppler spectrum of a 4th order AR process is shown in Figure 4.2.

Of the two types of AR models presented above, the Jakes type model will yield the best estimation and prediction performance, since its Doppler spectrum features strong spectral components that makes it easy to extrapolate the channel into the future. Even better performance will follow from a Doppler spectrum with a single strong component, which will sometimes be the case with on-line AR parameter estimation on real channel measurements.

4.2.2 State space modelling

Once the poles of the AR process have been found, we seek to construct a state space model

$$\begin{aligned} \mathbf{x}_{t+1}^{(c)} &= \mathbf{X}\mathbf{x}_t^{(c)} + \mathbf{Y}u_t^{(c)}, \\ h_t^{(c)} &= \mathbf{Z}\mathbf{x}_t^{(c)}, \end{aligned} \tag{4.2.15}$$

where $h_t^{(c)}$ is the fading channel coefficient (either in the frequency domain or in the time domain), $\mathbf{x}_t^{(c)}$ is the state vector of length k , where k is the model order, and $\{u_t^{(c)}\}$ is the process noise² that “drives” the process. The superscript $(\cdot)^{(c)}$ indicates that we at this point consider modelling of individual channel coefficients. The variance of $u_t^{(c)}$, which ultimately dictates the variance of the coefficient $h_t^{(c)}$ is uninteresting at this moment – the focus here is on the time dynamics of $h_t^{(c)}$.

The matrices $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ may generally be time-varying, but throughout this chapter we shall assume that they are varying slowly enough that they may be considered static. In Chapter 8, we will show that channel models may typically be held static for at least half a second at a time.

As shall be evident later, it is favourable to choose a state space model on diagonal form, so that \mathbf{X} is diagonal. This can be accomplished by putting the AR poles on the diagonal of \mathbf{X} . The elements of $\mathbf{X} \in \mathbb{C}^{k \times k}$, $\mathbf{Y} \in \mathbb{C}^{k \times 1}$, and $\mathbf{Z} \in \mathbb{C}^{1 \times k}$ are set as follows [43]:

$$\begin{aligned} \mathbf{X}[i, i] &= p_i, \\ \mathbf{Y}[i] &= \prod_{\substack{j=1, \dots, k \\ j \neq i}} (p_i - p_j)^{-1}, \\ \mathbf{Z}[i] &= p_i^{k-1}, \quad i = 1, \dots, k. \end{aligned} \tag{4.2.16}$$

The matrices \mathbf{Y} and \mathbf{Z} may be rescaled by an arbitrary factor without changing the fading characteristics of the coefficient $h_t^{(c)}$. While generally \mathbf{X} and \mathbf{Z} both have a reasonable scaling, the elements of \mathbf{Y} are often extremely large, the reason being that poles are frequently located close to the unit circle. It may therefore be a good idea to normalise \mathbf{Y} , e.g. by the magnitude of its largest element.

Note that, to be consistent with notation used in later models, we use matrix notation for \mathbf{Y} and \mathbf{Z} , although they are vectors in the model (4.2.15).

²The process noise $\{u_t^{(c)}\}$ is the same as the process $\{u_t\}$ in (4.2.9) except for a time shift. This time shift does not matter however since the process is white.

EXAMPLE 4.2.1 Channel coefficient model

Assume that we want to model a fading channel coefficient (i.e. a tap in an impulse response or the complex-valued scalar channel for a single sub-channel) experienced by a terminal moving at a speed of 72 km/h. The coefficient is the sum of a large number of reflections coming from scatterers evenly distributed around the terminal, all at approximately the same distance. The carrier frequency is $f_c=3$ GHz and the sampling period (the period between pilot-bearing OFDM symbols) is $t_p = 200\mu\text{s}$. The model order is set to $k = 2$.

The Doppler frequency is here $f_D = f_c v/c_0=200$ Hz, and the normalized angular Doppler frequency is $\Omega_D = 2\pi f_D t_p = 0.2513$. The geometry of the problem motivates a Jakes Doppler spectrum. In the approximating second order AR model, we choose the poles $p_1 = 0.95 \exp(j\Omega_D)$ and $p_2 = 0.95 \exp(-j\Omega_D)$, which gives strong emphasis to the maximum Doppler frequency f_D (and its negative counterpart), as does Jakes' model. The radius 0.95 was chosen arbitrarily but should be close to unity.

According to (4.2.16), we set

$$\mathbf{X} = \begin{pmatrix} p_1 & 0 \\ 0 & p_2 \end{pmatrix} = \begin{pmatrix} 0.9202 + 0.2363j & 0 \\ 0 & 0.9202 - 0.2363j \end{pmatrix}, \quad (4.2.17)$$

$$\mathbf{Y} = \begin{pmatrix} (p_1 - p_2)^{-1} \\ (p_2 - p_1)^{-1} \end{pmatrix} = \begin{pmatrix} -2.1164j \\ 2.1164j \end{pmatrix}, \quad (4.2.18)$$

and

$$\mathbf{Z} = (p_1 \ p_2) = (0.9202 + 0.2363j \ 0.9202 - 0.2363j). \quad (4.2.19)$$

(Note that while \mathbf{Y} in many cases may have very large elements, requiring a normalization, this is not the case here.) A realization of $\{h_t^{(c)}\}$ from the state space model $\mathbf{x}_{t+1}^{(c)} = \mathbf{X}\mathbf{x}_t^{(c)} + \mathbf{Y}u_t^{(c)}$, $h_t^{(c)} = \mathbf{Z}\mathbf{x}_t^{(c)}$ is displayed in Figure 4.3. The complex white noise $u_t^{(c)}$ was given unit variance. As can be seen from Figure 4.3, the wavelength is about 0.1 m, as should be since the carrier frequency is 3 GHz. We also plot the power spectrum, which exhibit strong peaks at $\pm f_D$, just like the Jakes spectrum. Note that the general amplitude of $h_t^{(c)}$ is high, since the poles are located near the unit circle. However, scaling is not an issue at this point.

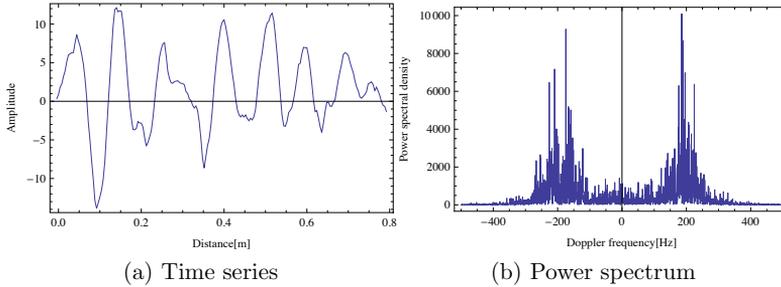


Figure 4.3: Realization of the one-coefficient state space model.

4.2.3 Multipath channel model

To model the signal coming from one input, we need to model each fading channel coefficient in that signal. We will consider tracking w parallel subchannels, over which pilots are transmitted. The ability to track these subchannels will depend on how correlated they are, which in turn depends on spacing in frequency and also on the frequency properties – mainly the coherence bandwidth – of the channel. The parameter w , which henceforth shall be called the *filter width*, is a design parameter with which the system designer can trade off channel estimation performance for complexity. In Chapters 5 and 6 we shall study how different settings of w affect channel estimation and prediction performance.

Two different modelling approaches

There are two approaches to modelling the fading channel coefficients. Either we use *impulse response modelling* and model and track the significant taps in the impulse response, or we use *subchannel modelling* and model and track the w parallel subchannels. Which scheme to choose depends on a number of factors. In a real system, it is generally easier to use subchannel modelling since that model can be obtained relatively easily from channel measurements. Using impulse response modelling in a real system, on the other hand, requires estimating the number, gains, and delays of the significant taps in the impulse response, which typically requires the use of subspace methods for system identification of linear dynamic systems.

Let us denote the number of significant taps in the impulse response by m . When setting up a model with the purpose of using it for analyzing a system's performance, the issue of deciding which modelling method to use is largely determined by which of the numbers m and w is smallest. As a rule of thumb, if $m < w$ we would use impulse response modelling. If the converse is true we would use subchannel modelling. The numerical results

will be exactly the same, but there are significant computational benefits to be made by making the appropriate choice. We will in the following use the integer parameter x to denote either m or w , depending on the method chosen. Correspondingly, we use the notion *channel coefficient* to denote a subchannel or a significant tap in an impulse response, whichever applies.

If impulse response modelling is to be used, we need to assume some knowledge about the power delay profile of the channel. This knowledge consists of the delays and the powers of the m distinct taps. It is represented as two matrices \mathbf{R}_γ and \mathbf{D} . The first, \mathbf{R}_γ , is a diagonal $m \times m$ matrix with each diagonal entry representing the power (variance) of a tap. For example, assuming that the impulse response has three significant and mutually independent taps with the respective variances 6 dB, 6 dB, and 3 dB, so that the total power is 10 dB, we would set $\mathbf{R}_\gamma = \text{diag}(4, 4, 2)$.

The matrix \mathbf{D} has dimensions $N \times m$, where N is the number of samples in an OFDM symbol, defined so that

$$\mathbf{g}_t^{(m)} = \mathbf{D}\mathbf{h}_{\text{time},t}^{(m)}, \quad (4.2.20)$$

where $\mathbf{h}_{\text{time},t}^{(m)}$ holds the most significant taps in the sampled impulse response $\mathbf{g}_t^{(m)}$. The superscript $(\cdot)^{(m)}$ indicates that we now consider multipath modelling. The elements of \mathbf{D} are given by

$$\mathbf{D}[i, j] = f(it_s - \tau_j), \quad (4.2.21)$$

where $f(t)$ is the compound of the transmitter and the receiver pulse shapes, t_s is the system sampling period, and τ_j is the delay of the j :th tap in $\mathbf{h}_{\text{time},t}^{(m)}$. Assuming that $f(t)$ is a raised cosine pulse and that all $\{\tau_j\}$ are integer multiples of t_s , \mathbf{D} is a sparse matrix containing exactly m ones and zeros otherwise. For further details on discrete-time channel modelling, see e.g. [54].

A block-diagonal state-space model

The x channel coefficients (in the time or frequency domain) are modelled by setting up a block-diagonal state space model³

$$\begin{aligned} \mathbf{x}_{t+1}^{(m)} &= \text{diag}(\mathbf{X}, \dots, \mathbf{X})\mathbf{x}_t^{(m)} + \text{diag}(\mathbf{Y}, \dots, \mathbf{Y})\mathbf{u}_t^{(m)} \\ &= \mathbf{A}\mathbf{x}_t^{(m)} + \mathbf{B}\mathbf{u}_t^{(m)}, \\ \mathbf{h}_t^{(m)} &= \mathcal{T} \times \text{diag}(\mathbf{Z}, \dots, \mathbf{Z})\mathbf{x}_t^{(m)} = \mathcal{T}\mathbf{C}\mathbf{x}_t^{(m)}. \end{aligned} \quad (4.2.22)$$

Although the state space model (4.2.22) allows different models to be used for the respective channel coefficients, we will throughout this thesis assume

³For a specification of the involved matrix dimensions, see Appendix 4.H.

that the same model is employed for all channel coefficients. Since x channel coefficients are modelled simultaneously, the state vector $\mathbf{x}_t^{(m)}$ has length⁴ kx . $\mathbf{A} \in \mathbb{C}^{kx \times kx}$ is diagonal and $\mathbf{B} \in \mathbb{C}^{kx \times x}$ and $\mathbf{C} \in \mathbb{C}^{x \times kx}$ are block diagonal matrices with x blocks each. Note that the vector $\mathbf{C}\mathbf{x}_t$ are the x fading coefficients. However, since we are interested in modelling the w subchannels, regardless of whether we use impulse response modelling or subchannel modelling, we introduced a matrix \mathcal{T} with the purpose of transforming $\mathbf{C}\mathbf{x}^{(m)}$ to the frequency domain, if necessary.

When subchannel modelling is used, no such transform is needed and we set $\mathcal{T} = \mathbf{I}_w$. When impulse response modelling is employed so that $\mathbf{C}\mathbf{x}_t^{(m)} = \mathbf{h}_{\text{time},t}^{(m)}$, we set $\mathcal{T} = N^{1/2}\mathcal{F}_w\mathbf{D}$, where \mathcal{F}_w is the partial Fourier matrix with elements

$$\mathcal{F}_w[i, j] = N^{-1/2}e^{-2\pi js_i j/N}, \quad i = 0, \dots, w-1, j = 0, \dots, N-1, \quad (4.2.23)$$

where s_i is the index for subchannel i . Then, $\mathbf{h}^{(m)} = N^{1/2}\mathcal{F}_w\mathbf{g}^{(m)}$, in accordance with (4.2.7).

For an appropriate scaling of $\mathbf{h}_t^{(m)}$ we need to determine the x -by- x process noise covariance matrix $\|\mathbf{u}_t^{(m)}\|^2 \triangleq \mathbf{L}$. Let

$$\mathbf{h}'_t = \mathbf{C}\mathbf{x}_t^{(m)}, \quad (4.2.24)$$

so that either $\mathbf{h}'_t = \mathbf{h}_t^{(m)}$ or $\mathbf{h}'_t = \mathbf{h}_{\text{time},t}^{(m)}$, depending on the type of modelling. We first need to establish the channel coefficient covariance matrix $\mathbf{R}_{h'} \triangleq \|\mathbf{h}'_t\|^2$. For impulse response modelling this is simply $\mathbf{R}_{h'} = \mathbf{R}_\gamma$. For subchannel modelling, $\mathbf{R}_{h'}$ can often be estimated directly from noisy measurements of the time-frequency channel, which means that we do not have any need for the matrices \mathbf{R}_γ and \mathbf{D} . Alternatively, $\mathbf{R}_{h'}$ can be derived from the time domain information through $\mathbf{R}_{h'} = N\mathcal{F}_w\mathbf{D}\mathbf{R}_\gamma\mathbf{D}^*\mathcal{F}_w^*$.

It is generally non-trivial to calculate the process noise covariance matrix \mathbf{L} from the channel coefficient covariance matrix $\mathbf{R}_{h'}$. Fortunately, the special state space structure used here permits us to set \mathbf{L} explicitly by using the following theorem.

Theorem 4.2.1 (Scaling of covariance matrices). *Consider a state space model*

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t, \\ \mathbf{h}'_t = \mathbf{C}\mathbf{x}_t, \end{cases} \quad t \geq 0, \quad (4.2.25)$$

⁴The model order may differ between coefficients, so that the length of $\mathbf{x}^{(m)}$ is actually $\sum_{i=1}^x k_i$, where k_i is the model order of channel coefficient i . For brevity we shall however denote the length of $\mathbf{x}^{(m)}$ as kx . This notational convention is later extended to the system model, when several users are considered.

with multivariate circular symmetric complex Gaussian variables $\{\mathbf{u}, \mathbf{x}_0\}$ such that

$$\langle [\mathbf{u}^T \mathbf{x}_0^T]^T, [\mathbf{u}^T \mathbf{x}_0^T \mathbf{1}]^T \rangle = \begin{pmatrix} \mathbf{L} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Gamma_0 & \mathbf{0} \end{pmatrix} \quad (4.2.26)$$

and matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ such that

$$\begin{aligned} \mathbf{A} &= \text{diag}(\mathbf{a}), \\ \mathbf{B} &= \text{diag}(\mathbf{Y}_1, \dots, \mathbf{Y}_x), \\ \mathbf{C} &= \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_x), \end{aligned} \quad (4.2.27)$$

where, for some positive integers $\{k_1, \dots, k_x\}$, \mathbf{a} is a vector of length $\sum_{i=1}^x k_i$ whose complex elements have magnitudes strictly less than one, and where the $\{\mathbf{Y}_i\}$ are column vectors of the respective lengths $\{k_i\}$ and the $\{\mathbf{Z}_i\}$ are row vectors of the same respective lengths $\{k_i\}$. If, for a positive definite matrix $\mathbf{R}_{h'}$, we set

$$\mathbf{L} = \mathbf{R}_{h'} \oslash \mathbf{C}(\mathbf{B}\mathbf{1}\mathbf{B}^* \oslash (\mathbf{1} - \mathbf{a}\mathbf{a}^*))\mathbf{C}^*, \quad (4.2.28)$$

where $\mathbf{1}$ indicates a matrix of appropriate dimensions containing only ones, and the symbol \oslash indicates element-wise division, and we also set

$$\Gamma_0 = \mathbf{B}\mathbf{L}\mathbf{B}^* \oslash (\mathbf{1} - \mathbf{a}\mathbf{a}^*), \quad (4.2.29)$$

then the model will be stationary, so that the state covariance $\|\mathbf{x}_t\|^2 = \Gamma_t$ has a constant value $\bar{\Gamma}$ obeying

$$\bar{\Gamma} = \mathbf{A}\bar{\Gamma}\mathbf{A}^* + \mathbf{B}\mathbf{L}\mathbf{B}^*, \quad (4.2.30)$$

and the covariance matrix for the channel coefficients will be constant with value $\mathbf{R}_{h'}$:

$$\|\mathbf{h}'_t\|^2 = \mathbf{R}_{h'}. \quad (4.2.31)$$

Proof. Since $\mathbf{A} = \text{diag}(\mathbf{a})$, it holds that $\mathbf{A}\bar{\Gamma}\mathbf{A}^* = \mathbf{a}\mathbf{a}^* \odot \bar{\Gamma}$ (for any matrix $\bar{\Gamma}$), where \odot is element-wise multiplication. The solution $\bar{\Gamma}$ to the Lyapunov equation (4.2.30) is therefore given by

$$(\mathbf{1} - \mathbf{a}\mathbf{a}^*) \odot \bar{\Gamma} = \mathbf{B}\mathbf{L}\mathbf{B}^* \Rightarrow \bar{\Gamma} = \mathbf{B}\mathbf{L}\mathbf{B}^* \oslash (\mathbf{1} - \mathbf{a}\mathbf{a}^*). \quad (4.2.32)$$

Since $\Gamma_t = \bar{\Gamma}$ for all t , with $\Gamma_0 = \bar{\Gamma}$ as a special case, the result (4.2.29) follows.

To derive the result (4.2.28) we partition $\bar{\Gamma}$ into blocks $\bar{\Gamma}_{i,j}$ so that each $\bar{\Gamma}_{i,j}$ is a $k_i \times k_j$ block, where $1 \leq i, j \leq x$. We also partition the vector \mathbf{a} into sub-vectors $\{\mathbf{a}_i\}$ so that \mathbf{a}_i is a k_i -vector. From (4.2.30), it then holds that

$$\begin{aligned} \bar{\Gamma}_{i,j} &= \mathbf{a}_i \mathbf{a}_j^* \odot \bar{\Gamma}_{i,j} + \mathbf{L}[i, j] \mathbf{Y}_i \mathbf{Y}_j^* \\ \Rightarrow \bar{\Gamma}_{i,j} &= \mathbf{L}[i, j] \mathbf{Y}_i \mathbf{Y}_j^* \oslash (\mathbf{1} - \mathbf{a}_i \mathbf{a}_i^*). \end{aligned} \quad (4.2.33)$$

From (4.2.25), it follows that $\mathbf{R}_{h'} = \mathbf{C}\bar{\Gamma}\mathbf{C}^*$. We then find the elements of \mathbf{L} from

$$\begin{aligned} \mathbf{R}_{h'}[i, j] &= \mathbf{Z}_i(\mathbf{L}[i, j]\mathbf{Y}_i\mathbf{Y}_j^* \circledast (\mathbf{1} - \mathbf{a}_i\mathbf{a}_j^*))\mathbf{Z}_j^* \\ \Rightarrow \mathbf{L}[i, j] &= \mathbf{R}_{h'}[i, j]/(\mathbf{Z}_i(\mathbf{Y}_i\mathbf{Y}_j^* \circledast (\mathbf{1} - \mathbf{a}_i\mathbf{a}_j^*))\mathbf{Z}_j^*), \end{aligned} \quad (4.2.34)$$

from which the result (4.2.28) follows. \square

In accordance with Theorem 4.2.1, we set

$$\mathbf{L} = \mathbf{R}_{h'} \circledast \mathbf{C}(\mathbf{B}\mathbf{1}\mathbf{B}^* \circledast (\mathbf{1} - \mathbf{a}\mathbf{a}^*))\mathbf{C}^*, \quad (4.2.35)$$

where \mathbf{a} are the diagonal elements of \mathbf{A} . To force model stationarity, we also set

$$\Gamma_0 \triangleq \|\mathbf{x}_0^{(m)}\|^2 = \mathbf{B}\mathbf{L}\mathbf{B}^* \circledast (\mathbf{1} - \mathbf{a}\mathbf{a}^*). \quad (4.2.36)$$

The single-input channel model $\{\mathbf{A}, \mathbf{B}, \mathcal{TC}, \mathbf{L}, \Gamma_0\}$ now models the kx fading channel coefficients by means of the state space model (4.2.22), with a white noise vector $\mathbf{u}_t^{(m)}$ of dimension x as input.

EXAMPLE 4.2.2 Multipath channel model – modelling in the impulse response domain

A frequency selective channel is in this example well described by a two-tap model, i.e. $m = 2$. Both taps have the same fading behaviour as described in Example 4.2.1. The second tap has one third of the power of the first tap, and the delay between the two is $\tau_1 - \tau_0 = 1 \mu\text{s}$. We may assume that the first tap has delay $\tau_0 = 0$. The SNR is 10 dB. The system uses square-root raised cosine pulses with roll-off factor 0 at both transmitter and receiver, so that the compound pulse shape is a Nyquist pulse [55]. We wish to model two parallel subchannels, which have separation $\Delta f = 80 \text{ kHz}$, by modelling in the impulse response domain.

First we set $\mathbf{A} = \text{diag}(\mathbf{X}, \mathbf{X})$, $\mathbf{B} = \text{diag}(\mathbf{Y}, \mathbf{Y})$, and $\mathbf{C} = \text{diag}(\mathbf{Z}, \mathbf{Z})$, with $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ set according to Example 4.2.1.

Then we need to establish the transformation matrix $\mathcal{T} = N^{1/2}\mathcal{F}_w\mathbf{D}$. Since the pulses are Nyquist pulses and the Fourier transform of a Nyquist pulse is a complex sinusoid of finite duration, we have that $\mathcal{T}[i, j] = e^{j\omega_i\tau_j}$, where $\omega_1 - \omega_0 = 2\pi\Delta f$ and we may assume that $\omega_0 = 0$. This means that

$$\mathcal{T} = \begin{pmatrix} 1 & 1 \\ 1 & e^{j\omega_1\tau_1} \end{pmatrix} \approx \begin{pmatrix} 1 & 1 \\ 1 & e^{j/2} \end{pmatrix}. \quad (4.2.37)$$

Finally, we need to determine the process noise covariance matrix \mathbf{L} . Since the two taps are modelled in the impulse response domain and the

total SNR is 10, we have $\mathbf{R}_\gamma = \text{diag}(7.5, 2.5)$ (assuming that the measurement noise has unit power). From (4.2.28), we then find

$$\mathbf{L} = \mathbf{R}_{h'} \oslash \mathbf{C}(\mathbf{B}\mathbf{1}\mathbf{B}^* \oslash (\mathbf{1} - \mathbf{a}\mathbf{a}^*))\mathbf{C}^* = \begin{pmatrix} 0.08947 & 0 \\ 0 & 0.02982 \end{pmatrix}, \quad (4.2.38)$$

using $\mathbf{R}_{h'} = \mathbf{R}_\gamma$ and \mathbf{a} holding the diagonal elements of \mathbf{A} . To make the model stationary, we can also set the prior state covariance matrix according to (4.2.29).

Making a realization from the model $\mathbf{x}_{t+1}^{(m)} = \mathbf{A}\mathbf{x}_t^{(m)} + \mathbf{B}\mathbf{u}_t^{(m)}$, $\mathbf{h}_t^{(m)} = \mathcal{T}\mathbf{C}\mathbf{x}_t^{(m)}$, with $\|\mathbf{u}_t^{(m)}\|^2 = \mathbf{L}$, it is found that the sample variance of each of the two subchannels in $\mathbf{h}^{(m)}$ is close to 10, as we expect. The covariance between the two should be $7.5 + 2.5e^{\pm j\omega_1\tau_1}$, as can be seen by studying the off-diagonal elements in the subchannel covariance matrix $\mathcal{T}\mathbf{R}_\gamma\mathcal{T}^*$. A numerical investigation gives the sample covariance $9.6462 \pm 1.1960j$ between the two subchannels, in close agreement with the theory.

EXAMPLE 4.2.3 Multipath channel model – modelling in the subchannel domain

The SNR in an OFDM system is assumed to be 10 dB, and the fading environment is assumed to be well described by Example 4.2.1. We want to model two subchannels ($w = 2$) in this system by modelling in the subchannel domain. By empirical study, we find the sample covariance between the two subchannels to be approximately $7.5 + 2.5e^{\pm j/2}$.

Assuming that the measurement noise has unit variance, we have the subchannel covariance matrix

$$\mathbf{R}_h = \begin{pmatrix} 10 & 7.5 + 2.5e^{-j/2} \\ 7.5 + 2.5e^{j/2} & 10 \end{pmatrix}, \quad (4.2.39)$$

and since we here model in the subchannel domain, we have from (4.2.24) that $\mathbf{R}_{h'} = \mathbf{R}_h$. Coincidentally, the covariance matrix for the subchannels $\mathbf{R}_{h'}$ used here is the same as the corresponding matrix $\mathcal{T}\mathbf{R}_\gamma\mathcal{T}^*$ in Example 4.2.2. Since we use subchannel domain modelling, the transformation matrix is now $\mathcal{T} = \mathbf{I}_2$. The process noise covariance \mathbf{L} is calculated according to (4.2.28), and we find

$$\mathbf{L} = \begin{pmatrix} 0.1193 & 0.1156 - 0.0143j \\ 0.1156 + 0.0143j & 0.1193 \end{pmatrix}. \quad (4.2.40)$$

Making a realization as in Example 4.2.2, we find the sample variance of both the subchannels to be close to 10, and their sample covariance is $9.6714 \pm 1.2070j$, again in close agreement with theory.

4.2.4 Multi-input channel model

The continuation from a single input model to a multi-input model is a simple matter of augmenting the state space with a third hierarchical block level⁵:

$$\begin{aligned}\mathbf{x}_{t+1} &= \text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_u)\mathbf{x}_t + \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_u)\mathbf{u}_t, \\ &\triangleq \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t \\ \mathbf{h}_t &= \text{diag}(\mathcal{T}_1\mathbf{C}_1, \dots, \mathcal{T}_u\mathbf{C}_u)\mathbf{x}_t \triangleq \mathbf{H}\mathbf{x}_t.\end{aligned}\tag{4.2.41}$$

Assuming independence between the u transmitting antennas, the process noise covariance matrix $\mathbf{Q} \triangleq \|\mathbf{u}_t\|^2$ and the prior $\Pi_0 \triangleq \|\mathbf{x}_0\|^2$ can be constructed accordingly:

$$\begin{aligned}\mathbf{Q} &= \text{diag}(\mathbf{L}_1, \dots, \mathbf{L}_u), \\ \Pi_0 &= \text{diag}(\Gamma_1, \dots, \Gamma_u).\end{aligned}\tag{4.2.42}$$

The subscripts indicate model matrices for u individual inputs.

Although in most cases the transformation matrices $\{\mathcal{T}_i\}_{i=1}^u$ would be the same for all inputs, the state space model is general enough to handle different transformation matrices, so that impulse response modelling may be used for some inputs and subchannel modelling for others. The model could also accommodate different numbers of subchannels for different inputs, but we are here interested in modelling u inputs over the *same* w subchannels. Note that although the model matrices are large ($\mathbf{F} \in \mathbb{C}^{kxu \times kxu}$, $\mathbf{G} \in \mathbb{C}^{kxu \times xu}$, $\mathbf{H} \in \mathbb{C}^{wu \times kxu}$), they are sparse and therefore require a limited amount of memory for storage.

Finally we model the measurements, in which the fading channel coefficients are observed in additive white noise through pilot symbols:

$$\mathbf{y}_t = \Phi_t \mathbf{h}_t + \mathbf{v}_t = \mathbf{J}_t \mathbf{x}_t + \mathbf{v}_t.\tag{4.2.43}$$

Here, $\mathbf{J}_t = \Phi_t \mathbf{H} \in \mathbb{C}^{w \times kxu}$, where $\Phi_t \in \mathbb{C}^{w \times wu}$, introduced in (4.2.3), holds potentially time-varying pilot symbols, and \mathbf{v}_t is white Gaussian noise with covariance matrix $\mathbf{R} \triangleq \|\mathbf{v}_t\|^2$. By conventionally setting $\mathbf{R} = \mathbf{I}_w$, we can interpret $\text{Tr}(\mathbf{R}_\gamma)$ for an input as that input's signal-to-noise (and interference) ratio, where $\text{Tr}(\cdot)$ is the matrix trace.

⁵For a specification of the involved matrix dimensions, see Appendix 4.H.

In a single input⁶ system, Φ_t is conveniently chosen as a diagonal matrix

$$\Phi_t = \begin{pmatrix} \phi_{1,1,t} & & \\ & \ddots & \\ & & \phi_{w,1,t} \end{pmatrix} \quad (4.2.44)$$

where the pilot symbols $\phi_{i,j,t}$ for subchannel i and input j at time t are generally complex-valued. In a multi-input system, Φ_t could either be set up to assign exclusive pilots to all the inputs (which is often intractable since it generates an excessive pilot overhead), or to assign pilots to many inputs simultaneously and at the same subchannels, in which case the pilots will overlap. We return to this matter in Chapter 6.

We now have a model

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t, \\ \mathbf{y}_t &= \mathbf{J}_t\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \quad (4.2.45)$$

where \mathbf{u}_t , \mathbf{v}_t , and \mathbf{x}_0 are zero-mean Gaussian, white, and

$$\|[\mathbf{u}_t^T, \mathbf{v}_t^T, \mathbf{x}_0^T]^T\|^2 = \text{diag}(\mathbf{Q}, \mathbf{R}, \Pi_0), \quad (4.2.46)$$

where $\mathbf{Q} > 0$, $\mathbf{R} > 0$, and $\Pi_0 > 0$, that accurately models all the kxu channel coefficients in the system. Appendix 4.H summarizes the dimensions and structures of all matrices that the model (4.2.45) depends on.

EXAMPLE 4.2.4 MIMO channel model

A terminal receives two inputs, both with fading channels according to Example 4.2.1. The inputs are both received on two subchannels. These subchannels have correlation properties as described by Examples 4.2.2 and 4.2.3 (the two examples are identical in this respect). The measured pilot symbols from the two inputs overlap in time and frequency, but we may choose the values of the transmitted pilot symbols. We here wish to model the received noisy signal as measured by the terminal.

For this purpose, we compute $\{\mathbf{A}, \mathbf{B}, \mathcal{TC}, \mathbf{L}, \Gamma_0\}$ as in Example 4.2.2 or 4.2.3 and set the model matrices

$$\begin{aligned} \mathbf{F} &= \text{diag}(\mathbf{A}, \mathbf{A}), & \mathbf{G} &= \text{diag}(\mathbf{B}, \mathbf{B}), & \mathbf{H} &= \text{diag}(\mathcal{TC}, \mathcal{TC}), \\ \mathbf{Q} &= \text{diag}(\mathbf{L}, \mathbf{L}), & \mathbf{R} &= \mathbf{I}_2, \end{aligned}$$

⁶Note that the notion *input* refers to a data stream from one transmitting antenna element, received over w subchannels, and that the inputs are here represented by the matrix Φ_t . This should be distinguished from the elements in the process noise \mathbf{u}_t , which are sometimes referred to as inputs in state-space contexts.

We also set the prior state covariance matrix $\Pi_0 = \text{diag}(\Gamma_0, \Gamma_0)$ to force model stationarity.

The total model thus has process noise of dimension $xu = 4$, $kxu = 8$ states, $xu = 4$ channel coefficients, and a measurement signal of dimension $w = 2$. It remains to set the pilot symbol matrix Φ_t , which may be time varying. The pilot patterns used by the two inputs should be mutually orthogonal. In a flat fading environment, choosing orthogonal pilot patterns is adequate to separate the channels for the two inputs completely. Here, since the channels are frequency selective, it is a good idea to also make the patterns orthogonal over time for each of the two subchannels. For example, we could let Φ_t alternate between the two values

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \quad (4.2.47)$$

With this choice of Φ_t , the regressor matrix $\mathbf{J}_t = \Phi_t \mathbf{H}$ has period 2. In Chapter 6, we will see how time-varying pilots can greatly improve channel estimation and prediction performance when $u > w$.

4.2.5 A few remarks

Although the model (4.2.45) is quite general, we have made a few restricting assumptions that one may want to relax.

Rice components

We have throughout this chapter modelled the fading channel as a channel having zero-mean. The channel impulse response may however have a static non-zero component, especially in line-of-sight (LOS) scenarios when a phase locked loop locks to the LOS component. The static component, say $\mathbf{g}^{(av.)}$, may then be estimated and the corresponding frequency response average $\mathbf{h}^{(av.)} = N^{1/2} \mathcal{F}_w \mathbf{g}^{(av.)}$ calculated. The total channel is then $\mathbf{h}_t^{(tot.)} = \mathbf{h}^{(av.)} + \mathbf{h}_t$, and we model and track the zero-mean component \mathbf{h}_t .

Antenna coupling

In an $n_R \times n_T$ MIMO system, we assumed in the beginning of Section 4.2 that the n_R receiving antennas are uncorrelated. Furthermore, in Section 4.2.4, we also assumed that the n_T transmitting antennas are uncorrelated. However, due to insufficient antenna separation, electrical coupling, and/or local

geometry, antenna correlations may be present at both the transmitting and the receiving sides of the system.

Theorem 4.2.1, which we used to set the process noise covariance matrix so as to produce a correct correlation among the modelled channel coefficients, can be trivially extended to handle correlations among antennas. The theorem was applied at the multipath modelling level (Section 4.2.3), but can easily be “moved up” to the multi-input modelling level (Section 4.2.4). Replace the system model (4.2.25) in the theorem with $\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t$, $\mathbf{h}' = \text{diag}(\mathbf{C}_1, \dots, \mathbf{C}_u)\mathbf{x}_t$ and assume that the $n_R n_T$ -by- $n_R n_T$ channel coefficient correlation matrix $\mathbf{R}_{h'}$ = $\|\mathbf{h}'\|^2$ has been estimated. Solve for $\mathbf{Q} = \|\mathbf{u}_t\|^2$ instead of \mathbf{L} and $\Pi_0 = \|\mathbf{x}_0\|^2$ instead of Γ_0 . The matrix structures comply with the theorem also for this larger state space model.

4.3 Channel inference

4.3.1 The Kalman filter recursions

From the total multi-input system model (4.2.41)–(4.2.43), characterized by matrices $\{\mathbf{F}, \mathbf{G}, \mathbf{J}_t, \mathbf{Q}, \mathbf{R}, \Pi_0\}$, we may now infer the unknown fading channel coefficients by way of the Kalman filter (KF) recursions. See Table 4.1.

In Table 4.1, $\mathbf{R}_{e,t}$ is the covariance matrix for the innovations, and $\mathbf{K}_{f,t}$ is the *Kalman filter gain* which expresses the tradeoff between taking new measurements into account and relying on old knowledge. The matrix $\mathbf{W}_t = \mathbf{J}_t \mathbf{P}_t$ is introduced as an intermediate result to make the computations efficient.

The KF does not only calculate point estimates of the state vector \mathbf{x}_t . Rather, it produces the full pdf

$$p(\mathbf{x}_t | \mathbf{y}_0, \dots, \mathbf{y}_t, I) = \mathcal{CN}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \quad (4.3.9)$$

for the state vector, given as recent measurement as possible. The mean value of this pdf, $\hat{\mathbf{x}}_{t|t}$, is called the *filtered estimate* of \mathbf{x}_t and is the optimal estimate given measurements up to time t . As a by-product, the KF also produces the *one-step prediction* pdf

$$p(\mathbf{x}_t | \mathbf{y}_0, \dots, \mathbf{y}_{t-1}, I) = \mathcal{CN}(\mathbf{x}_t; \hat{\mathbf{x}}_t, \mathbf{P}_t), \quad (4.3.10)$$

which gives the optimal estimate given measurements up to time t . Note that we use, for brevity, the short notation

$$\hat{\mathbf{x}}_t \triangleq \hat{\mathbf{x}}_{t|t-1} \quad \text{and} \quad \mathbf{P}_t \triangleq \mathbf{P}_{t|t-1}. \quad (4.3.11)$$

The covariance matrix for the state vector \mathbf{x}_t given measurements up to and including \mathbf{y}_t , and the covariance matrix for the state vector \mathbf{x}_t given

Table 4.1: The Kalman recursions with numerical complexity for impulse response modelling (imp. mod. cplx.) and subchannel (subc. mod. cplx.) modelling, respectively, in number of arithmetic complex operations (one operation=one multiplication and one addition). The total number of states is $n = kxu$, where k is the model order per channel coefficient, x is either w (filter width) or m (number of taps in the impulse response), and u is the number of inputs.

initialize:

$$\hat{\mathbf{x}}_0 = \mathbf{0}, \mathbf{P}_0 = \mathbf{\Pi}_0$$

iterate:

operation	imp. mod. cplx.	subc. mod. cplx.	
$\mathbf{J}_t = \Phi_t \mathbf{H}$	nw	n	(4.3.1)
$\mathbf{W}_t = \mathbf{J}_t \mathbf{P}_t$	$n^2 w$	n^2	(4.3.2)
$\mathbf{R}_{e,t} = \mathbf{W}_t \mathbf{J}_t^* + \mathbf{R}$	$nw^2/2$	$nw/2$	(4.3.3)
$\mathbf{K}_{f,t} = \mathbf{W}_t^* \mathbf{R}_{e,t}^{-1}$	$nw^2 + w^3/6$	$nw^2 + w^3/6$	(4.3.4)
$\hat{\mathbf{x}}_{t t} = \hat{\mathbf{x}}_t + \mathbf{K}_{f,t}(\mathbf{y}_t - \mathbf{J}_t \hat{\mathbf{x}}_t)$	$nw + nw$	$n + nw$	(4.3.5)
$\mathbf{P}_{t t} = \mathbf{P}_t - \mathbf{K}_{f,t} \mathbf{W}_t$	$n^2 w/2$	$n^2 w/2$	(4.3.6)
$\hat{\mathbf{x}}_{t+1} = \mathbf{F} \hat{\mathbf{x}}_{t t}$	n	n	(4.3.7)
$\mathbf{P}_{t+1} = \mathbf{F} \mathbf{P}_{t t} \mathbf{F}^* + \mathbf{G} \mathbf{Q} \mathbf{G}^*$	$n^2/2$	$n^2/2$	(4.3.8)

measurements up to and including \mathbf{y}_{t+1} , are denoted $\mathbf{P}_{t|t}$ and \mathbf{P}_t , respectively. Note that these matrices also constitute the covariance matrices for the *state estimation errors*:

$$p(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t} | \mathbf{y}_0, \dots, \mathbf{y}_t, I) = \mathcal{CN}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}; \mathbf{0}, \mathbf{P}_{t|t}), \quad (4.3.12)$$

$$p(\mathbf{x}_t - \hat{\mathbf{x}}_t | \mathbf{y}_0, \dots, \mathbf{y}_t, I) = \mathcal{CN}(\mathbf{x}_t - \hat{\mathbf{x}}_t; \mathbf{0}, \mathbf{P}_t). \quad (4.3.13)$$

From the state estimate vector we can compute the channel coefficients estimate by $\hat{\mathbf{h}}_{t|t} = \mathbf{H} \hat{\mathbf{x}}_{t|t}$, which may then be used for channel equalization or as input to a soft channel decoder. When the KF is employed as an analyzing tool for assessing different aspects of a system without carrying out any simulations, the state estimation update equations (4.3.5) and (4.3.7) should be omitted entirely.

An important property of the KF recursions is that the error covariance matrices $\mathbf{P}_{t|t}$ and \mathbf{P}_t depend only on the model matrices and not on the measurements $\{\mathbf{y}_t\}$. This means that although we cannot calculate any estimates in the absence of measurements, we can calculate the *distribution* of

estimates without measurements being present. And more importantly, we may determine the distribution of *errors*, i.e. the differences between estimated and true channel coefficients.

The distribution of errors for the channel coefficients in \mathbf{h}_t , given measurements up to time t_0 , is zero mean circular symmetric complex Gaussian with variance, or mean square error (MSE), given by the diagonal elements of $\mathbf{H}\mathbf{P}_{t|t_0}\mathbf{H}^*$. As a measure of quality for channel estimations or predictions, we will use the *normalized mean square errors* (NMSE):

$$\text{vector of NMSE values for } \mathbf{h}_t = \text{diag}(\mathbf{H}\mathbf{P}_{t|t_0}\mathbf{H}^*) \oslash \text{diag}(\mathbf{R}_h), \quad (4.3.14)$$

given $\mathbf{y}_0, \dots, \mathbf{y}_{t_0}$

where \oslash means element-wise division, $\text{diag}(\mathbf{M})$ is a vector holding the diagonal elements of the matrix \mathbf{M} , and \mathbf{R}_h is the subchannel covariance matrix.

In the above, we have used the general notation $\mathbf{P}_{t|t_0}$ in place of $\mathbf{P}_{t|t}$ to emphasize that (4.3.14) holds whatever the amount of measurement available, since the KF recursions can be extended to calculate any error covariance $\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t_0}\|^2 \triangleq \mathbf{P}_{t|t_0}$. If $t > t_0$ we have a *prediction* problem. Prediction error covariance matrices can be calculated iteratively through

$$\mathbf{P}_{t+c+1|t} = \mathbf{F}\mathbf{P}_{t+c|t}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*. \quad (4.3.15)$$

It is straightforward to show that if $|\mathbf{F}| < 1$, then $\mathbf{P}_{t+c|t} \rightarrow \Pi_0$ as $c \rightarrow \infty$, so that the KF “falls back” to a state of maximum ignorance when measurements become uninformative. When this happens, we see from the relation $\mathbf{R}_h = \mathbf{H}\Pi_0\mathbf{H}^*$ that the NMSE is 0 dB.

When $t < t_0$ we have a smoothing problem. Kalman smoothing is not as straightforward as prediction and there are many variants of smoothing formulas. In Chapter 3 we briefly presented the so-called Bryson-Frazier formulas for Kalman smoothing. For a comprehensive presentation of Kalman smoothing, see [49], [56].

4.3.2 The periodic filter

We have assumed in (4.2.41)–(4.2.42) that the model matrices $\{\mathbf{F}, \mathbf{G}, \mathbf{H}, \mathbf{Q}, \mathbf{R}\}$ are time-invariant. The matrix \mathbf{F} is, by construction, diagonal, with the poles of the models for the respective channel coefficients as elements. These poles are chosen to be stable, meaning that the eigenvalues of \mathbf{F} are strictly inside the unit circle. When also the pilot matrix Φ_t is static, the KF will settle to a stationary state after a few iterations in the recursions loop, which means that all covariance matrices in the KF recursions become constant, see Theorem 4.G.1. It is then convenient to define the stationary covariance matrices

$$\begin{aligned} \mathbf{P}_{t|t} &\rightarrow \bar{\mathbf{P}}_f, & t &\rightarrow \infty \\ \mathbf{P}_t &\rightarrow \bar{\mathbf{P}}_p, & t &\rightarrow \infty \end{aligned} \quad (4.3.16)$$

However, in the system model (4.2.45) we have allowed Φ_t to be time varying. A special case occurs when Φ_t is periodic with period b . The filter will then settle to a periodic state, see Appendix 4.G. Analogous to (4.3.16), we may then define two sets of matrices $\{\bar{\mathbf{P}}_{f,0}, \dots, \bar{\mathbf{P}}_{f,b-1}\}$ and $\{\bar{\mathbf{P}}_{p,0}, \dots, \bar{\mathbf{P}}_{p,b-1}\}$, each of size b , defined by

$$\bar{\mathbf{P}}_{f,j} \triangleq \mathbf{P}_{ib+j|ib+j} \quad \text{and} \quad \bar{\mathbf{P}}_{p,j} \triangleq \mathbf{P}_{ib+j}, \quad i \geq i_0 \quad (4.3.17)$$

where $j \in [0, b)$, and i_0 is assumed to be large enough that the filter has reached a periodic state.

4.3.3 Modelling channel estimates

When analysing any particular communications system, we are interested in seeing how estimation and prediction errors are distributed, i.e. we want to calculate their *frequency* distribution. While a probability density distribution is not generally the same as a frequency distribution, this equality holds when the filter is stationary or periodic [43].

EXAMPLE 4.3.1 Filtered estimates model

The UE in Example 4.2.4 uses an optimal filter based on a perfect channel model to infer the fading MIMO channel. Filtered estimates are used as channel estimates. We wish to statistically describe and produce such filtered estimates when the filter operates in a stationary state.

Statistically correct estimates can be produced by calculating a synthetic channel and then running a KF on it, but it is more straight-forward to use a model for the estimates themselves. First, note that \mathbf{J}_t from Example 4.2.4 has period 2. We find the 2-periodic stabilizing solution $\{\bar{\mathbf{P}}_{p,0}, \bar{\mathbf{P}}_{p,1}\}$ by iterating the *discrete-time Riccati difference equation*

$$\mathbf{P}_{t+1} = \mathbf{F}\mathbf{P}_t\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* - \mathbf{F}\mathbf{P}_t\mathbf{J}_t^*(\mathbf{J}_t\mathbf{P}_t\mathbf{J}_t^* + \mathbf{R})^{-1}\mathbf{J}_t\mathbf{P}_t\mathbf{F}^* \quad (4.3.18)$$

a few times. Defining $\bar{\mathbf{R}}_{e,0} \triangleq \mathbf{J}_0\bar{\mathbf{P}}_{p,0}\mathbf{J}_0^* + \mathbf{R}$, $\bar{\mathbf{R}}_{e,1} \triangleq \mathbf{J}_1\bar{\mathbf{P}}_{p,1}\mathbf{J}_1^* + \mathbf{R}$, and $\bar{\mathbf{K}}_{p,0} \triangleq \mathbf{F}\bar{\mathbf{P}}_{p,0}\mathbf{J}_0^*\bar{\mathbf{R}}_{e,0}^{-1}$, $\bar{\mathbf{K}}_{p,1} \triangleq \mathbf{F}\bar{\mathbf{P}}_{p,1}\mathbf{J}_1^*\bar{\mathbf{R}}_{e,1}^{-1}$, we can set up the model for the one-step predictions (see (3.4.42)),

$$\hat{\mathbf{x}}_{t+1} = \mathbf{F}\hat{\mathbf{x}}_t + \bar{\mathbf{K}}_{p,t \bmod 2}\mathbf{e}_t \quad (4.3.19)$$

$$\mathbf{y}_t = \mathbf{J}_t\hat{\mathbf{x}}_t + \mathbf{e}_t \quad (4.3.20)$$

where $t \bmod 2$ is the remainder of division of t by 2, and \mathbf{e}_t is white with time-varying covariance matrix $\bar{\mathbf{R}}_{e,t}$. Filtered channel estimates are then calculated from the one-step state predictions:

$$\hat{\mathbf{h}}_{t|t} = \mathbf{H}(\hat{\mathbf{x}}_t + \mathbf{F}^{-1}\bar{\mathbf{K}}_{p,t}\mathbf{e}_t). \quad (4.3.21)$$

The estimation errors for the filtered estimates are white in time with covariance matrix $\mathbf{H}\bar{\mathbf{P}}_{f,t}\mathbf{H}^*$, where $\bar{\mathbf{P}}_{f,t}$ is given by (4.3.6):

$$\bar{\mathbf{P}}_{f,t} = (\mathbf{I} - \bar{\mathbf{P}}_{p,t}\mathbf{J}_t^*\bar{\mathbf{R}}_{e,t}^{-1}\mathbf{J}_t)\bar{\mathbf{P}}_{p,t}. \quad (4.3.22)$$

4.3.4 Optimal channel gain prediction

Consider a single channel coefficient h_t in the channel coefficient vector \mathbf{h}_t . The squared magnitude channel coefficient, $|h_t|^2$, together with the noise variance, determines the signal-to-noise ratio and has thus a direct bearing on system performance. It is therefore of interest to predict the squared magnitude of a channel coefficient rather than its complex value. From basic statistical theory (see Appendix 4.A and 4.B) we know that if a parameter h has pdf $\mathcal{CN}(h; \hat{h}, \sigma^2)$, then its squared magnitude $z \triangleq |h|^2$ has a non-central χ^2 pdf with two degrees of freedom:

$$p(z|\hat{h}, \sigma^2) = \frac{1}{\sigma^2} \exp\left(-\frac{z + |\hat{h}|^2}{\sigma^2}\right) I_0\left(\frac{2\sqrt{z}|\hat{h}|}{\sigma^2}\right), \quad z \geq 0, \quad (4.3.23)$$

with mean value $|\hat{h}|^2 + \sigma^2$ and variance $\sigma^2(2|\hat{h}|^2 + \sigma^2)$. Here, I_0 is the zeroth order modified Bessel function of the first kind. The optimal point estimate⁷ of a coefficient's squared magnitude $|h_c|^2$ is therefore $|\hat{h}_c|^2 + \sigma_c^2$ as suggested by Ekman *et al* in [57], and not – as one perhaps would have thought – simply $|\hat{h}|^2$, as suggested in e.g. [58]. (It should be noted that a different but equivalent expression for the χ^2 -distribution (4.3.23) is used in [57] and [58]. See Appendix 4.C.)

On vector form we can therefore compactly write the channel power gain estimates as follows:

$$\begin{array}{l} \text{vector of optimal squared} \\ \text{magnitude estimates} \end{array} = \text{diagonal elements of } (\hat{\mathbf{h}}_{t|t_0}\hat{\mathbf{h}}_{t|t_0}^* + \mathbf{H}\mathbf{P}_{t|t_0}\mathbf{H}^*), \quad (4.3.24)$$

where we have assumed that measurements up to and including time t_0 are available.

⁷With $z = |h|^2$, the optimal estimate \hat{z} of z is meant in the sense that a quadratic loss function is used, so that the expected value of the error $(z - \hat{z})^2$ is minimized, *cf.* Section 3.2.4.

4.3.5 A special case: The GCG algorithm

We will generally take the regressor matrix \mathbf{J}_t in (4.2.45) to be periodic, since the filter will then converge to a periodic filter. But in fact, under special circumstances a time-invariant model yielding a stationary filter can be constructed even when \mathbf{J}_t is not periodic. Take the single-input case (4.2.43) with $u = 1$ so that the pilot matrix Φ_t (4.2.44) is invertible. A “fictitious” measurement signal

$$\bar{\mathbf{y}}_t = \Phi_t^{-1} \mathbf{y}_t = \mathbf{H} \mathbf{x}_t + \Phi_t^{-1} \mathbf{v}_t \quad (4.3.25)$$

can then be constructed from the measurement vector (4.2.43). The fictitious noise $\Phi_t^{-1} \mathbf{v}_t$ has covariance matrix $\Phi_t^{-1} \mathbf{R} \Phi_t^{-*}$ which is generally time varying. Since Φ_t is selected to be diagonal when $u = 1$, the same covariance matrix can be written $\Psi_t \odot \mathbf{R}$, where \odot is the elementwise multiplication operator, and the elements of Ψ_t are

$$\Psi_t[i, j] = \phi_{i,t}^{-1} (\phi_{j,t}^*)^{-1}, \quad (4.3.26)$$

where $\phi_{i,t}$ is the pilot for subchannel i and time t . By letting $\phi_{i,t} = \exp(j\omega t)$, so that pilots have unit modulus and the same pilot symbol is used on all subchannels, the model becomes time static for any value of the noise covariance matrix \mathbf{R} . If we assume that the elements in \mathbf{v}_t are mutually uncorrelated with variance σ_v^2 , we have $\mathbf{R} = \sigma_v^2 \mathbf{I}$. It then suffices to choose $\Phi_t^{-1} = \Phi_t^*$, i.e. unit modulus pilots that are allowed to differ between subchannels, to make the model time-invariant.

For legacy reasons, we have sometimes referred to filters using the fictitious measurement signal (4.3.25) as *General Constant Gain* (GCG) algorithms; see e.g. [59],[60].

4.4 Numerical complexity

To evaluate the feasibility of using optimal filters as proposed in this thesis, we need to assess the complexity of the KF. The complexity is mainly determined by the number of states n , which is the product of the model order k , the number of modelled channel coefficients x (i.e. the impulse response length m or the filter width w), and the number of inputs u .

The KF recursions update the filtered state estimate vector $\hat{\mathbf{x}}_{t|t}$, the state prediction vector $\hat{\mathbf{x}}_{t+L|t}$, the filtered state estimation error covariance matrix $\mathbf{P}_{t|t}$, and the state prediction error covariance matrix $\mathbf{P}_{t+L|t}$. The numerical complexity of a general KF is on the order of n^3 [61], but due to the block diagonal structures used for fading channel modelling in this thesis and the

fact that \mathbf{F} is diagonal, the complexity is reduced considerably as compared to the general KF.

In Table 4.1, we list the complexity, counted in complex arithmetic operations, for each operation in the KF recursion loop when either subchannel modelling or impulse response modelling is being used. One complex arithmetic operation constitutes one complex multiplication and one complex addition. Note that inversion of a positive definite w -by- w matrix can be done by Cholesky factorization and requires about $w^3/6$ operations [62].

When L -step prediction is used, the time update equations (4.3.7) and (4.3.8) furthermore have to be calculated $L - 1$ extra times in each loop in the recursions (the first prediction step comes “free of charge” from the Kalman recursions), imposing a total extra cost of $(L-1)(n^2/2+n)$ arithmetic operations on the numerical complexity of the filter.

Using the block-diagonal model matrices suggested in this work leads to an alleviation of the numerical complexity as compared to the general KF. This alleviation is mainly found in the time update equations. Normally, (4.3.7) would require n^2 operations, but require only n operations here, since \mathbf{F} is diagonal. The state prediction error covariance update (4.3.8) would take $3n^3/2 + nw^2 + n^2w/2$ operations (assuming $\mathbf{Q} \in \mathbb{C}^{w \times w}$) if general and full model matrices were used, but here it can ideally be done in $n^2/2$ operations. We have then assumed that the matrix $\mathbf{G}\mathbf{Q}\mathbf{G}^*$ has been precomputed. For the rest of the equations in the KF cycle, the general KF has the same complexity as when impulse response modelling is used. When subchannel modelling is employed, complexity relaxation is achieved also for some of these equations. In Appendix 4.F, we investigate some alternative KF formulations but find that these do not lead to further computational gains for this particular model structure.

For L -step prediction using measurements \mathbf{y}_t of dimension w and $n = uwk$ states, the total number of complex arithmetic operations per update in the KF sums to

$$n^2 \left(\frac{w}{2} + \frac{3}{2} + \frac{L-1}{2} \right) + n \left(w^2 + \frac{3w}{2} + 3 + (L-1) \right) + \frac{w^3}{6} \quad (4.4.1)$$

complex arithmetic operations when subchannel modelling is used. The complexity is somewhat higher for impulse response modelling, see Table 4.1.

It should be noted that if the time varying matrix \mathbf{J}_t is periodic with period b , then the operation (4.3.1) in Table 4.1 only have to be carried out b times and the stored results can then be reused periodically. However, it is not evident that this saving in computation makes up for the additional memory access that would then be necessary.

If the total number of subchannels to estimate and predict is w_0 , and each filter covers a band of w subchannels, then w_0/w (rounded upwards)

KFs have to be run in parallel. However, if all of these bands use the same state-space model, which will be the case if the same set of inputs, all using the same set of pilots, are received in all bands, then the covariance matrices in the KF:s are identical and need only be updated for one band.

Although a large part of the computational burden is alleviated by the special matrix structures, the complexity is still quite high. However, as will be seen in Chapter 8, in most cases the filter will quickly settle to a stationary state, which makes it unnecessary to update the large covariance matrices. Therefore, after a short initial phase of a few tens of iterations, the filter has converged and requires only $n(w + 2)$ or $n(2w + 1)$ complex operations per update for the estimation. The prediction $\hat{\mathbf{x}}_{t+L|t} = \mathbf{F}^{L-1}\hat{\mathbf{x}}_{t+1}$ adds a mere n operations per update if the value of \mathbf{F}^{L-1} is kept from one iteration to the other. The filter needs to be re-initiated every time the model parameters are updated, but this happens on a slow time scale (on the order of seconds).

4.5 Summary and discussion

We have in Section 4.2.1 discussed AR models that are parameterized by the maximal Doppler frequency f_D , such as the models adjusted to a Jakes fading Doppler spectrum or to a flat Doppler spectrum. Such models can be held fixed as long as the (estimated) maximal Doppler frequency remains almost constant.

For autoregressive models that are adjusted to the detailed fading statistics, the duration over which such models can be held constant will depend on the duration over which the local fading environment remains almost constant, in terms of strengths and angles of main propagation/ reflection/ scattering paths. This time is on the order of at least hundreds of milliseconds for most utilized combinations of UE velocities and distances to the nearest significant scatterers. The stationarity of the noise and interference environment is of importance. If the noise vector contains bursty interference components, this will require frequent re-adjustments of the Kalman estimators, to retain the best performance.

The remainder of the chapter discussed how to construct state space representations of MIMO-OFDM time-varying channels. It was described how these models can be used in conjunction with a KF to produce probability density functions for channel estimates and predictions. The framework outlined in this chapter will be used in subsequent chapters to study channel estimation, prediction, and link adaptation in various settings.

4.A Some useful distributions

4.A.1 Definitions

The Gaussian distribution

The Gaussian distribution with mean value μ and variance σ^2 is defined

$$\mathcal{N}(x; \mu, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (4.A.1)$$

The circular symmetric complex Gaussian distribution

The circular symmetric complex Gaussian distribution, or sometimes, for short, just the Gaussian distribution, with mean value μ and variance σ^2 is defined

$$\mathcal{CN}(h; \mu, \sigma^2) \triangleq \frac{1}{\pi\sigma^2} \exp\left(-\frac{|h - \mu|^2}{\sigma^2}\right). \quad (4.A.2)$$

The notion ‘circular symmetric complex’ means that the real and imaginary parts are independent and that both have the same variance $\sigma^2/2$.

The non-central χ^2 distribution

Let $z = |h|^2$. If $p(h|\mu, \sigma^2) = \mathcal{CN}(h; \mu, \sigma^2)$, then $p(z|\mu, \sigma^2)$ is a non-central χ^2 -distribution with two degrees of freedom:

$$p(z|\mu, \sigma^2) = \chi^2(z; |\mu|^2, \sigma^2) \triangleq \frac{1}{\sigma^2} \exp\left(-\frac{z + |\mu|^2}{\sigma^2}\right) I_0\left(\frac{2\sqrt{z}|\mu|}{\sigma^2}\right), \quad (4.A.3)$$

where $I_0(\cdot)$ denotes the zeroth order modified Bessel function of the first kind. (For a proof, see Appendix 4.B.) It should be noted that many references implicitly assume $\sigma^2 = 1$, simplifying the above expression somewhat. In this thesis, we will however require the variance to be variable.

When $\mu = 0$, the gain z has an exponential distribution:

$$p(z|\sigma^2) = \chi^2(z; 0, \sigma^2) = \frac{1}{\sigma^2} \exp\left(-\frac{z}{\sigma^2}\right). \quad (4.A.4)$$

The Rice distribution

Let $a = |h|$. If $p(h|\mu, \sigma^2) = \mathcal{CN}(h; \mu, \sigma^2)$, then $p(a|\mu, \sigma^2)$ is a Rice distribution:

$$p(a|\mu, \sigma^2) = \frac{2a}{\sigma^2} \exp\left(-\frac{a^2 + |\mu|^2}{\sigma^2}\right) I_0\left(\frac{2a|\mu|}{\sigma^2}\right). \quad (4.A.5)$$

(This follows from a simple change of variables of (4.A.3).)

For the special case $\mu = 0$, the magnitude a has a Rayleigh distribution:

$$p(a|\sigma^2) = \frac{2a}{\sigma^2} \exp\left(-\frac{a^2}{\sigma^2}\right). \quad (4.A.6)$$

4.A.2 Change of variables

The noncentral χ^2 distribution

For the non-central χ^2 -distribution and a parameter α independent of z , it holds that

$$\chi^2(z; |\hat{h}|^2, \sigma^2) dz = \chi^2(\alpha z; \alpha|\hat{h}|^2, \alpha\sigma^2) d(\alpha z), \quad (4.A.7)$$

since

$$\begin{aligned} \frac{1}{\sigma^2} \exp\left(-\frac{z + |\hat{h}|^2}{\sigma^2}\right) I_0\left(\frac{2|\hat{h}|\sqrt{z}}{\sigma^2}\right) dz = \\ \frac{\alpha}{\alpha\sigma^2} \exp\left(-\frac{\alpha z + \alpha|\hat{h}|^2}{\alpha\sigma^2}\right) I_0\left(\frac{2\alpha|\hat{h}|\sqrt{\alpha z}}{\alpha\sigma^2}\right) d(\alpha z). \end{aligned} \quad (4.A.8)$$

4.B Posterior distribution for the channel power

Denoting the real and imaginary part of a complex number h (e.g. a channel coefficient) by h_r and h_i , respectively, we seek the pdf for $z = h_r^2 + h_i^2$ (the channel power), conditioned on that h has a circular symmetric complex Gaussian distribution with mean value $\mu = \mu_r + j\mu_i$ and variance σ^2 . This is done by calculating $P(t < z|\mu, \sigma^2, I)$ and then taking the derivative with respect to z . The prior assumption gives that h_r and h_i are uncorrelated and that both has variance $\sigma^2/2$. Let $f(z) = p(z|\mu, \sigma^2, I)$. Then

$$\begin{aligned} \int_{t < z} f(t) dt &= \int_{h_r^2 + h_i^2 < z} \mathcal{N}(h_r; \mu_r, \sigma^2/2) \mathcal{N}(h_i; \mu_i, \sigma^2/2) dh_r dh_i \\ &= \int_{h_r^2 + h_i^2 < z} \frac{1}{\pi\sigma^2} e^{-(h_r - \mu_r)^2 - (h_i - \mu_i)^2} / \sigma^2 dh_r dh_i. \end{aligned}$$

This can be rewritten by use of the law of cosines, which says that the sides in a triangle are related as $a^2 = b^2 + c^2 - 2bc \cos \theta$, where θ is the angle between sides b and c . Changing to polar coordinates, we now have

$$\int_{t < z} f(t) dt = \int_0^{\sqrt{z}} \int_0^{2\pi} \frac{1}{\pi\sigma^2} e^{(-|\mu|^2 - r^2 + 2|\mu|r \cos \theta) / \sigma^2} d\theta r dr.$$

Using $I_0(z) = \frac{1}{\pi} \int_0^\pi \exp(z \cos \theta) d\theta$, where $I_0(z)$ is the modified Bessel function of the first kind, the above evaluates to

$$\int_{t < z} f(t) dt = \int_0^{\sqrt{z}} 2 \frac{e^{-(|\mu|^2 + r^2)/\sigma^2}}{\sigma^2} I_0(2|\mu|r/\sigma^2) r dr.$$

Finally, taking the derivative with respect to z and noting that $\frac{d}{dz} \int^{\sqrt{z}} f(t) dt = f(\sqrt{z})/2\sqrt{z}$, we have

$$p(z|\mu\sigma I) = e^{-(|\mu|^2 + z)/\sigma^2} I_0(2|\mu|\sqrt{z}/\sigma^2)/\sigma^2.$$

This is the non-central χ^2 -distribution with mean value $|\mu|^2 + \sigma^2$ and variance $\sigma^2(2|\mu|^2 + \sigma^2)$. Note that μ and σ^2 are the mean value and variance, respectively, of the circular symmetric complex Gaussian distribution $p(h_r + jh_i|\mu, \sigma^2, I)$.

4.C Alternative formulations of the non-central χ^2 -distribution

In pioneering works in the field of channel prediction ([4],[63],[64]), a pdf

$$f_\gamma(\gamma|\hat{\gamma}) = \frac{U(\gamma)U(\hat{\gamma} - \bar{\gamma}\sigma^2/\sigma_h^2)}{\bar{\gamma}\sigma^2/\sigma_h^2} \exp\left(-\frac{\gamma + \hat{\gamma} - \bar{\gamma}\sigma^2/\sigma_h^2}{\bar{\gamma}\sigma^2/\sigma_h^2}\right) \times I_0\left(\frac{2}{\bar{\gamma}\sigma^2/\sigma_h^2} \sqrt{\gamma(\hat{\gamma} - \bar{\gamma}\sigma^2/\sigma_h^2)}\right), \quad (4.C.1)$$

is used to describe an agent's state of knowledge about the future SNR γ when the channel is subject to Rayleigh fading. Here we will show that this pdf in fact can be identified as a noncentral χ^2 -distribution. In (4.C.1), $\hat{\gamma}$ is the predicted SNR, $\bar{\gamma}$ is the mean SNR, σ^2 is the total error variance for the predicted complex channel gain, σ_h^2 is the total channel variance, and $U(\cdot)$ is the Heaviside step function. In terms of the channel coefficient h , the predicted coefficient \hat{h} , the transmitted signal power S , and the noise power σ_v^2 , these quantities can be written

$$\gamma = \frac{S|h|^2}{\sigma_v^2}, \quad \hat{\gamma} = \frac{S(|\hat{h}|^2 + \sigma^2)}{\sigma_v^2}, \quad \bar{\gamma} = \frac{S\sigma_h^2}{\sigma_v^2}. \quad (4.C.2)$$

Note that $\hat{\gamma} - \bar{\gamma}\sigma^2/\sigma_h^2 = S|\hat{h}|^2/\sigma_v^2$. With $c = S/\sigma_v^2$ we can express (4.C.1) as

$$f_\gamma(\gamma|\hat{\gamma}) = \frac{U(c|h|^2)U(c|\hat{h}|^2)}{c\sigma^2} \exp\left(-\frac{c|h|^2 + c|\hat{h}|^2}{c\sigma^2}\right) I_0\left(\frac{2\sqrt{c|h|^2 c|\hat{h}|^2}}{c\sigma^2}\right), \quad (4.C.3)$$

which now clearly takes the form of a noncentral χ^2 -distribution:

$$f_\gamma(\gamma|\hat{\gamma}) = \chi^2(c|h|^2; c|\hat{h}|^2, c\sigma^2). \quad (4.C.4)$$

A change of variables (see Appendix 4.A.2) from $c|h|^2$ to $|h|^2$ gives

$$\chi^2(c|h|^2; c|\hat{h}|^2, c\sigma^2)d(c|h|^2) = \chi^2(|h|^2; |\hat{h}|^2, \sigma^2)d(|h|^2). \quad (4.C.5)$$

The squared channel coefficient $z = |h|^2$ hence has pdf

$$p(z|\hat{h}, \sigma^2) = \frac{1}{\sigma^2} \exp\left(-\frac{z + |\hat{h}|^2}{\sigma^2}\right) I_0\left(\frac{2|\hat{h}|\sqrt{z}}{\sigma^2}\right), \quad z > 0, \quad (4.C.6)$$

which is the form used in this thesis.

4.D Iterative channel estimation

In PSAM, it is assumed that pilot symbols are fully known to the receiver. However, a part of the training symbols may be allowed to carry a small amount of feedback information. By letting these “control” symbols belong to a small constellation of M symbols, each control symbol can carry $\log_2 M$ bits, while retaining a function as training symbol. This works as follows. Primary channel estimates based on true pilots are first produced for the control symbol sub-locations. The control symbols are then detected based on the primary estimates. The detected symbols are fed back into the channel estimator/predictor in a secondary phase under the pretention that they are true, thus functioning as regular pilot symbols. The information carried by control symbols can be used for feedback information such as CSI feedback or modulation format recommendations.

The procedure described above was proposed in the Wireless IP project [65] and then also in work on the WINNER I downlink [66]. It is a special case of the concept of *iterative channel estimation* (ICE). Early iterative channel estimators used uncoded symbols [67]. Later, channel codes have been used to minimize the risk for erroneous decisions. The most powerful iterative channel estimation technique known to date is when turbo codes are used as channel codes. This technique is known as *turbo equalization* [68]. To find the most probable channel coefficients from coded noisy symbols, iterative expectation maximization (EM) algorithms have proved to be efficient, see e.g. [69] and simplified algorithms in [70].

Most often, the control symbols are correctly detected and will then operate as any other pilot. Occasionally though, a control symbol is interpreted as the wrong symbol which introduces an unexpected error. We now examine

the impact of such errors. Let the measurement of a control symbol in the secondary phase be

$$y = (\hat{s} + \tilde{s})h + v, \quad (4.D.1)$$

where $s = \hat{s} + \tilde{s}$ is the true control symbol and \hat{s} is the (possibly wrongly) detected control symbol. Here, h is the unknown channel coefficient that we wish to estimate, v is white additive noise whose variance is assumed known, and $\hat{s}h$ has variance E_s , where E_s is the mean received symbol energy. The term $\tilde{s}h$ here acts as an extra noise term and we are therefore interested in knowing its variance.

If the SNR is moderately high, we may assume that when an error occurs, then $\tilde{s}h = d_{min}$, where d_{min} is the Euclidean distance between the true received symbols sh and the received symbol s_1h lying closest to sh . The symbol error rate P_e , i.e. the probability for $d \neq 0$, can then be upper bounded by the union bound [71]

$$P_e \leq (M - 1)Q \left(\frac{d_{min}}{\sqrt{N_0/2}} \right). \quad (4.D.2)$$

We then have $\text{var}(\tilde{s}h) \leq |d_{min}|^2 P_e$.

To take a specific example we may mention the Wireless IP project [65], in which an early adaptive OFDM downlink was designed in which 10 % of the sub-symbols were used for the channel estimation and prediction. Two thirds of these sub-symbols were allocated for control symbols, while the others were known reference symbols. The control symbols were chosen to be QPSK symbols. In the QPSK constellation, adjacent symbols have separation $\sqrt{2E_s}$, which means that the noise contribution from erroneous detection, normalized by the received signal energy, is $2P_e$. For example, $P_e < 10^{-2}$ means that the normalized noise contribution must be at most -17 dB. If the additive noise v has considerably higher power than this value, then the impact of erroneous detection can be neglected.

4.E A simplified prediction model

To simplify the simulating of a communication system that uses prediction, it was suggested in [72] that a prediction of a channel coefficient can be simply modelled as

$$\hat{h} = ch + \tilde{h}, \quad (4.E.1)$$

where h is the true channel coefficient and $c = 1 - \sigma_h^2/\sigma_h^2$, and σ_h^2 and σ_h^2 are the channel variance and the prediction error as produced by an optimal predictor, respectively. \tilde{h} is here white noise with variance $c(1 - c)\sigma_h^2$ and uncorrelated with h .

This simple model can be used to model the predicted channel by generating a scalar estimate \hat{h} within each time-frequency resource block of an OFDM system, requiring one appropriately scaled Gaussian number to be generated per resource block. This model is useful e.g. in system-level simulations, where simplified models of the performance of channel estimation and prediction algorithms are needed for efficient simulation of a large number of communication sessions. However, the model 4.E.1 does not fully comply with the correlation properties featured by an optimal predictor. We now analyze how the simplified model differs from an optimal predictor.

For an optimal predictor, it holds that the prediction error $\tilde{h} = h - \hat{h}$ is independent of the optimal prediction \hat{h} rather than of h . It then holds from the variance $\sigma_{\tilde{h}}^2$ of the optimal prediction that

$$\sigma_{\tilde{h}}^2 = \sigma_h^2 - \sigma_{\hat{h}}^2. \quad (4.E.2)$$

It is easy to verify that

$$\|\{h, \hat{h}, \tilde{h}\}\|^2 = \begin{pmatrix} \sigma_h^2 & \sigma_{\hat{h}}^2 & \sigma_{\tilde{h}}^2 \\ \sigma_{\hat{h}}^2 & \sigma_{\hat{h}}^2 & 0 \\ \sigma_{\tilde{h}}^2 & 0 & \sigma_{\tilde{h}}^2 \end{pmatrix}. \quad (4.E.3)$$

For the model (4.E.1) it holds that

$$\|\{h, \hat{h}, \tilde{h}\}\|^2 = \begin{pmatrix} \sigma_h^2 & c\sigma_h^2 & 0 \\ c\sigma_h^2 & c\sigma_h^2 & c(1-c)\sigma_h^2 \\ 0 & c(1-c)\sigma_h^2 & c(1-c)\sigma_h^2 \end{pmatrix} = \begin{pmatrix} \sigma_h^2 & \sigma_{\hat{h}}^2 & 0 \\ \sigma_{\hat{h}}^2 & \sigma_{\hat{h}}^2 & c\sigma_{\tilde{h}}^2 \\ 0 & c\sigma_{\tilde{h}}^2 & c\sigma_{\tilde{h}}^2 \end{pmatrix}. \quad (4.E.4)$$

Thus, the statistical properties (4.E.4) of the model (4.E.1) differ only very slightly from the statistical properties (4.E.3) of a realizable predictor. The model (4.E.1) should therefore be adequate to use in most simulation situations.

4.F Alternative KF formulations

It was seen in Section 4.4 that the special matrix structures considered here help to reduce the numerical complexity considerably, as compared to the general KF. There are however formulations for the KF other than the standard formulation that could potentially reduce the numerical complexity even further. We study a few of these alternative formulations below.

4.F.1 The CKMS recursions

A strength of the KF recursions is that they are indifferent to whether or not the coefficient matrices are time varying. We have used this fact and

Table 4.2: Numerical complexity of the general CKMS algorithm and the corresponding relaxations emerging from block-structured matrices. subchannel modelling is assumed (the complexity for impulse response modelling is parenthesized). The term “unchanged” refers to the fact that the expression in question is the same as for the general case.

operation	complexity (general case)	complexity (here)
$\mathbf{U}_i = \mathbf{F}\mathbf{L}_i$	$n^2\alpha$	$n\alpha$
$\mathbf{V}_i = \mathbf{J}\mathbf{L}_i$	$nw\alpha$	$uw\alpha k(nw\alpha)$
$\mathbf{K}_{p,i} = \mathbf{K}_i \mathbf{R}_{e,i}^{-1}$	$nw^2 + w^3/6$	unchanged
$\mathbf{K}_{i+1} = \mathbf{K}_i - \mathbf{U}_i \mathbf{R}_{r,i}^{-1} \mathbf{V}_i^*$	$n\alpha^2 + nw\alpha/2 + \alpha^3/6$	unchanged
$\mathbf{L}_{i+1} = \mathbf{U}_i - \mathbf{K}_{p,i} \mathbf{V}_i$	$nw\alpha$	unchanged
$\mathbf{R}_{e,i+1} = \mathbf{R}_{e,i} - \mathbf{V}_i \mathbf{R}_{r,i}^{-1} \mathbf{V}_i^*$	$w\alpha^2 + \alpha w^2/2$	unchanged
$\mathbf{R}_{r,i+1} = \mathbf{R}_{r,i} - \mathbf{V}_i^* \mathbf{R}_{e,i}^{-1} \mathbf{V}_i$	$w^2\alpha + w\alpha^2/2$	unchanged
$\mathbf{P}_{i+1} = \mathbf{P}_i - \mathbf{L}_i \mathbf{R}_{r,i}^{-1} \mathbf{L}_i^*$	$n\alpha^2 + n^2\alpha/2$	unchanged

allowed the regressor matrix \mathbf{J}_t to be time varying. However, if we let \mathbf{J}_t be static so that all model matrices are constant, one would expect that part of the computational burden of the KF recursions could be lifted from the channel estimator if this fact was used in some way. The *Chandrasekhar, Kailath, Morf, Sidhu* (CKMS) recursions [49] are an exact reformulation of the standard KF recursions for the special case of constant model parameters, that generally require fewer operations per update than the standard KF.

Instead of propagating $\mathbf{P}_{t|t}$ and \mathbf{P}_t directly, the CKMS formulation propagates four smaller matrices \mathbf{K}_t , \mathbf{L}_t , $\mathbf{R}_{e,t}$, and $\mathbf{R}_{r,t}$. The matrices \mathbf{L}_t and $\mathbf{R}_{r,t}$ are initiated at $t = 0$ by a spectral factorisation

$$-\mathbf{L}_0 \mathbf{R}_{r,0}^{-1} \mathbf{L}_0^* = \mathbf{F} \mathbf{\Pi}_0 \mathbf{F}^* + \mathbf{G} \mathbf{Q} \mathbf{G}^* - \mathbf{K}_0 \mathbf{R}_{e,0}^{-1} \mathbf{K}_0^* - \mathbf{\Pi}_0, \quad (4.F.1)$$

with $\mathbf{K}_0 = \mathbf{F} \mathbf{\Pi}_0 \mathbf{H}^*$ and $\mathbf{R}_{e,0} = \mathbf{R} + \mathbf{H} \mathbf{\Pi}_0 \mathbf{H}^*$. The spectral factorisation yields a matrix \mathbf{L}_0 of size $n \times \alpha$, and a matrix $\mathbf{R}_{r,0}$ of size $\alpha \times \alpha$. The lower the α , the more beneficial the use of CKMS as compared to the standard KF formulation. The exact CKMS recursions and their corresponding numerical complexity is presented in Table 4.2. Unfortunately, very little is gained by exploiting the special matrix structures. Taking into account that the spectral factorization (4.F.1) has to be carried out every time the model is updated, it is unlikely that the CKMS filter would be useful to us in the present context.

4.F.2 Array algorithms

The KF can also be implemented as an *array algorithm*. One then updates so called *square-root factors* of $\mathbf{P}_{t|t}$ or \mathbf{P}_t . To exemplify, we show a so called *information* array filter, where inverses of the state estimation error covariances are propagated.

Construct the left hand side of (4.F.2) below and use QR decomposition to form

$$\begin{pmatrix} \mathbf{R}^{-*/2} & \mathbf{0} & \mathbf{0} \\ -\mathbf{F}^{-*}\mathbf{J}_t^*\mathbf{R}^{-*/2} & \mathbf{F}^{-*}\mathbf{P}^{-*/2} & \mathbf{0} \\ \mathbf{Q}^{*/2}\mathbf{G}^*\mathbf{F}^{-*}\mathbf{J}_t^*\mathbf{R}^{-*/2} & -\mathbf{Q}^{*/2}\mathbf{G}^*\mathbf{F}^{-*}\mathbf{P}_t^{-*/2} & \mathbf{I} \\ -\mathbf{y}_t^*\mathbf{R}^{-*/2} & \hat{\mathbf{x}}_t^*\mathbf{P}_t^{-*/2} & \mathbf{0} \end{pmatrix} = \Theta \begin{pmatrix} \mathbf{R}_{e,t}^{-*/2} & -\mathbf{K}_{p,t}^*\mathbf{P}_{i+t}^{-*/2} & (*) \\ \mathbf{0} & \mathbf{P}_{i+t}^{-*/2} & (*) \\ \mathbf{0} & \mathbf{0} & (*) \\ -\mathbf{e}_t^*\mathbf{R}_{e,t}^{-*/2} & \hat{\mathbf{x}}_{t+1}^*\mathbf{P}_{t+1}^{-*/2} & (*) \end{pmatrix}, \quad (4.F.2)$$

where Θ is a unitary matrix satisfying $\Theta\Theta^* = I$. Comparing the blocks at position $\{4, 2\}$ in the matrices in (4.F.2), it is seen that this operation performs an updating of $\hat{\mathbf{x}}_t^*\mathbf{P}_t^{-*/2}$ to $\hat{\mathbf{x}}_{t+1}^*\mathbf{P}_{t+1}^{-*/2}$. The $(*)$ are elements whose values are unimportant here.

Array algorithms can be implemented in such a way that they take virtually the same number of arithmetic operation to iterate as the standard KF. Also, they are often more numerically stable than more straightforward KF formulations. However, as was the case with the CKMS algorithm, they too suffer from not being able to exploit the block-structures of the matrices presented in this thesis. Since we have not seen any tendency of numerical misdemeanour in the algorithms presented in the main text, we find no reason to use array KF implementations here.

4.G Convergence of the periodic filter

Consider the model

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t, \\ \mathbf{y}_t &= \mathbf{J}_t\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \quad (4.G.1)$$

where \mathbf{u}_t , \mathbf{v}_t , and \mathbf{x}_0 are Gaussian, zero-mean, white, and uncorrelated, and

$$\|[\mathbf{u}_t^T, \mathbf{v}_t^T, \mathbf{x}_0^T]^T\|^2 = \text{diag}(\mathbf{Q}, \mathbf{R}, \Pi_0), \quad (4.G.2)$$

where $\mathbf{Q} > 0$, $\mathbf{R} > 0$, and $\Pi > 0$.

Theorem 4.G.1 (Filter convergence for time invariant models). *Assume the state space model (4.G.1) with a constant measurement matrix $\mathbf{J}_t \triangleq \mathbf{J}$ so that the model is time invariant. If the eigenvalues of \mathbf{F} are strictly inside the unit circle, the associated Kalman filter converges so that the filter gain $\mathbf{K}_{f,t}$ approaches a steady state \mathbf{K}_f , i.e. $\mathbf{K}_{f,t} \rightarrow \mathbf{K}_f$ as $t \rightarrow \infty$.*

Proof. See [56, Ch. 4]. □

Theorem 4.G.2 (Filter convergence for periodic models). *Assume the state space model (4.G.1) with periodic matrices $\{\mathbf{J}_t\}$ such that $\mathbf{J}_t = \mathbf{J}'_{t \bmod b}$, i.e. $\mathbf{J}_t \in \{\mathbf{J}'_0, \dots, \mathbf{J}'_{b-1}\}$ for known matrices \mathbf{J}_i , $0 \leq i < b$. Then if the eigenvalues of \mathbf{F} are strictly inside the unit circle, the associated Kalman filter converges to a periodic steady state, so that the Kalman filter gain $\mathbf{K}_{f,t}$ periodically assumes some values $\{\mathbf{K}'_{f,0}, \dots, \mathbf{K}'_{f,b-1}\}$, i.e. $\mathbf{K}_{f,ib+j} = \mathbf{K}'_{f,j}$ as $i \rightarrow \infty$ for all $j \in [0, b)$.*

Proof. We aim to prove that the one-step prediction error covariance matrix $\mathbf{P}_t = \|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}\|^2$ periodically assumes some values $\{\mathbf{P}'_0, \dots, \mathbf{P}'_{b-1}\}$, i.e. $\mathbf{P}_{ib+j} = \mathbf{P}'_j$ as $i \rightarrow \infty$ for all $j \in [0, b)$. Since $\mathbf{K}_{f,t}$ is a function of \mathbf{P}_t , it must then hold that $\mathbf{K}_{f,ib+j} = \mathbf{K}'_{f,j}$ as $i \rightarrow \infty$ for all $j \in [0, b)$.

Construct the time-invariant augmented state space model

$$\begin{aligned}
 \begin{bmatrix} \mathbf{x}_{t+b} \\ \vdots \\ \mathbf{x}_{t+2b-1} \end{bmatrix} &= \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{F} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{F}^2 \\ & & \vdots & \vdots \\ \mathbf{0} & & \mathbf{0} & \mathbf{F}^b \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t+b-1} \end{bmatrix} \\
 &+ \begin{bmatrix} \mathbf{G} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{FG} & \mathbf{G} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}^2\mathbf{G} & \mathbf{FG} & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \\ \mathbf{F}^{b-1}\mathbf{G} & & & & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{t+b-1} \\ \vdots \\ \mathbf{u}_{t+2b-2} \end{bmatrix} \\
 &= \mathbf{F}^{(aug.)} \mathbf{x}_t^{(aug.)} + \mathbf{G}^{(aug.)} \mathbf{u}_t^{(aug.)} \\
 \begin{bmatrix} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t+b-1} \end{bmatrix} &= \begin{bmatrix} \mathbf{J}'_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}'_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}'_{b-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t+b-1} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_{t+b-1} \end{bmatrix} \\
 &= \mathbf{J}^{(aug.)} \mathbf{x}_t^{(aug.)} + \mathbf{v}_t^{(aug.)},
 \end{aligned} \tag{4.G.3}$$

where $t = 0, b, 2b, \dots$. The zero matrices $\mathbf{0}$ in $\mathbf{F}^{(aug.)}$, $\mathbf{G}^{(aug.)}$, and $\mathbf{J}^{(aug.)}$ indicate $\mathbf{0}_{n \times n}$, $\mathbf{0}_{n \times xu}$, and $\mathbf{0}_{w \times n}$, respectively, where n , xu , and w are the

dimensions for \mathbf{x}_t , \mathbf{u}_t , and \mathbf{y}_t , respectively. The augmented vectors $\mathbf{u}_t^{(aug.)}$, $\mathbf{v}_t^{(aug.)}$, and $\mathbf{x}_0^{(aug.)}$ are Gaussian, independent, white, and zero-mean with time-invariant covariance matrices

$$\begin{aligned}\|\mathbf{u}^{(aug.)}\|^2 &= \text{diag}(\mathbf{Q}, \dots, \mathbf{Q}), \\ \|\mathbf{v}^{(aug.)}\|^2 &= \text{diag}(\mathbf{R}, \dots, \mathbf{R}),\end{aligned}\tag{4.G.4}$$

and

$$\|\mathbf{x}_0^{(aug.)}\|^2 = \begin{bmatrix} \Pi_0 & \Pi_0\mathbf{F}^* & \Pi_0\mathbf{F}^{*2} & \Pi_0\mathbf{F}^{*3} & \dots \\ \mathbf{F}\Pi_0 & \Pi_1 & \Pi_1\mathbf{F}^* & \Pi_1\mathbf{F}^{*2} & \dots \\ \mathbf{F}^2\Pi_0 & \mathbf{F}\Pi_1 & \Pi_2 & \Pi_2\mathbf{F}^* & \dots \\ \mathbf{F}^3\Pi_0 & \mathbf{F}^2\Pi_1 & \mathbf{F}\Pi_2 & \Pi_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},\tag{4.G.5}$$

where Π_t is given recursively by $\Pi_{t+1} = \mathbf{F}\Pi_t\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*$, $t \geq 0$. The model (4.G.3) is identical to (4.G.1) except that it models the state vectors in batches of b instances of the original state vector. It is easy to verify that the eigenvalue decomposition of $\mathbf{F}^{(aug.)}$ is

$$\mathbf{F}^{(aug.)} = \mathbf{V}_F \mathbf{D}_F \mathbf{V}_F^{-1}\tag{4.G.6}$$

with

$$\mathbf{V}_F = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \mathbf{F}^{-b+1} \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \vdots & \ddots & \ddots & \mathbf{F}^{-1} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_n \end{bmatrix}, \text{ and } \mathbf{V}_F^{-1} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & \mathbf{0} & -\mathbf{F}^{-b+1} \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \vdots & \ddots & \ddots & -\mathbf{F}^{-1} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_n \end{bmatrix},\tag{4.G.7}$$

and

$$\mathbf{D}_F = \begin{bmatrix} \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{F}^b \end{bmatrix},\tag{4.G.8}$$

where zero matrices $\mathbf{0}$ indicate $\mathbf{0}_{n \times n}$. Since the eigenvalues of \mathbf{F} lie strictly inside the unit circle, it is clear that also the eigenvalues of $\mathbf{F}^{(aug.)}$ are inside the unit circle. In light of Theorem 4.G.1, the ‘‘one-step’’ prediction error covariance of the Kalman filter associated with (4.G.3),

$$\mathbf{P}_t^{(aug.)} = \left\| \begin{bmatrix} \mathbf{x}_{t+b} \\ \vdots \\ \mathbf{x}_{t+2b-1} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{x}}_{t+b|t+b-1} \\ \vdots \\ \hat{\mathbf{x}}_{t+2b-1|t+b-1} \end{bmatrix} \right\|^2, \quad t = 0, b, 2b, \dots,\tag{4.G.9}$$

will then converge to a steady state solution $\mathbf{P}^{(aug.)}$. Evidently, the upper left $n \times n$ -block of $\mathbf{P}_t^{(aug.)}$ can be identified as \mathbf{P}_t , where $t = ib$, $i \in \mathbb{N}$, which hence converges to a constant value \mathbf{P}'_0 . Since \mathbf{P}_t for other time instants can be calculated by the Ricatti equation

$$\mathbf{P}_{t+1} = \mathbf{F}\mathbf{P}_t\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* - \mathbf{F}\mathbf{P}_t\mathbf{J}_t^*(\mathbf{J}_t\mathbf{P}_t\mathbf{J}_t^* + \mathbf{R})^{-1}\mathbf{J}_t\mathbf{P}_t\mathbf{F}, \quad (4.G.10)$$

and all matrices in (4.G.10) are b -periodic, periodicity holds for all \mathbf{P}_t . \square

4.H Model matrix structures

Here we summarize the sizes and structures of the matrices in the system model (4.2.41):

\mathbf{X}	$\in \mathbb{C}^{k \times k}$	diagonal
\mathbf{Y}	$\in \mathbb{C}^{k \times 1}$	full
\mathbf{Z}	$\in \mathbb{C}^{1 \times k}$	full
\mathbf{R}_γ	$\in \mathbb{R}^{m \times m}$	diagonal
\mathbf{D}	$\in \mathbb{R}^{N \times m}$	1 non-zero element/column or full
\mathbf{A}	$\in \mathbb{C}^{kx \times kx}$	diagonal
\mathbf{B}	$\in \mathbb{C}^{kx \times x}$	k non-zero elements/column, 1 non-zero element/row
\mathbf{C}	$\in \mathbb{C}^{x \times kx}$	1 non-zero element/column, k non-zero elements/row
\mathcal{F}_w	$\in \mathbb{C}^{w \times N}$	full
\mathcal{T}	$= \mathbf{I}_w$	diagonal (subchannel modelling)
\mathcal{T}	$\in \mathbb{C}^{w \times m}$	full (impulse response modelling)
\mathcal{TC}	$\in \mathbb{C}^{w \times kw}$	1 non-zero element/column, k non-zero elements/row (subchannel modelling)
\mathcal{TC}	$\in \mathbb{C}^{w \times km}$	full (impulse response modelling)
\mathbf{R}_h	$\in \mathbb{C}^{w \times w}$	full (subchannel modelling)
\mathbf{R}_h	$\in \mathbb{R}^{m \times m}$	diagonal (impulse response modelling)
\mathbf{L}	$\in \mathbb{C}^{w \times w}$	full (subchannel modelling)
\mathbf{L}	$\in \mathbb{R}^{m \times m}$	diagonal (impulse response modelling)
\mathbf{F}	$\in \mathbb{C}^{kxu \times kxu}$	diagonal
\mathbf{G}	$\in \mathbb{C}^{kxu \times xu}$	k non-zero elements/column, 1 non-zero element/row
\mathbf{H}	$\in \mathbb{C}^{wu \times kwu}$	1 non-zero element/column, k non-zero elements/row (subchannel modelling)
\mathbf{H}	$\in \mathbb{C}^{wu \times kmu}$	w non-zero elements/column, km non-zero elements/row (impulse response modelling)
Φ_t	$\in \mathbb{C}^{w \times wu}$	1 non-zero element/column, u non-zero elements/row
\mathbf{J}_t	$\in \mathbb{C}^{w \times kwu}$	1 non-zero element/column, ku non-zero elements/row (subchannel modelling)
\mathbf{J}_t	$\in \mathbb{C}^{w \times kmu}$	full (impulse response modelling)

Chapter 5

A channel estimation case study

The next generation wireless systems complying with IMT-advanced requirements need to operate in widely different deployment and usage scenarios. Support for flexible resource allocation is therefore important. This chapter illustrates a case study in which we investigate how channel estimation performs in a number of different OFDMA subchannel allocation schemes. We evaluate three types of channel estimation methods, which differ in the amount of data they use for producing estimates. We also investigate the impact of using various training signal designs (pilot patterns), in order to assess what the optimal strategy is when designing pilot patterns. The study is undertaken in various fading environments and for different user equipment (UE) velocities, based on the framework outlined in Chapter 4. The results show that channel estimation can perform well enough for time-frequency localized resources as small as 22 sub-symbols with two pilots in many important scenarios. The results provided can be used to identify appropriate subchannel allocations for the next generation OFDMA based wireless systems.

5.1 Introduction

Orthogonal Frequency Division Multiple Access (OFDMA) is a promising multiple access technique for future mobile communication systems because of its potential for high spectral efficiency. In a wireless multiuser system, mobile users will generally experience frequency selective and time varying (fading) radio channels. In some situations such as high Signal-to-Noise Ratios (SNR) and modest vehicular speeds, the channel quality can be accurately predicted to enable utilization of the channel frequency selectivity and time-variability for link adaptation and multi-user scheduling gains [65]. In other scenarios such as low SNR, high speeds and/or multicast transmis-

sion, the multi-path fading channel has to be mitigated by diversity-based transmission techniques to obtain robust transmission with low outage probability [32]. Good frequency diversity is enabled by allocating distributed subchannels for each user, since the fading of these resources then becomes almost uncorrelated. However, under such circumstances a channel estimator cannot take advantage of frequency correlation in the channel estimation procedure. Thus, there is a trade-off for the overall performance of a diversity based OFDMA system between diversity gains and channel estimation performance. The performance of channel estimation depends on the acceptable pilot overhead and deployed channel estimation algorithm, which may be limited by computational complexity.

In the EU FP6 WINNER II project [31], a Discrete Fourier Transform (DFT) precoded distributed subchannel scheme was defined, denoted Block Interleaved Frequency Division Multiple Access (B-IFDMA), for the diversity-based uplink mode [73], [74]. The corresponding scheme for the downlink without DFT precoding is denoted Block Equidistant Frequency Division Multiple Access (B-EFDMA). The channel estimation performance for these schemes has been studied within the WINNER II project assuming a frequency multiplexed pilot grid and 2D Wiener filter based channel estimation, [75], [76], [77]. With a dedicated frequency multiplexed pilot grid, the channel estimation problem becomes the same for both schemes.

In this chapter we complement the results in [75], [76], [77] by exploring how channel estimation performs depending on how many measurements are used in the estimation process. We compare different channel estimators in terms of their ability to correctly recover channel coefficients under different subchannel allocation schemes and different pilot patterns. The performance is evaluated in various fading environments and for different UE velocities. The overall goal of the investigations in this chapter is to guide in the selection of a proper subchannel allocation and pilot pattern for diversity-based OFDMA systems in the different evaluated scenarios.

5.2 System model

Orthogonal Frequency Division Multiplexing (OFDM) divides the radio channel resource into a time-frequency grid. In OFDMA, different subchannels and/or OFDM symbols can be allocated to different users. The way resources are assigned to users is here referred to as the *block allocation* (BA).

Six different BAs proposed by the WINNER II project for diversity-based transmission are studied [74]. All BAs investigated here are variants of B-IFDMA and B-EFDMA, presented in [73], [74]. Five of the six BAs use a 4-subchannels-by-3-OFDM-symbols block as smallest common unit, here

Table 5.1: Dimensions for the six BAs used in this chapter. In a dimension specification $a \times b$, a refers to frequency and b refers to time.

name	bin size (basic blocks)	bin size (sub-symbols)
B-IFDMA 1x1	1×1	4×3
B-IFDMA 1x2	1×2	4×6
B-IFDMA 2x1	2×1	8×3
B-IFDMA 2x2	2×2	8×6
LFDMA	2×2	8×12
IFDMA	–	1×12

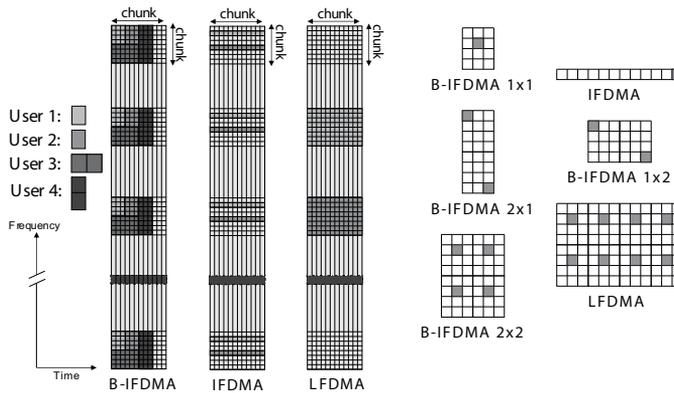


Figure 5.1: Illustration of investigated block allocations. The left side shows examples of allocations for multiple users. B-IFDMA 1x1, B-IFDMA 1x2, and B-IFDMA 2x1 are represented by users 1, 3, and 4 in the far left figure, respectively. (B-IFDMA 2x2 is not presented in the left figure.) The right side shows optimal or near-optimal pilot patterns as proposed in Section 5.4.

denoted a *basic block*. These BAs, which we call B-IFDMA 1x1, B-IFDMA 1x2, B-IFDMA 2x1, B-IFDMA 2x2, and LFDMA (localized frequency division multiple access), assign to each user, at each scheduling slot of 24 OFDM symbols, multiple ‘bins’ consisting of one basic block, two basic blocks stacked in frequency, two basic blocks stacked in time, two-by-two basic blocks, and two-by-four basic blocks, respectively. In the sixth BA, IFDMA (interleaved frequency division multiple access), each bin consists of one single subchannel with 12 OFDM symbols. The dimensions for all six BAs can be seen in Table 5.1.

At each scheduling slot, each user is allocated one or more bins, which

are allocated equidistantly in frequency with a spacing considerably larger than the channel coherence bandwidth. See Figure 5.1. The data rate for each user can thus be adjusted by varying the bin size and the number of bins per scheduling slot. To efficiently support robust transmission of small packets, a small bin size is desirable. As shown in [73, Figure 9], there is a large diversity gain for B-IFDMA and IFDMA compared to LFDMA.

While small bins scattered over a large portion of the frequency band ensures data integrity, it impedes channel estimation; all channel coefficients within a bin are fairly well correlated, but the correlation between coefficients in different bins is small. Hence we would expect B-IFDMA 1x1 and IFDMA to yield a comparably poor channel estimation performance while LFDMA should give good estimation results.

The bins are also separated in time. We assume half-duplex, so that each scheduling slot consists of 12 OFDM symbols for the uplink, followed by 12 OFDM symbols for the downlink. The scheduling slot period is therefore 24 OFDM symbols which means that there is, e.g. a gap of 18 OFDM symbols between consecutive B-IFDMA 1x1 blocks, but only a gap of 12 OFDM symbols between consecutive LFDMA blocks.

We here assume the use of *persistent scheduling* in the uplink, which means that the resource allocation for any given user is static and does not change over time. By assuming so, we guarantee that measurements of past bins are always available, so that the channel estimator has historic data at its disposal¹. Eight blocks of 12 sub-symbols are allocated per user at each scheduling slot. This means, for example, that eight bins per slot are allocated when B-IFDMA 1x1 and IFDMA are used, and one bin per slot is allocated when LFDMA is used.

The system parameter values that we use are those proposed by the WINNER project for urban base coverage [78]; the FFT bandwidth is 80 MHz divided on 2048 subchannels, yielding a subchannel spacing of approximately 39 kHz and an OFDM symbol duration of 25.6 μs (+3.2 μs for the cyclic prefix). The carrier frequency is set to 3.7 GHz.

If the cyclic prefix is at least as long as the channel's delay spread, and if the influences from non-linear system components are negligible, then inter-carrier and inter-symbol interference is avoided. The received signal \mathbf{y}_t over

¹No such assumption is required for the downlink, since pilots can there be assumed to be available at all pilot sub-locations, regardless of whether payload is transmitted and to whom it is transmitted. We assume single-antenna transmissions. In general, when using multi-antenna beamforming, two types of pilots are required: antenna-specific pilots (common reference signals), and beam-specific pilots (demodulation reference signals). The pilots discussed here are demodulation reference signals, used for coherent reception, but the distinction is not needed in the single transmit antenna case.

a portion of w parallel subchannels can then be modelled by (see Chapter 4)

$$\mathbf{y}_t = \Phi_t \mathbf{h}_t + \mathbf{v}_t, \quad (5.2.1)$$

where w is 1, 4, or 8 depending on which BA is used. The w -by- w diagonal matrix Φ_t holds the symbols transmitted over the w subchannels, the w -vector \mathbf{h}_t is the fading channel coefficients that we seek to estimate, and the elements of the w -vector \mathbf{v}_t is mutually uncorrelated circular symmetric complex Gaussian noise with unit variance. Choosing the number w of parallel subchannels to take into each measurement is a tradeoff between complexity and performance; measuring many subchannels at a time allows the channel estimator to take frequency correlation into account which improves the estimate, but it also increases numerical complexity.

The frequency-domain vector of channel coefficients \mathbf{h}_t are modelled on block-diagonal state-space form:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t, \\ \mathbf{h}_t &= \mathbf{H}\mathbf{x}_t, \end{aligned} \quad (5.2.2)$$

where the block-diagonal matrices \mathbf{F} , \mathbf{G} , and \mathbf{H} each consist of w blocks. Each such triplet $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ of blocks (or submatrices) is a model of order k that models one subchannel. Thus, with the terminology introduced in Chapter 4, we here use subchannel modelling, not impulse response modelling. In this chapter we assign a time-invariant model to the dynamics of the subchannels, thus assuming constant velocity and Doppler spectrum.

The subchannel model, the same for all w subchannels, is set up in such a way that it accommodates a certain Doppler spectrum. We study two different fading statistics. One, which we term *rich* fading, means to represent a scenario typical for cities, where sideways reflexes are generated by nearby buildings. The other, termed *Jakes-like* fading, represents sub-urban conditions with the mobile user surrounded by many equidistant scatterers. See Section 4.2.1 for details on how to construct these models. It is generally easier to perform channel estimation in the Jakes-like case than in rich fading. The two assumed Doppler spectra are plotted in Fig. 5.2. In the figure, the velocity is set to 50 km/h, giving a practical maximum Doppler frequency of just below 200 Hz for that velocity at a 3.7 GHz carrier frequency, although the models will allow some amount of higher frequencies.

The correlation between subchannels, which is crucial to channel estimation performance, is described by the channel model. The model used here is the WINNER II C2 Non-Line-Of-Sight (NLOS) Power Delay Profile (PDP)[79], reproduced in Section 2.3.2. These correlation properties are built into the channel model by scaling the w -by- w covariance matrix for the process noise \mathbf{u}_t in the model (5.2.2) in such a way that the covariance matrix

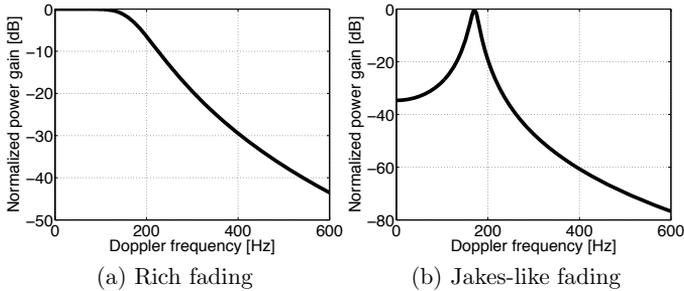


Figure 5.2: Doppler spectra for the two fading scenarios.

for the vector of channel coefficients \mathbf{h}_t is the Fourier transform of the PDP (see Theorem 4.2.1 in Chapter 4).

5.3 Pilot-aided channel estimation

Pilot-aided channel estimation inserts known symbols, called *pilots*, into the time-frequency grid. The specific choice of sub-locations for these pilots is referred to as the *pilot pattern*. By considering only the pilots in the measurements, disregarding all the payload symbols and inserting zeros into Φ_t at payload sub-locations, the pilot matrix Φ_t in (5.2.1) becomes known but time-variant.

It is of interest to keep the *pilot overhead* low, i.e. the ratio between number of pilots and the total number of sub-symbols. In all the BAs that we study here, the pilot overhead is $1/12$.

Since persistent scheduling is used in the uplink case, the channel estimator has access to a continuous flow of pilot measurements on which it can base its channel estimates. We are interested in channel estimation after a whole bin has been received, so that measurements all the way to the end of the bin are available. We consider three cases as displayed in Figure 5.3:

B-MMSE filters, which stands for *Block-MMSE* filters, base their estimates only on measurements within the bin. Historic data are not considered.

Kalman (KF) filters continuously produce estimate from the most recent measurement and take all past measurements into account. However, no “future” measurements are used, so that estimates for channel coefficients early in the bin are mostly based on measurements from earlier bins.

Smoothed Kalman filters use all available measurements and can be produced only when all the pilots of a bin have been received.

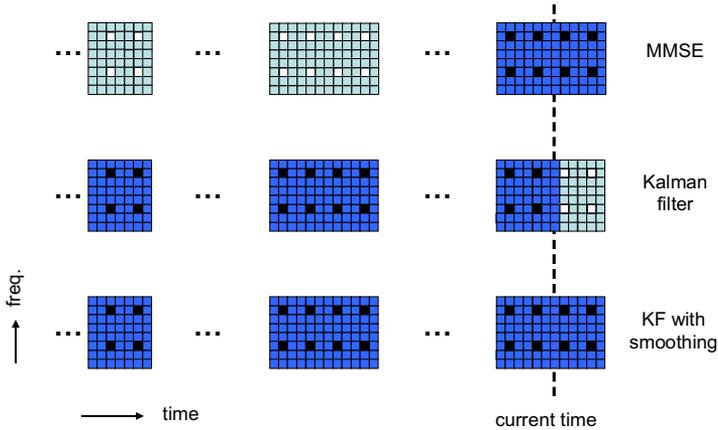


Figure 5.3: Schematic illustration of the three types of channel estimators examined. Here, the LFDMA block allocation is used. Black squares indicate pilot measurements that are used for forming channel estimates, while white squares are pilots disregarded by the respective estimators.

Figure 5.3 illustrates how the different types of estimators use available measurements. Below, we describe briefly how the respective estimators operate.

5.3.1 Non-smoothed Kalman filter

The state-space model (5.2.2) enables us to directly write down the optimal Kalman observer of the state vector \mathbf{x}_t (see Section 4.3):

$$\hat{\mathbf{x}}_{t|t} = \underbrace{(\mathbf{I} - \mathbf{K}_{f,t}\mathbf{H})\mathbf{F}}_{\mathbf{A}_t} \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{K}_{f,t}\mathbf{y}_t. \quad (5.3.1)$$

The so called *filtered* state estimate $\hat{\mathbf{x}}_{t|t}$ is the minimum mean square estimate *given all measurements up to time t* . We assume in this chapter that enough pilot data has then been received that the channel estimator has settled to a steady state. The Kalman gain $\mathbf{K}_{f,t}$, which is generally time-dependent and needs to be updated by a computationally demanding Riccati equation, will then be periodic-constant since the pilot matrix Φ_t is periodic, see Section

4.3.2. The periodic series of matrices $\{\mathbf{K}_{f,t}\}$ may therefore be precomputed. For the same reason, the matrix \mathbf{A}_t is also known. Once the state estimation vector has been computed, channel coefficient estimates are calculated by

$$\hat{\mathbf{h}}_{t|t} = \mathbf{H}\hat{\mathbf{x}}_{t|t}. \quad (5.3.2)$$

5.3.2 Smoothed Kalman filter

The KF estimate (5.3.1) may be improved upon by smoothing the estimates. Assume that the bin begins at time $t_1 + 1$ and ends at time t_2 . Since measurements up to and including time t_2 is available, we may produce the *smoothed* estimates $\hat{\mathbf{h}}_{t_1+1|t_2}, \hat{\mathbf{h}}_{t_1+2|t_2}, \dots, \hat{\mathbf{h}}_{t_2|t_2}$ instead of $\hat{\mathbf{h}}_{t_1+1|t_1+1}, \hat{\mathbf{h}}_{t_1+2|t_1+2}, \dots, \hat{\mathbf{h}}_{t_2|t_2}$ as the non-smoothed KF would produce. Smoothed Kalman estimates can be computed from the one-step predictions $\hat{\mathbf{x}}_{t|t-1}$, which are computed with the non-smoothed KF:

$$\hat{\mathbf{x}}_{t|t_2} = \hat{\mathbf{x}}_{t|t-1} + \sum_{j=t}^{t_2} \mathbf{B}_j(\mathbf{y}_j - \mathbf{H}\hat{\mathbf{x}}_{j|j-1}), \quad (5.3.3)$$

where the matrices \mathbf{B}_j may be precomputed here (see [56, Ch. 7]). Again, estimates of the channel coefficients are produced by (5.3.2).

5.3.3 B-MMSE

The last channel estimation algorithm that we examine is the most commonly suggested in the literature. It conducts channel estimation bin-wise, basing the estimates only on the measurements taken from the bin. The theory behind this type of estimator was presented in Section 3.1. Such algorithms go by many names²; we refer to it here, as in Chapter 3, as the *block minimum mean square error* (B-MMSE) estimator.

The B-MMSE estimator works as follows: given two Gaussian vectors \mathbf{h} and \mathbf{y} with cross-covariance R_{hy} and the latter with covariance R_y , the optimal least squares estimate of \mathbf{h} given \mathbf{y} is

$$\hat{\mathbf{h}} = \boldsymbol{\mu}_h + \mathbf{R}_{hy}\mathbf{R}_y^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \quad (5.3.4)$$

where $\boldsymbol{\mu}_h$ and $\boldsymbol{\mu}_y$ are the mean values of \mathbf{h} and \mathbf{y} , respectively, which we here set to all-zeros. The matrices \mathbf{R}_{hy} and \mathbf{R}_y can be relatively easily derived from the models (5.2.1) and (5.2.2) by setting up an extended state-space over all the time steps in the bin.

²The name *Wiener filters* is common in the literature. That notion however has ambiguous meaning so we refrain from using it here.

We here use an alternative method where we use the smoothed Kalman filter to produce the B-MMSE estimate. The vector of channel coefficients \mathbf{h}_t is associated with a prior distribution that represents the knowledge that we have about the channel before measurements have been received. According to Section 3.2.3, this prior distribution is set as $p(\mathbf{h}_t|I) = \mathcal{CN}(\mathbf{h}; \mathbf{0}, \mathbf{R}_h)$ for all t , where \mathbf{R}_h is the (known) covariance matrix for the channel coefficients over the w subchannels, with the SNR along is diagonal, since the noise has unit variance. By using this prior and running the smoothed KF over only one bin, so that it does not necessarily converge, the B-MMSE estimate is retrieved, because no past data is then utilized. Alternatively, the smoothed KF run to convergence but with the bins extensively separated in time can also be used to produce B-MMSE estimates. Both these approaches are identical to (5.3.4).

In the present investigation, we are not directly interested in the estimates (5.3.2) and (5.3.4), but rather by the *Normalized Mean Square Error* (NMSE) that they give rise to. Assuming a correct modelling of the actual fading channel conditions, NMSE values can be calculated directly from the model without conducting simulation, see Section 4.3 and Equation (4.3.14).

Of the three estimation methods, the smoothed KF will always have the best performance, since it uses all past measurements. The B-MMSE estimator and the non-smoothed KF will outperform one another depending on situation; while the B-MMSE estimator only uses local data and keeps no record of measurements of previous bins, the non-smoothed KF learns from history. However, the B-MMSE filter will use all data that it has available even for estimating the channel coefficients located early in the bin. As a contrast, the non-smoothed KF will generally give poor performance for early channel coefficients, since it at that point has *only* historic measurements to base its decision on.

Before evaluating the respective channel estimation methods on the six different BAs, we need to define a performance criterion with which the channel estimators can be evaluated.

5.3.4 Criterion

For any particular channel coefficient h (an element in \mathbf{h}_t), the channel estimator produces an estimate \hat{h} (e.g. an element in $\hat{\mathbf{h}}_{t|t}$) with an estimation error $\tilde{h} \triangleq h - \hat{h}$, so that h has the pdf $\mathcal{CN}(h; \hat{h}, \sigma_{\tilde{h}}^2)$, where $\sigma_{\tilde{h}}^2$ is the variance of the white zero-mean process \tilde{h} . An element in (5.2.1) can then be written

$$y_t = s_t h_t + v_t = s_t \hat{h}_t + s_t \tilde{h}_t + v_t. \quad (5.3.5)$$

Here, s_t is a diagonal element in Φ_t , i.e. a pilot, and v_t is additive complex Gaussian noise. The channel equalizer and detector then operate on the

signal $s_t \hat{h}_t$. Assuming that the pilots have mean power S , the experienced effective SNR after the estimator is

$$\text{SNR}_{\text{eff}} = \frac{S\sigma_{\hat{h}}^2}{S\sigma_h^2 + \sigma_v^2}, \quad (5.3.6)$$

where σ_v^2 is the disturbance noise variance and σ_h^2 is the (prior) variance of the estimate \hat{h} . We have then assumed that the estimate \hat{h}_t is uncorrelated with v_t . See Section 7.2.3 for a further discussion on this subject. The effective SNR should be compared to the SNR

$$\gamma \triangleq S\sigma_h^2/\sigma_v^2 \quad (5.3.7)$$

experienced when no channel estimation error is present, where σ_h^2 is the variance of the channel coefficient h and assumed known. We demand that the channel estimator must not decrease the effective SNR by more than some factor c . We may, without loss of generality, assume that $S = 1$ and $\sigma_v^2 = 1$. Assuming an optimal estimator so that $\sigma_{\hat{h}}^2 = \sigma_h^2 + \sigma_{\tilde{h}}^2$ and seeking a constraint on the NMSE, the following inequality on the NMSE $\xi \triangleq \sigma_{\tilde{h}}^2/\sigma_h^2$ must hold:

$$\frac{\sigma_{\hat{h}}^2}{\sigma_v^2 + \sigma_{\hat{h}}^2} = \frac{\sigma_h^2 - \sigma_{\tilde{h}}^2}{\sigma_v^2 + \sigma_{\hat{h}}^2} = \frac{1 - \xi}{\gamma^{-1} + \xi} \geq c\gamma \quad (5.3.8)$$

Rearranging the above gives us the lower acceptable value for the NMSE ξ :

$$\xi \leq \frac{1 - c}{1 + c\gamma}. \quad (5.3.9)$$

Note that this expression has the asymptotes $(1 - c)$ for low SNR values and $\gamma^{-1}(1 - c)/c$ for high SNR values.

Particularly, if we can accept a decrease of the effective SNR of 3 dB compared to the SNR γ , we set $c = 1/2$ and find the condition on ξ to be

$$\xi \leq (2 + \gamma)^{-1} \quad (5.3.10)$$

In this chapter we will for illustration accept a reduction of the effective SNR of 3 dB. We therefore use the criterion (5.3.10) as performance limit. As a good approximation, one may use the limit $\xi \leq \gamma^{-1}$ when $\gamma \geq 12$ dB.

5.4 Experimental results

The BAs are tested for two different fading statistics (see Section 5.2) and three different velocities: 5 km/h, 50 km/h, and 250 km/h. Perfect model match is assumed, so that the model (5.2.2) of the fading statistics correctly

represents the statistical properties of the real channel. We also assume stationary pilot transmission conditions (that is, static channel model matrices except for the pilot matrix) in which transients from earlier conditions have faded. As mentioned earlier, this would require a continuous pilot transmission (corresponding to persistent scheduling of payload data) from UEs over uplinks. All of the above six conditions (two types of fading statistics times three different velocities) are tested for each and every one of the six BAs, giving a total of 36 scenarios.

The algorithms are tested over a range of SNRs³ and measured in terms of the resulting NMSE. The total error contributed by the SNR and the NMSE ultimately results in a bit error rate (BER) or a frame error rate (FER). For a given decoding scheme, the ideal BER/FER curve is that produced when $\text{NMSE} = -\infty$ dB. When $\text{NMSE} > 0$, the BER/FER curve is then displaced from the ideal curve. Using the criterion (5.3.10), this displacement is less than 3 dB. Further, results in [77] and [76] suggest that the displacement can be improved down to less than a mere 1 dB displacement when iterative channel estimation (see Appendix 4.D) is used. Motivated by this, we use (5.3.10) as criterion for acceptable channel estimation performance.

Results for the 36 scenarios that we have investigated are illustrated in Figures 5.5–5.10. To give an overview of all the results, we summarize in

	B-IFDMA	IFDMA	B-IFDMA 1x1	B-IFDMA 1x2	B-IFDMA 2x1	LFDMA
5 km/h, Jakes-like fading						
5 km/h, rich fading						
50 km/h, Jakes-like fading						
50 km/h, rich fading						
250 km/h, Jakes-like fading						
250 km/h, rich fading						

Figure 5.4: Summary of 36 combinations and block allocations, velocities, and fading statistics. Dark grey indicates that none of the three channel estimators suggested here meet the performance criterion. Light grey indicates that B-MMSE does not meet the criterion. White indicates that both B-MMSE and smoothed KF, but not necessarily non-smoothed KF, meet the criterion. Detailed results for twelve combinations are presented in other figures, as indicated by the numbers and letters.

Figure 5.4 which of the 36 scenarios that do not meet the performance limit over the whole SNR range for any of the estimation methods, when certain “hand-picked” pilot patterns are used (see below). Also, we indicate for which

³We here assume matched filter reception so that $\text{SNR} = E_s/N_0$, where E_s is the symbol energy and N_0 is the power spectral density for the total noise [71].

scenarios that the smoothed KF but not the B-MMSE filter will comply with the NMSE criterion over the whole range of SNR values.

5.4.1 Impact of pilot pattern

Although the assumed pilot overhead is fixed to $1/12$, it is not evident where exactly the pilots should be placed in the time-frequency grid. We therefore conduct exhaustive searches over all possible patterns in order to find the optimal pattern for each BA. We have also hand-picked one pattern for each BA that performs well in all scenarios. These patterns are illustrated in Figure 5.1. Exhaustive searches could not be performed for the last two BAs. Instead, 100 random patterns were set to compete.

The channel estimation results are presented in Figures 5.5–5.10. The performance for the best and worst pilot patterns for the respective estimation methods are plotted, along with the hand-picked pattern. The main conclusion here is that the choice of pilot pattern has a large impact on the estimation performance. The patterns presented in Figure 5.1 have close to optimal performance in all situations.

5.4.2 Impact of UE velocity and fading statistics

We examine how UE velocity and fading statistics affect channel estimation performance. We find that KF estimation is generally superior to B-MMSE, except for high velocities and rich fading, where smoothed KF and B-MMSE have the same performance. It is also important to use smoothing when using KF for the ‘harsher’ scenarios (50 km/h with rich fading, and 250 km/h).

5.4.3 Impact of block allocation choice

The choice of block allocation greatly impacts estimation performance. For low velocities, all six BAs considered here meet the performance criterion over the whole SNR range, although KF estimation needs to be used for B-IFDMA 1×1 . However at higher velocities, the performance for BAs using small bins quickly deteriorates, especially in rich fading scenarios. Evidently, the use of B-IFDMA 1×1 should be questioned in the studied scenario, and IFDMA may need smoothed KF estimation, which has a very high numerical complexity. Small bins may however be attractive when transmitting small packets in situations when a low outage probability is more important than the power efficiency of the transmission, or in the downlink when pilots from adjacent users can be used. We also note that KF with smoothing must be used for the 24-symbols-per-bin schedules at high velocities in order to

meet the performance boundary, and this only works in Jakes-like fading environments.

5.5 Conclusions

In this chapter we have investigated channel estimation for OFDMA based resource allocation schemes. Within the framework of Block Interleaved Frequency Division Multiple Access (B-IFDMA) we showed that channel estimation can perform well for small time-frequency localized resources, here called bins. Small bins are useful for robust frequency-diversity based transmission of small packets. Combined with persistent scheduling, we find that smoothed Kalman filtering estimation is generally superior to B-MMSE estimation, except for high velocities and rich fading, where smoothed Kalman filtering and B-MMSE filters have the same performance. With Kalman filtering, bins as small as 22 channel symbols and two pilots can be used even at high velocities if the fading statistics of the radio channel is favourable.

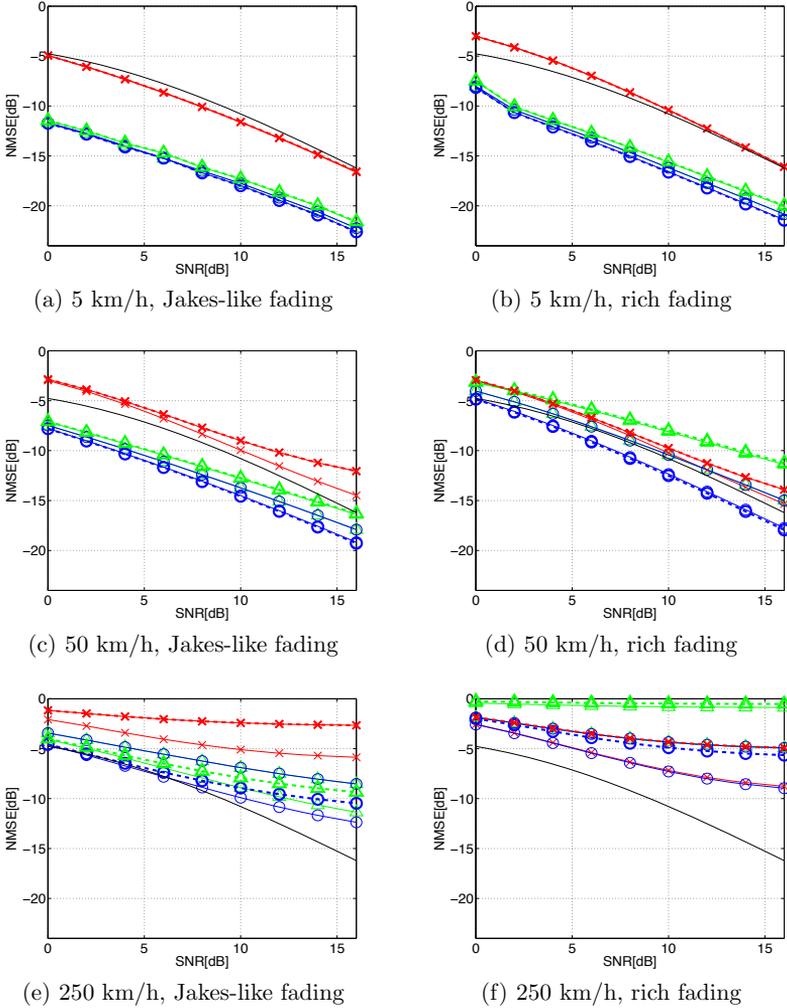


Figure 5.5: Estimation performance for IFDMA at different velocities and with different fading statistics. Simulations were carried out over every possible one-pilots pattern. Worst and best performance curves are indicated by thin solid lines for Kalman filtering (KF) with smoothing (blue, circles), KF without smoothing (green, triangles), and B-MMSE (red, crosses). Thick dashed lines indicate performance for the hand-picked pattern illustrated in Figure 5.1. However, for the B-MMSE estimator, the hand-picked pattern is the one where the pilot is in the middle of the bin, and not at the end. The solid line without markers is the acceptance boundary for the criterion (5.3.10).

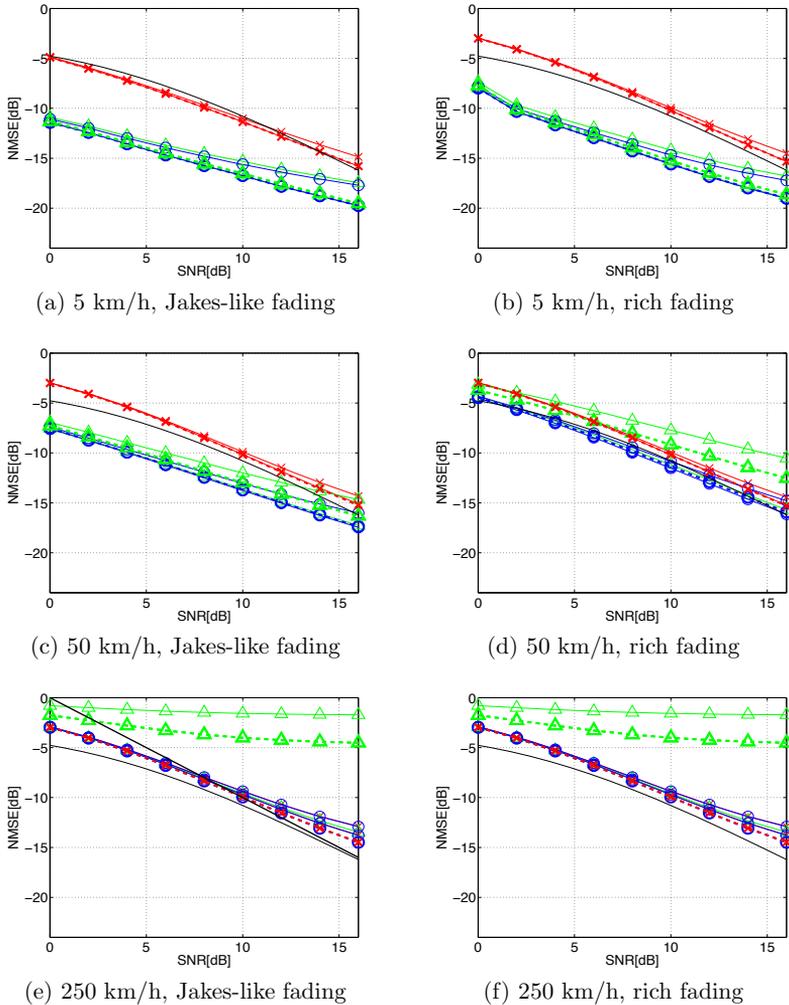


Figure 5.6: Estimation performance for B-IFDMA 1x1-mode at different velocities and with different fading statistics. Simulations were carried out over every possible one-pilots pattern. Worst and best performance curves are indicated by thin solid lines for Kalman filtering (KF) with smoothing (blue, circles), KF without smoothing (green, triangles), and B-MMSE (red, crosses). Thick dashed lines indicate performance for the hand-picked pattern illustrated in Figure 5.1. The solid line without markers is the acceptance boundary for the criterion (5.3.10).

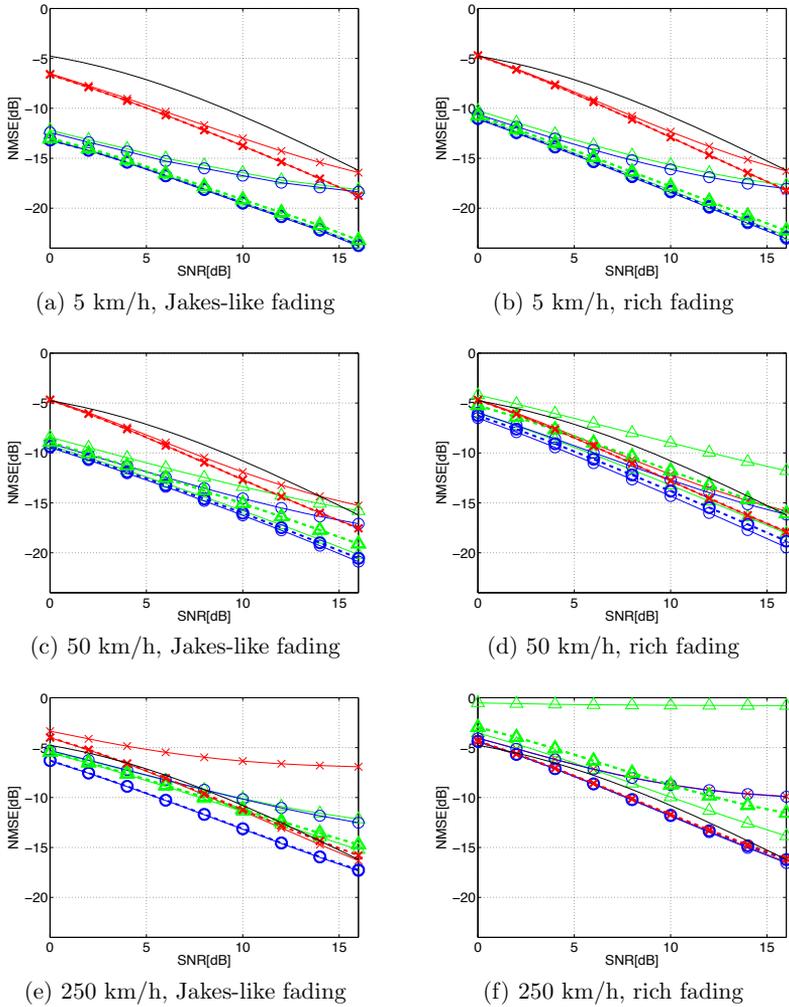


Figure 5.7: Estimation performance for B-IFDMA 1x2-mode at different velocities and with different fading statistics. Simulations were carried out over every possible two-pilots pattern. Worst and best performance curves are indicated by thin solid lines for Kalman filtering (KF) with smoothing (blue, circles), KF without smoothing (green, triangles), and B-MMSE (red, crosses). Thick dashed lines indicate performance for the ‘opposite corners’ pattern illustrated in Figure 5.1. The solid line without markers is the acceptance boundary for the criterion (5.3.10).

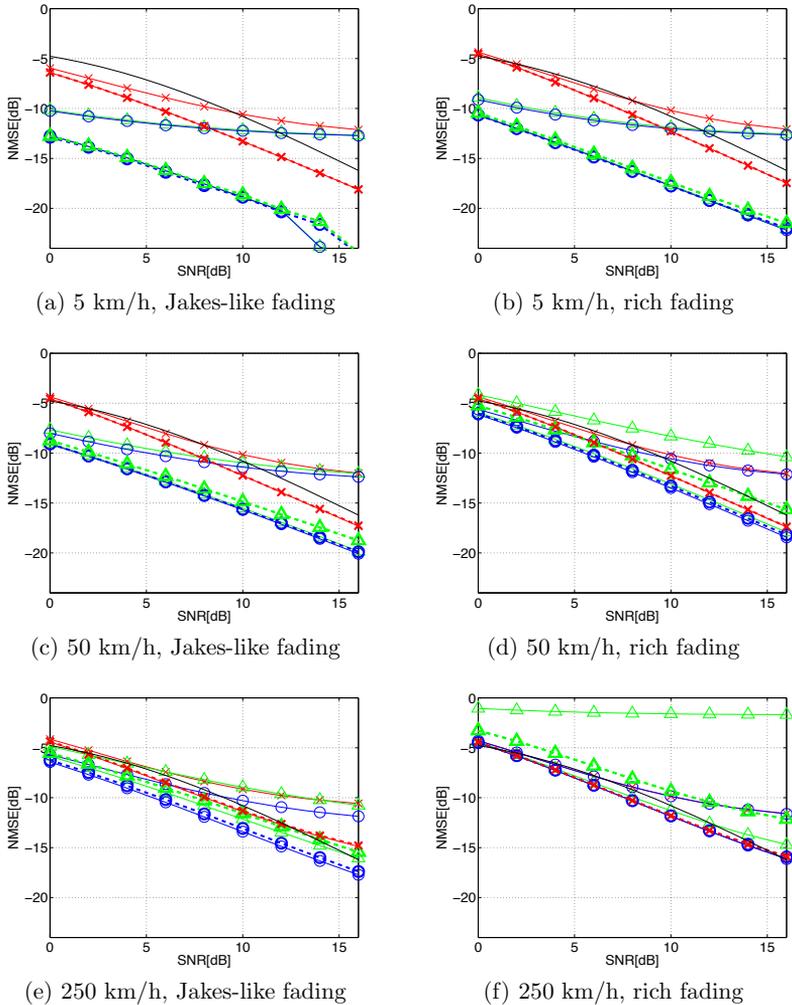


Figure 5.8: Estimation performance for B-IFDMA 2x1-mode at different velocities and with different fading statistics. Simulations were carried out over every possible two-pilots pattern. Worst and best performance curves are indicated by thin solid lines for Kalman filtering (KF) with smoothing (blue, circles), KF without smoothing (green, triangles), and B-MMSE (red, crosses). Thick dashed lines indicate performance for the ‘opposite corners’ pattern illustrated in Figure 5.1. The solid line without markers is the acceptance boundary for the criterion (5.3.10).

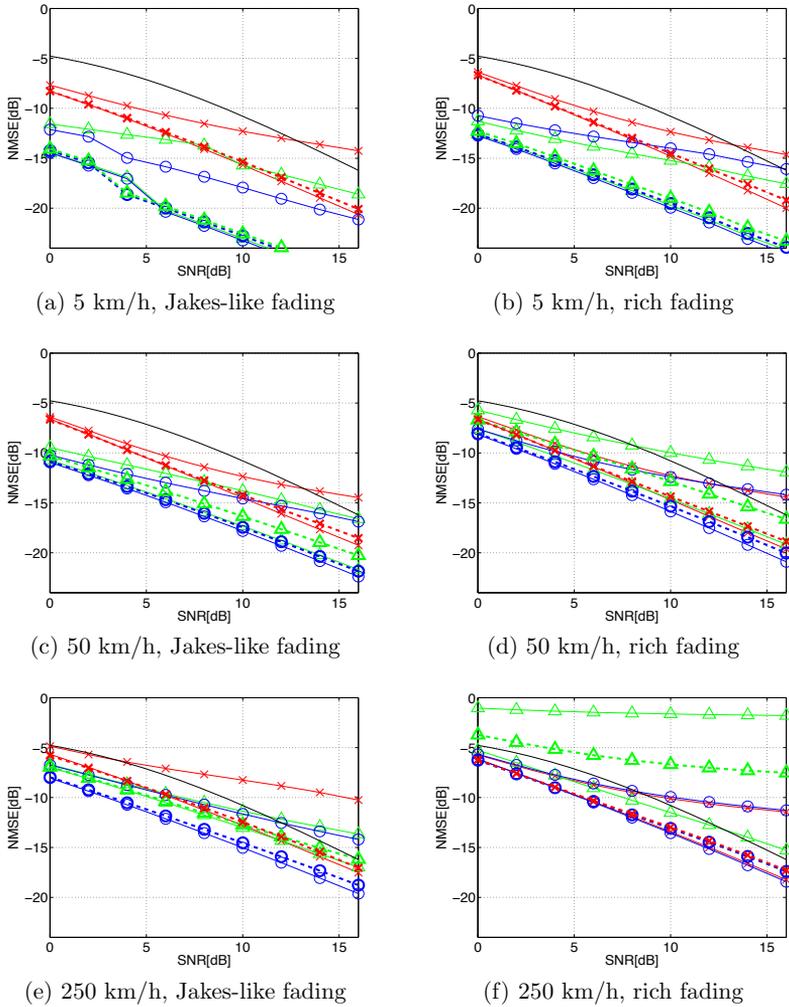


Figure 5.9: Estimation performance for B-IFDMA 2x2-mode at different velocities and with different fading statistics. Simulations were carried out over 100 randomly picked pilot patterns. Worst and best performance curves are indicated by thin solid lines for Kalman filtering (KF) with smoothing (blue, circles), KF without smoothing (green, triangles), and B-MMSE (red, crosses). Thick dashed lines indicate performance for the hand-picked pattern illustrated in Figure 5.1. The solid line without markers is the acceptance boundary for the criterion (5.3.10).

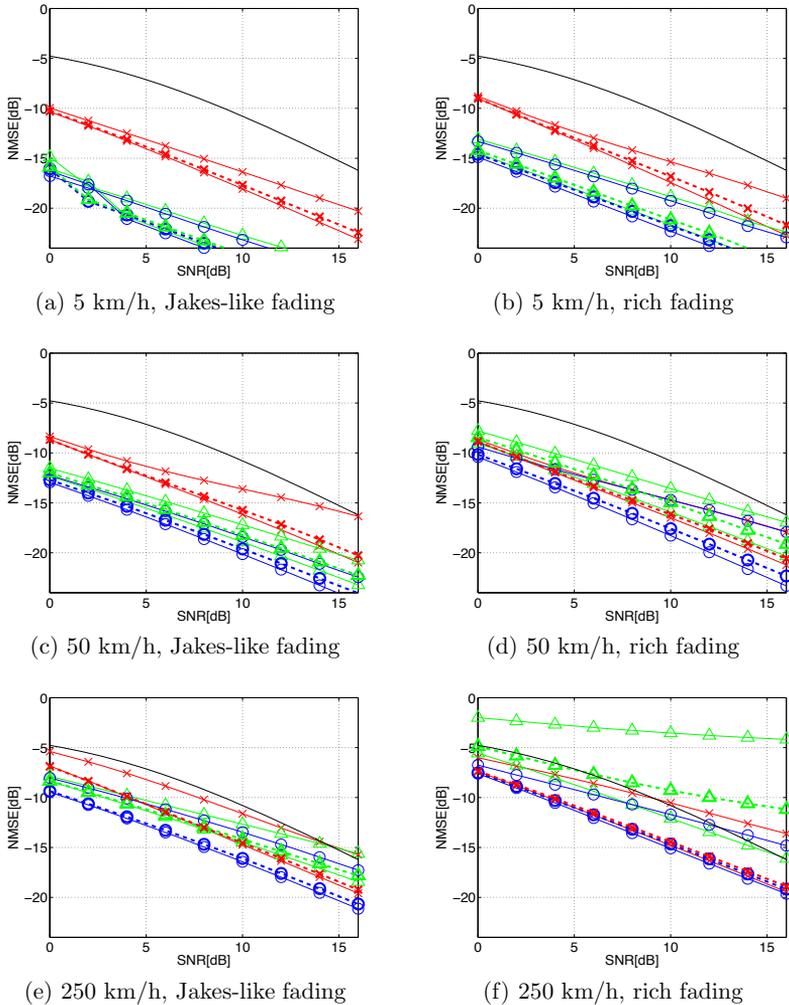


Figure 5.10: Estimation performance for LFDMA at different velocities and with different fading statistics. Simulations were carried out over 100 randomly picked pilot patterns. Worst and best performance curves are indicated by thin solid lines for Kalman filtering (KF) with smoothing (blue, circles), KF without smoothing (green, triangles), and B-MMSE (red, crosses). Thick dashed lines indicate performance for the hand-picked pattern illustrated in Figure 5.1. The solid line without markers is the acceptance boundary for the criterion (5.3.10).

Chapter 6

The OFDMA uplink design

Frequency-adaptive multiuser scheduling in OFDM utilizes the frequency-selective small-scale fading to allocate subchannels with advantageous signal-to-noise ratio (SNR) to each user. Due to channel time-variability and delays of the transmission control loop, this will in general require channel prediction. FDD (Frequency Division Duplex) multi-input transmissions, such as uplinks with multiple users, pose the most challenging prediction problem: all subchannels that may potentially be allocated must here be predicted for all involved user equipments (UEs), based on pilots transmitted from all UEs. This poses challenges with respect to prediction accuracy, estimator complexity and pilot overhead. This chapter explores the design, performance and complexity of Kalman predictors used for multi-user uplink prediction, in the context of the EU WINNER project baseline design system. One conclusion is that multi-user uplink prediction that is useful at vehicular velocities in 4G systems operating at 3-5 GHz is indeed feasible, if very fast resource allocation is implemented, so that short prediction horizons are adequate. However, the channel predictability depends crucially on the local fading environment, so predictors should be based on models of the Doppler spectrum for each terminal.

Although we here formulate the problem as a multiuser uplink problem, the present investigation is also valid for multi-input downlinks in coordinated multi-point (CoMP) scenarios, where several base stations transmit simultaneously to one UE.

6.1 Introduction

The use of adaptive OFDMA on wide-band channels enables the allocation of UEs to the frequency and time regions that are most advantageous for them. A multiuser scheduler may allocate the transmission to/from each UE

to appropriate frequency bands by utilizing the channel variations due to the small-scale frequency selective fading. Channels to/from each UE will in general vary independently. Substantial multiuser scheduling gains can be attained if each of u UEs is allocated to the fraction $1/u$ of the time-frequency regions having the highest SNR for that particular UE. Assume that all of the u UEs experience equi-powered Rayleigh fading channels that can be perfectly predicted and tracked by the receiver, and the scheduler always allocates the resources to the UE having the best channel. Then, each UE will on average experience an SNR of $\gamma \sum_{i=1}^u i^{-1}$, where γ is the mean SNR that a UE would experience were that UE alone in the system [32]. Additional (but smaller) gains are obtained by also using link adaptation that is adjusted to each allocated subchannel.

Due to these potential gains, frequency-adaptive transmission in OFDMA downlinks has been of interest in the research community, for example in the European beyond-3G WINNER project [31]. OFDMA is used in uplinks in the WiMAX system [42]¹.

One complication in systems that use adaptive OFDMA is that for moving terminals, measurements of the channel gain for a subchannel will quickly become outdated. Figure 6.1 illustrates the variation of received power with time and frequency for one particular UE and fading pattern.

Frequency-adaptive transmission for vehicular UEs would therefore require low latency control loops for the transmission and also efficient channel prediction of the SNR at the instant of transmission. Designs for downlinks up to 100 km/h at 2 GHz carrier have been investigated within the Swedish Wireless IP project [81] and downlink designs targeted at up to 70 km/h at 5 GHz have been proposed and evaluated within the WINNER project [66], [82].

For FDD downlinks and TDD (time division duplex) uplinks and downlinks, channel prediction can be based on downlink pilots that are transmitted by the base station to all terminals within a sector or beam. (In TDD, prediction of the downlink channel gains can be used also for the uplink gains, due to the channel reciprocity.) Channel predictors at each terminal may then predict the frequency selective channel over a band of interest [4], [60]. Suitably compressed messages are reported to the scheduler at the base station, so that the scheduler may allocate the sub-bands among the UEs.

The potential multiuser scheduling gain is as large in FDD uplinks as in the cases outlined above, but the channel prediction problem becomes much harder: FDD uplinks cannot be predicted based on downlink pilots, because

¹In the 3GPP LTE standard [80],[33], OFDMA is used in downlinks, but not in uplinks. There, larger contiguous frequency blocks are allocated to each user. This scheme has been selected because it enables the use of signalling waveforms with low envelope variations, which facilitate the design of low-cost terminals with high power efficiency.

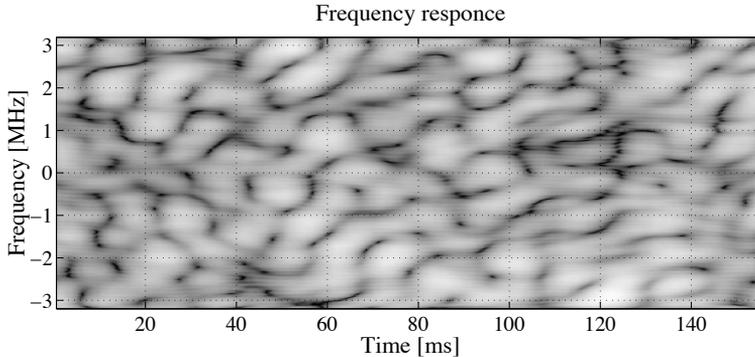


Figure 6.1: Time-frequency representation of an estimated channel obtained from measurement data on a 6.4 MHz channel at a 1880 MHz carrier. White colour denotes high power whereas dark colour denotes low power. The dynamic range and the speed of the mobile is approximately 40 dB and 50 km/h, respectively. The coherence bandwidth is 0.6 MHz in this example. This particular channel was studied in [4].

uplinks and downlinks work at significantly different carrier frequencies. Predictors at the base station will have to predict the channels from all UEs that compete for a set of transmission resources. To support this prediction, all of these UEs would have to send special-purpose pilots within all resources of interest, at an appropriate channel sampling rate². The need for such special uplink pilots leads to two problems that become severe when the number of competing UEs is large:

1. If uplink pilots are transmitted in sub-locations that are exclusively allotted to each UE, then the overhead of earmarked pilot sub-locations relative to payload could become unacceptably large.
2. If u UEs compete for a set of resources, then each of them will on average obtain only u^{-1} of the resources but will still have to transmit pilots in all of them. For large values of u , this pilot power overhead will overshadow the multiuser scheduling throughput gains.

The multiuser scheduling gains increase significantly with u for small values

²If only the pilots that are embedded in uplink payload transmissions are used, then the channel sampling would depend on the availability and scheduling of the uplink transmissions. Extrapolation to other frequency bands beyond the correlation bandwidth could not be performed. Also, the sampling in any given band could not be relied upon to have sufficient rate to support a reliable prediction of the channel in that band.

of u but slower when u increases³. This makes it possible to handle both of the problems outlined above by partitioning the total bandwidth into a number of *competition bands*, each with a limited number of competitors (typically $u = 8$ or less). The competition bands should be composed of frequency resources that are well spread out over the uplink bandwidth, to sample the available frequency selectivity.

Our problem here will therefore be to design and assess channel predictors that work on uplink pilots that are transmitted from u UEs, within a subset of OFDM subchannels that constitute a competition band.

We will discuss various design aspects for Kalman-based schemes that produce MMSE estimates of the complex frequency-domain channel gains. Compared to Wiener filtering [30], [83], Kalman estimators provide better initial transient accuracy and are the optimal estimators for linear signal models and Gaussian noise. They make optimal use of the previous time history of measured fading channels, using in effect a time window that grows with decreasing vehicle velocity (increasing coherence time of the channel). In contrast, Wiener-based OFDM channel estimators are mostly implemented using the pilots within a time-frequency window of fixed size. Kalman predictors also provide the prediction covariances. This uncertainty information can be used by the scheduler and by the link adaptation, to e.g. attain a target bit error rate at a given prediction variance [64].

The Kalman predictors can be implemented in the time domain, by tracking impulse response coefficients and then transforming them to predicted frequency domain channels. Alternatively, they can be implemented in the frequency domain. Details regarding these different kind of modelling is found in Chapter 4. The performances of both these schemes have been evaluated, and are identical. The computational complexity of the two schemes will however differ in general, see Sections 4.3 and 4.4. We here show the results from the frequency domain modelling implementation.

Initial results on Kalman-based predictors for FDD uplinks were previously reported in [84] and [66]. We will here discuss the effects of using different kinds of pilot sequences and examine the impact of channel fading statistics. The computational complexity, which determines the feasibility in fast adaptation feedback loops, will also be discussed.

6.2 System model

Prediction performance will be evaluated with respect to the baseline system design of the WINNER FDD mode [85]. This design has a system sampling

³When using proportional fair scheduling, the multiuser scheduling gain scales as $\log(\log(u))$ with the number of active users u .

period of 12.5 ns, giving a FFT bandwidth of 80 MHz. The signal bands are 45 MHz in both uplinks and downlinks. Each OFDM symbol is 2048 samples plus an additional 256 samples for the cyclic prefix. The subchannel width is 39.06 kHz and the OFDM symbol + guard duration is 28.8 μ s.

The time-frequency radio resource is divided into blocks (here denoted *chunks*) of 8 subchannels (312.5 kHz) by 12 OFDM symbols (345.6 μ s). A time duration of a chunk, 345.6 μ s, is denoted a slot. A chunk constitutes the unit for frequency-adaptive resource allocation. The chunk size is selected to make the channel moderately flat within chunks. Uplink pilot symbols known to the receiver facilitate the prediction. They are here assumed located on the first of the 12 OFDM symbols and no payload data is transmitted on this symbol, meaning that the pilot overhead is 1/12. We here assume a full-duplex FDD uplink, so uplink pilots will be transmitted within each slot.

To prepare for frequency adaptive uplink transmission, the UE is allocated a competition band and begins to send pilots in this band. Based on these pilots, the predictor estimates the parameters of an autoregressive model of a predetermined order k (AR k) so that it well describes the temporal correlation (Doppler spectrum) of the channel. The frequency correlation of the fading is also estimated. The model parameters (the AR coefficients and the frequency correlation) need to be estimated on exclusive pilots transmitted from a UE at a time, see Chapter 8. However, as will be demonstrated in Chapter 8, model parameter estimation only has to take place on a slow time scale (order of seconds). In this chapter, we make the assumption that AR parameters have been found that fully comply with the behaviour of the channels.

Channel predictions are then produced for this UE's channel. When a packet for uplink transmission arrives, the UE sends a transmission request during slot j . The scheduler may grant the request and sends the allocation information over a downlink control channel during slot $j + 1$. The transmission then commences over the uplink in slot $j + 2$. The required prediction horizon is two slots, or 0.7 ms, or $L = 2$ pilot samples. This tight control loop requires an update of the channel prediction by the end of slot j , that is based on pilots up until those transmitted at the beginning of slot j . The prediction, the scheduling, and the downlink control transmission together have to be executed within less than 1.5 slot durations (0.5 ms).⁴

We construct a linear filter that uses measurements of w parallel pilot-

⁴The feasibility of such a low control loop latency was studied within the WINNER project, to investigate the limits of performance of adaptive OFDMA transmission. The control of scheduled transmissions in systems that work according to the 3GPP LTE Release 8 standard is much longer, 5-6 ms. The attainable vehicular velocities in LTE would be correspondingly lower, at equal carrier wavelengths.

bearing subchannels. Details regarding this can be found in Chapter 4. Here we present a quick overview of the modelling technique. The fading channel coefficients of u simultaneous UEs⁵ are modelled by (see Section 4.2.4 in the case of subchannel modelling)

$$\begin{aligned}\mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t, \\ \mathbf{h}_t &= \mathbf{H}\mathbf{x}_t,\end{aligned}\tag{6.2.1}$$

where

$$\begin{aligned}\mathbf{F} &= \text{diag}_u(\text{diag}_w(\mathbf{X})), \\ \mathbf{G} &= \text{diag}_u(\text{diag}_w(\mathbf{Y})), \\ \mathbf{H} &= \text{diag}_u(\text{diag}_w(\mathbf{Z})).\end{aligned}\tag{6.2.2}$$

Here, $\text{diag}_n(\cdot)$ denotes a block-diagonal matrix with n blocks, and \mathbf{h}_t is a vector holding the uw fading channel coefficients. Each triplet $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ models the fading statistics of one channel coefficient with an autoregressive model of order 4. The four poles of this model are so chosen as to represent an estimated Doppler spectrum. They are in general different for different users. Here we use the same model for all channel coefficients across all terminals. The shape of the Doppler spectrum depends on the fading environment. We here assume that the Doppler spectrum describes the fading behaviour of the channel perfectly.

The pilots $\{\phi_{1,k,t}, \dots, \phi_{w,k,t}\}$ for each UE k at time t are represented by a time-variant regressor matrix Φ_t ,

$$\Phi_t = \begin{pmatrix} \phi_{1,1,t} & & & \phi_{1,u,t} & & \\ & \ddots & & & \ddots & \\ & & \phi_{w,1,t} & & & \\ & & & & & \phi_{w,u,t} \end{pmatrix}.\tag{6.2.3}$$

The measurement vector \mathbf{y}_t of the w received subchannels at the pilot sublocations is then modelled by

$$\mathbf{y}_t = \Phi_t \mathbf{h}_t + \mathbf{v}_t.\tag{6.2.4}$$

Here, \mathbf{v}_t represents noise and interference. Note that, unless we set most pilot symbols to zero, the received signal at a subchannel will be affected by the channels from multiple UEs.

⁵Channels for multiple transmit antennas/spatial streams from one UE may be modelled in the same way as channels from different UEs are modelled, with the difference that in that case, it may suffice to estimate one set of model parameters, since all streams from one and the same UE will feature the same channel statistics.

The correlation between the channel coefficients is expressed by the process noise covariance matrix $\mathbf{Q} = \|\mathbf{u}_t\|^2$, while $\mathbf{R} = \|\mathbf{v}_t\|^2$ is the noise covariance matrix. Here, we will assume that the noise has zero-mean and unit variance on each subchannel and is uncorrelated between subchannels, so that $\mathbf{R} = \mathbf{I}$. Theorem 4.2.1 describes how to appropriately scale the matrix \mathbf{Q} .

With the state and measurement equations (6.2.1) and (6.2.4), optimal inferences about the channel coefficients \mathbf{h}_t are obtained by the Kalman equations, see Section 4.3. The Kalman equations produce *filtered state estimates* $\hat{\mathbf{x}}_{t|t}$ based on all measurements up to and including time t . The L -step prediction estimate of \mathbf{h}_t is then obtained from $\hat{\mathbf{x}}_{t|t}$ by $\hat{\mathbf{h}}_{t+L|t} = \mathbf{H}\mathbf{F}^L\hat{\mathbf{x}}_{t|t}$.

The attainable prediction accuracy will depend on the prediction horizon scaled in carrier wavelengths, l , which in turn depends on the UE velocity v , the prediction horizon Lt_p expressed in seconds, where t_p is the sampling period of the filter, and the carrier wavelength λ , via the relation

$$l = vLt_p/\lambda. \quad (6.2.5)$$

In the assumed WINNER baseline design, the sampling period t_p equals the slot duration (345.6 μs), $\lambda = 8.1$ cm (3.7 GHz uplink carrier), and the prediction range is $L = 2$ steps.

The results in this chapter are evaluated on two channel models: A flat (non-frequency selective) channel, and a frequency selective non-line-of sight channel for urban environments (WINNER II C2 NLOS channel) with power delay profile given in Section 2.3.2.

When not explicitly stated otherwise, we set the velocity of the terminals to 50 km/h, the SNR to 12 dB, and the filter width w to 8 subchannels (one chunk width). The prediction horizon is set to two steps (slots). Performance is expressed in terms of the mean value of the normalized mean square error (NMSE), see Equation (4.3.14), averaged over all uw channel coefficients. These coefficients correspond to pilot sub-locations which are irrelevant for payload data detection. When the payload is to be detected, channel estimates need to be interpolated (and possibly extrapolated) to payload sub-locations. It can be shown that channel estimates acquired for these sub-locations have approximately the same quality as the estimates produced for the pilot sub-locations, see Appendix 6.A.1. Motivated by this fact, we use NMSE values for pilot sub-locations as a performance metric in this study.

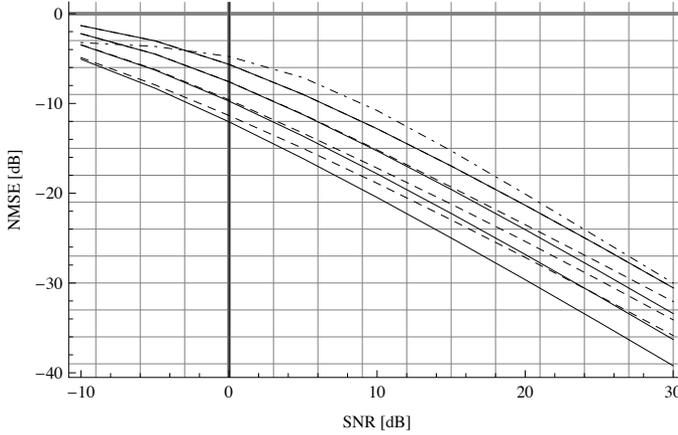


Figure 6.2: NMSE versus SNR for one UE tracking w parallel subchannels, where $w = \{1, 2, 4, 8\}$ (upper to lower curves). The maximum normalized Doppler frequency is $f_D t_p = 0.06$. Solid lines represent a flat fading channel. Dashed lines represent the frequency selective WINNER II C2 NLOS channel. Gridlines are spaced by 3 dB to elucidate the fact that a doubling of the filter width leads to a 3 dB performance increase for flat fading channels. The dash-dotted line indicates the filter estimation performance boundary (5.3.10) from Chapter 5.

6.3 Channel prediction for multiple users

In this section we discuss how to predict multiple fading channels based on a noisy superposition of measurements of the channels. As a prerequisite, we begin by studying some basic properties of an optimal filter taking parallel subchannels into account.

6.3.1 Filter width and Doppler frequency

We study how the choice of filter width w affects estimation performance. Since the noise is white by assumption, one expects that for flat fading channels, a doubling of the filter width would lead to a NMSE decrease of approximately 3 dB. This is indeed the result as seen in Figure 6.2, which illustrates how the NMSE performance for channel estimation (note that no prediction is used here) changes when the filter width is doubled. We use a Doppler spectrum that is approximately flat in the frequency range $[-f_D, f_D]$, where f_D is the maximum Doppler frequency. Results for both flat fading and frequency selective channels are presented. Evidently, the doubling gain is somewhat less than 3 dB for frequency selective channels. The general rule is that the denser the pilots are packed frequency-wise, and the less frequency selective the channel is, the closer to 3 dB will the doubling gain be.

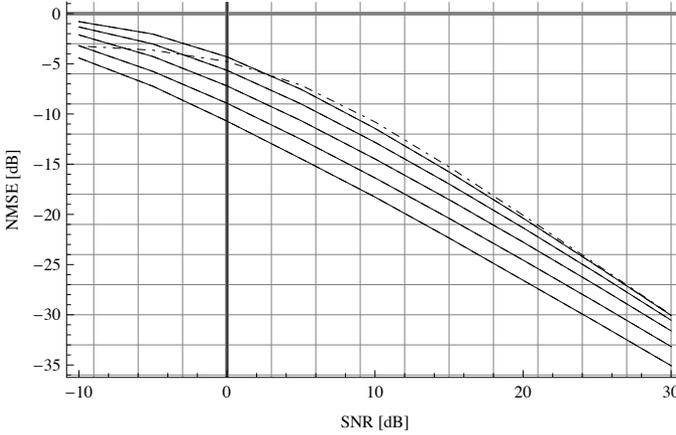


Figure 6.3: NMSE versus SNR for one UE tracking a single subchannel subjected to a fading channel with normalized Doppler frequency $f_D t_p = \{0.0075, 0.015, 0.03, 0.06, 0.12\}$, which correspond approximately to velocities $\{6, 13, 25, 50, 100\}$ km/h for the WINNER system parameters. For completeness, the dash-dotted line indicates the filter estimation performance boundary (5.3.10) from Chapter 5.

A similar result is achieved for the time domain when the normalized Doppler frequency changes. See Figure 6.3, which illustrates how the NMSE performance for channel estimation improves when the Doppler frequency is halved. Analogous to the doubling of filter width, we see a performance increase of a little less than 3 dB when the Doppler frequency is halved. What this says is that a low velocity and/or a low carrier frequency will be beneficial to channel estimation performance. In the same way will a dense pilot spacing influence estimation performance in a positive way. Note that the smaller the Doppler frequency, the closer to 3 dB will the performance increase be when the Doppler frequency is halved.

The velocity and the carrier frequency are parameters that a communications system engineer rarely can change at will. The duration between pilot bearing OFDM symbols, t_p , is however a design parameter whose value should be taken into careful consideration when designing a system, as should the filter width w . Note however that the choices of t_p and w will influence the numerical complexity of the Kalman filter (KF). The complexity of the KF increases linearly with t_p but cubically with w .

6.4 Pilot strategy

The filter width w , i.e. the number of simultaneous subchannels to be tracked, which is the same as the dimensionality of the measurements \mathbf{y}_t , may be adjusted for performance/complexity tradeoff. A competition band that comprises c predicted subchannels will then require the use of c/w Kalman predictors run in parallel, assuming that c/w is an integer.

The proportion of sub-symbols that are pilots, i.e. the pilot overhead, is typically somewhere between $1/20$ and $1/10$ for each transmit antenna, see for example Figure 2.5. In a multi-user setting, if each UE were assigned exclusive pilots, the grid of sub-symbols could be dominated by pilots already when as few as five or six UEs share the bandwidth. This is unacceptable. We therefore allocate a small subset of sub-locations for pilots, irrespective of how many UEs are taking part in the competition for the channel. If multiple UEs are present, then they somehow have to share the space allotted to pilots.

The pilot matrix $\{\Phi_t\}$ is the design parameter that tells how this should be done. Should the pilot sub-symbols transmitted by each UE be placed on all w subchannels that are tracked, hence making the pilots from the different UEs *overlap*? Or should we instead use *dedicated* pilots, so that each UE concentrates its pilot energy to one single subchannel, not transmitting anything on the remaining $w - 1$ subchannels?

Assuming that the number of UEs u in the competition band is less or equal to the number of subchannels w , we may represent the pilots used by w simultaneous UEs by a $w \times u$ matrix Ψ , where each column in Ψ contains the complex-valued time-frequency pilots for one UE:

$$\Psi = \begin{pmatrix} \phi_{1,1} & \cdots & \phi_{1,u} \\ \vdots & & \vdots \\ \phi_{w,1} & \cdots & \phi_{w,u} \end{pmatrix}, \quad (6.4.1)$$

where $\phi_{i,j}$ is the pilot symbol transmitted on subchannel i for UE j . For brevity, we have excluded time indices. The matrix Ψ is a “compacted” version of the pilot matrix Φ (6.2.3). *Dedicated* pilots for w UEs are here simply obtained through $\Psi = \sqrt{w}\mathbf{I}_w$, where \mathbf{I} denotes the identity matrix. *Overlapping* pilots are in this chapter constructed from Hadamard matrices.

Although complex Hadamard matrices are possible to find, they have no advantage over real matrices in the present context. We will here use Sylvester’s construction which yields pilot symbols of either -1 or 1 , i.e. BPSK symbols. Hence we construct a $2^n \times 2^n$ Hadamard matrix by setting $\mathbf{U}_0 = 1$ and iterating

$$\mathbf{U}_{n+1} = \begin{pmatrix} \mathbf{U}_n & \mathbf{U}_n \\ \mathbf{U}_n & -\mathbf{U}_n \end{pmatrix}. \quad (6.4.2)$$

To get w overlapping pilot sequences, we set $\Psi = \mathbf{U}_{\log_2 w}$. Note that in the dedicated pilots case and the overlapping pilots case alike, we choose pilot symbols so that the total pilot energy per user over the filter width is w .

With pilot sequences of length w , it is not possible to construct more than w real or complex-valued orthogonal pilot sequences. If the number of UEs is larger than the filter width w , we therefore need to construct additional non-orthogonal pilot sequences from the orthogonal set Ψ . There is no general scheme for how to do this optimally. In this chapter, we construct 16 pilot sequences from 8 pilots through a linear mapping. The 16 pilot sequences are here given by the columns of $\Psi\mathbf{T}$, where the matrix \mathbf{T} that maps 8 orthogonal pilots onto 16 non-orthogonal pilots is

$$\mathbf{T} = \begin{pmatrix} 1 & \cdot & \alpha & \frac{1}{2} \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \cdot & \frac{1}{2} & \cdot & \cdot & \frac{1}{2} & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \frac{1}{2} & \cdot & \cdot & \frac{1}{2} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \alpha & \cdot & \cdot & \frac{1}{2} & \frac{1}{2} & \cdot & \cdot & \cdot & \frac{1}{2} \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \alpha & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \alpha & \cdot & \frac{1}{2} & \cdot & \cdot & \cdot & \frac{1}{2} & \cdot & \frac{1}{2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \alpha & \frac{1}{2} & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \\ \cdot & 1 & \alpha & \cdot & \cdot & \frac{1}{2} & \cdot & \frac{1}{2} & \cdot & \frac{1}{2} & \cdot \end{pmatrix}, \quad (6.4.3)$$

where (\cdot) indicates a zero element, and $\alpha = 1/\sqrt{8}$.

6.4.1 Overlapping versus dedicated pilots

In a general FDD uplink wireless scenario, the channels between UEs and a base station will be frequency selective. The base station has to estimate these channels to good precision in order to be able to schedule resources efficiently. On the one hand, the use of overlapping pilots will provide the base station with information of the entire filter bandwidth for all UEs simultaneously. On the other hand, the fact that UEs send pilots at the same sub-locations will degrade performance as compared to the scenario where each UE has dedicated sub-locations for their pilots. Conversely, the use of dedicated pilots will give poor information about how the channel varies over different subchannels.

We compare the performance of overlapping pilots against the performance of dedicated pilots. Figures 6.4 and 6.5 show NMSE versus number of UEs u for the respective choices of pilots. The filter width is set to $w = 8$, and the Doppler spectrum is the flat Doppler spectrum used in Section 6.3.1. Here we use a normalized maximum Doppler frequency of 0.06, corresponding e.g. to a velocity of 50 km/h at a carrier frequency of 3.7 GHz when the WINNER system parameters are used. Results for both flat-fading and frequency selective channels are presented.

First, we turn our attention to the case $u \leq w$. In the flat-fading case, all subchannels fade in unison and the pilots for UEs 1 through 8 are completely orthogonal. In a noise-free case, the w measurements provided at time t by equation (6.2.4) would then provide a solvable linear system of equations with respect to the $u \leq w$ different channel coefficients. This holds regardless of whether we use overlapping or dedicated pilots. For example, when $u = w$, we would have $\mathbf{y}_t = \Psi \bar{\mathbf{h}}_t$, where $\bar{\mathbf{h}}_t$ is a w -vector whose i :th element is the (single) channel coefficient for user equipment i . Since Ψ has full column rank when $u \leq w$, the system can be solved with respect to $\bar{\mathbf{h}}_t$.

When the channels are flat fading we therefore have the result that

- the choice of pilots is irrelevant as long as the pilots are orthogonal, and
- the performance does not degrade with an increasing number of UEs u as long as $u \leq w$.

The situation is different when the channels are frequency selective, i.e. when the WINNER II C2 NLOS channel model is used and the frequency spacing between pilot subchannels is 39.06 kHz. The importance of measuring over the entire filter bandwidth is evident when we compare the dashed lines in Figures 6.4 and 6.5 for UEs 1–8 and find the overlapping pilots to perform considerably better than the dedicated pilots. For the particular working point SNR=12 dB studied here, the gain is about 2 dB for one UE, and decreases when the number of simultaneous UEs increases.

The curves merge at the point $u = 8$, indicating that the choice of pilots is unimportant when the orthogonal set has been filled. This conclusion should however be drawn with care, because in the dedicated case, the NMSE values will vary considerably, from high (on peripheral subchannels), to low (on middle subchannels). When adaptive modulation is used, the modulation format is assumed to be chosen per resource block. One would therefore prefer as even a distribution of the NMSE values as possible, which is provided by the overlapping pilots.

The reason for the performances for dedicated pilots (dashed) actually *increasing* with u in Figure 6.4 is due to the way the pilot subchannels have been allocated to UEs in this experiment. The first UE here puts pilot energy on the first subchannel, which is on the border of the filter bandwidth, while UEs 4 and 5 invest their pilots in the middle of the bandwidth. The latter is the better tactic when we rate performance based on the mean value of the NMSE over all subchannels. This is the reason for the performance increase when UEs 2, 3 and so on are added to the system.

This illustrates that if dedicated (and time static) pilots are to be used, then one should assign one of the middle subchannels to the first UE to enter the system, and only assign border subchannels when necessary.

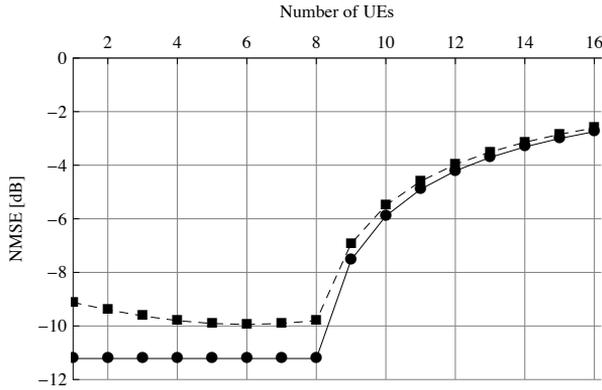


Figure 6.4: NMSE performance, averaged over the eight subchannels, versus number of UEs for two-step prediction at an SNR of 10 dB. Results for flat-fading (solid, circles) and frequency selective (dashed, squares) are presented. *Dedicated* and time-static pilots were used. The normalized Doppler frequency is 0.06, so that two-step prediction corresponds to 0.12 wavelengths, and the filter width is $w = 8$. The performance results are the same for all UEs.

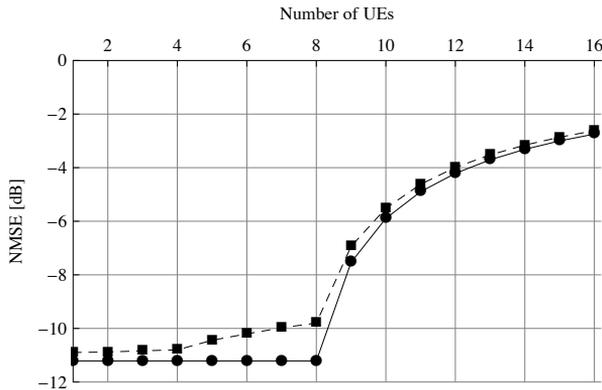


Figure 6.5: NMSE performance, averaged over the eight subchannels, versus number of UEs for two-step prediction at an SNR of 10 dB. Results for flat-fading (solid, circles) and frequency selective (dashed, squares) are presented. *Overlapping* and time-static pilots were used. The normalized Doppler frequency is 0.06, so that two-step prediction corresponds to 0.12 wavelengths, and the filter width is $w = 8$. The performance results are the same for all UEs.

When the number of UEs u to share a certain bandwidth is larger than the corresponding filter width w , it is not possible to find a set of u orthogonal pilots. As presented in Section 6.4, we then construct new pilots from the original set of w orthogonal pilots by a linear mapping. Restudying Figures 6.4 and 6.5 but now turning our attention to the case $u > w$, we note two facts. One is that the performance is unaffected by the choice of dedicated or overlapping pilots. The other fact is that the performance drop when we go from orthogonal to non-orthogonal pilots ($u = 8$ to $u = 9$) is considerable. It can be shown (see Appendix 6.A.2) that when the number of UEs u is less than or equal to the filter width w and all UEs have the same SNR γ and independent fading channels, then the filter estimation NMSE per UE, $\xi_{u \leq w}$, is

$$\xi_{u \leq w} = (1 + w\gamma)^{-1}. \quad (6.4.4)$$

When the number of UEs u is $w + 1$, where w is the filter width and all UEs have the same SNR γ and independent channels, the filter estimation NMSE per UE, $\xi_{u=w+1}$, is

$$\xi_{u=w+1} = \frac{1 + \gamma}{1 + (w + 1)\gamma}. \quad (6.4.5)$$

6.4.2 Time-varying pilots

The performance drop from (6.4.4) to (6.4.5) as $u \geq w$ can be avoided by providing the filter with more information about the time variability of the channels. In the case $u \leq w$, we have seen that the prediction performance increases if we spread out the pilot energy and let the pilots vary over the filter's frequency band. In the same manner we may design the pilots to make optimal use of previous channel samples. In the case of noiseless, *frequency-selective but time-invariant* channels (i.e. immobile terminals), we obtain a linear system of equations from the most recent w measurements

$$\begin{pmatrix} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-w+1} \end{pmatrix} = \underbrace{\begin{pmatrix} \Phi_{1,t} & \cdots & \Phi_{u,t} \\ \vdots & \vdots & \vdots \\ \Phi_{1,t-w+1} & \cdots & \Phi_{u,t-w+1} \end{pmatrix}}_{\mathbf{A}} \mathbf{h} \quad (6.4.6)$$

where $\Phi_{i,t}$ is the diagonal pilot matrix that holds the pilot sequence for UE i at time t , and \mathbf{h} holds the uw time-static channel coefficients. If the pilot sequences contained in the diagonal matrices $\{\Phi_{i,j}\}$ vary over time so that \mathbf{A} obtains full rank uw , then the system (6.4.6) becomes solvable with respect to the uw unknowns in \mathbf{h} . One would expect that time-varying pilots have the potential of improving the estimation also for time-varying channels and noisy measurements. Below, we study the utility of using time-varying pilots.

In Section 4.3.2, it was proved that when the channel models are static and the pilots vary periodically over time, then the KF will converge to a periodic filter. This property has a positive impact on the complexity of the predictor. We shall therefore consider periodic pilot sequences, which we denote *cyclic pilots*, and we here consider the specific case $w = 8$ and $u = 16$.

For *dedicated* pilots, cyclic pilots for UEs 1–8 are produced by rotating the original pilot matrix $\sqrt{8}\mathbf{I}_8$ one step to the left every time step, hence producing a period of eight time steps in the pilot cycle. The pilot sequences for UEs 9–16 are set as those for UEs 1–8, but with signs that alternate over time, so that the total 8-by-16 pilot matrix is

$$\Psi_t = \sqrt{8} \left[\text{rot}_{t \bmod 8} \{\mathbf{I}_8\} \quad (-1)^t \text{rot}_{t \bmod 8} \{\mathbf{I}_8\} \right], \quad (6.4.7)$$

where $\text{rot}_i \{\mathbf{A}\}$ is a cyclic rotation of the columns of \mathbf{A} i steps to the left, and $j \bmod 8$ is the remainder of division of j by 8.

Time varying *overlapping* pilots are once again constructed from Hadamard matrices. With $u = 16$ UEs, a 16-by-16 Hadamard matrix is constructed according to (6.4.2). With $w = 8$, the rows of this matrix are then partitioned into two sections. The first eight rows give the 16 pilot sequences for the first time step in the pilot cycle, and the last eight rows describe the second and last time step in the cycle. For other values of u and/or w , the matrix would be partitioned into u/w parts (assuming u/w is an integer), so that the pilot cycle has a period of u/w . Expression (6.4.8) shows the partitioning of a 16-by-16 Hadamard matrix into a period of 2. Note that this method of producing overlapping cyclic pilots results in the first w UEs actually having *static* pilots.

$$\left(\begin{array}{l} t = 0, 2, \dots \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ t = 1, 3, \dots \end{array} \left\{ \begin{array}{l} 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \\ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \\ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \\ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \\ 1 \ 1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ 1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \\ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \\ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \\ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \\ 1 \ 1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ 1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \\ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \\ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \\ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \\ 1 \ 1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ 1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \\ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \\ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \\ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \end{array} \right. \right) \quad (6.4.8)$$

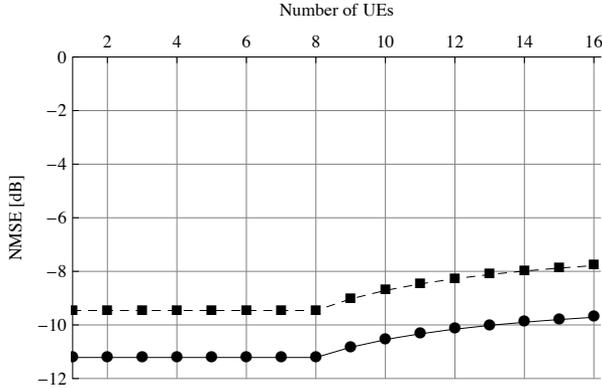


Figure 6.6: NMSE performance, averaged over the eight subchannels, versus number of UEs for two-step prediction at an SNR of 10 dB. Results for flat-fading (solid, circles) and frequency selective (dashed, squares) are presented. *Dedicated* pilots that vary over time with period 8 where used. The normalized Doppler frequency is 0.06, so that two-step prediction corresponds to 0.12 wavelengths, and the pilot cycle period corresponds to 0.48 wavelengths. The filter width is $w = 8$. The performance results are the same for all UEs.

The impact of using cyclic pilots is studied in Figures 6.6 and 6.7. Again, we use the flat Doppler spectrum from Section 6.3.1, with a normalized maximum Doppler frequency of 0.06. Results for both flat fading and frequency selective channels (WINNER II C2 NLOS) are presented. When the number of simultaneous UEs u is less or equal to eight, the only improvement is an averaging of the performance for the dedicated pilots as compared to the case when static pilot positions were used (Figure 6.4).

When $u \geq 9$ the improvement is more dramatic. The steep performance drop at $u = 9$ is now gone. Whereas the performance for flat fading channels dropped by about 8 dB from $u = 8$ to $u = 16$ when static pilots were used, cyclic pilots reduce this decrease in prediction performance to a mere 2 dB or less. Evidently, the use of cyclic pilots is highly important to maintain a high estimation performance when the number of UEs competing for a frequency band is larger than the bandwidth w .

The number of orthogonal cyclic pilot sequences that can be used is ultimately dictated by the size of a time-frequency region over which the channel is essentially static. As stated in Section 2.3.1, this size is usually a few thousand sub-locations and is approximately given by $(T_m B_d)^{-1}$, where T_m is the channel's multipath spread and B_d is the Doppler spread of the channel. Even if only one part in twenty of all sub-locations are allotted to pilots, and even if T_m and B_d are high, cyclic pilots can provide orthogonal pilot

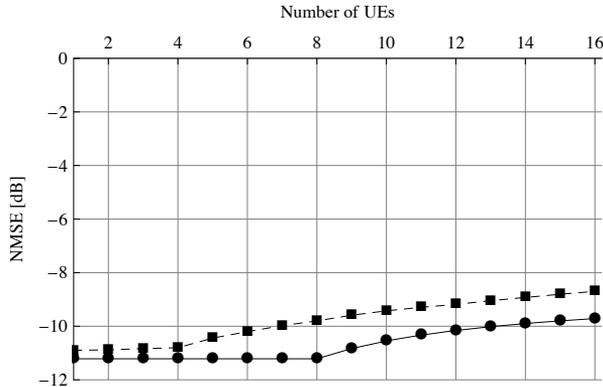


Figure 6.7: NMSE performance, averaged over the eight subchannels, versus number of UEs for two-step prediction at an SNR of 10 dB. Results for flat-fading (solid, circles) and frequency selective (dashed, squares) are presented. *Overlapping* pilots that vary over time with period 2 were used. The normalized Doppler frequency is 0.06, so that two-step prediction as well as the pilot cycle period correspond to 0.12 wavelengths. The filter width is $w = 8$. The performance results are the same for all UEs.

sequences for multiple UEs in most situations.

Note that (6.4.5) can be used as an upper bound on performance when the number of competing UEs is larger than what cyclic pilots can support with preserved orthogonality between UEs. If the time-frequency region over which the channel is essentially static holds five pilot sub-symbols, but six equally strong UEs compete, then their respective filter estimate NMSE performances are upper bounded by (6.4.5) with $w = 5$.

6.5 The impact of Doppler spectra on channel prediction

So far in this chapter, we have assumed a prediction range of $L = 2$ time steps, corresponding to 0.7 ms, or 0.12 wavelengths at 50 km/h and 3.7 GHz carrier frequency. We have also assumed that the temporal fading behaviour of the channel is described by a flat Doppler spectrum. This makes channel prediction inherently hard: Since the Doppler spectrum does not exhibit any strong frequency components, the fading channel data series do not contain distinct sinusoids that can easily be extrapolated into the future. We now turn to study how the prediction range and the Doppler spectrum affect prediction performance.

6.5.1 Criterion

It was found in [63] and [64] that when the prediction NMSE is below -10 dB, then link adaptation and/or scheduling can be successfully used for that resource. The performance degradation due to prediction errors is then relatively small. This result holds for uncoded systems. When small code blocks are used, it was found that somewhat higher prediction NMSEs can be accepted. In [82], a performance boundary of $0.15 \approx 8$ dB was used. When evaluating prediction performance, we will choose the lower limit and accept a maximum NMSE of -10 dB. Note that, unlike the performance boundary for channel estimation given in (5.3.10) in Chapter 5, the prediction performance bound does not depend on the SNR. Also, the limit -10 dB is an acceptable NMSE level *for link adaptation*. In other applications for channel prediction, e.g. to decide whether or not to use spatial multiplexing, this limit may not be appropriate.

6.5.2 Studies on different Doppler spectra

In Figures 6.8–6.10, we examine the prediction performance for a number of SNRs and channel properties. Four simultaneous UEs with equal average SNR and orthogonal overlapping and static pilot sequences were used. The filter width was $w = 4$. Results for three different Doppler spectra are given. In each figure, three pairs of curves are plotted. The pairs illustrate, in descending order of NMSE level, prediction performance when the SNR, the same for all UEs, is 6, 12, and 18 dB, respectively. Results are shown for flat fading channels (lower curve in each pair), and frequency selective channels (upper curves).

As is clear from the figures, favourable fading statistics are crucial for good prediction performance. We investigate four cases. The prediction range $L = 2$ set previously corresponds to $Lt_p v f_c / c_0 = 0.12$ wavelengths at velocity $v = 50$ km/h, pilot spacing $t_p = 345.6$ ms, and carrier frequency $f_c = 3.7$ GHz. With the flat Doppler spectrum (Figure 6.8) that has been used throughout the chapter so far, the performance criterion is just met at 12 dB but not at 6 dB. However, when the fading channel conforms to a Jakes-like Doppler spectrum (Figure 6.9), all SNR values meet the criterion at the prescribed prediction range.

For a flat Doppler spectrum, prediction at the -10 dB NMSE accuracy level is possible 0.08–0.15 wavelengths ahead. For a Jakes-like oscillatory channel, the attainable horizons are longer, 0.14–0.27 wavelengths, in the same SNR range 6–18 dB. These results can be compared to the case of just extrapolating the present (noisy) sample as prediction. The attainable horizon is then around 0.05 wavelengths, see e.g. Figure 3.7 in [66].

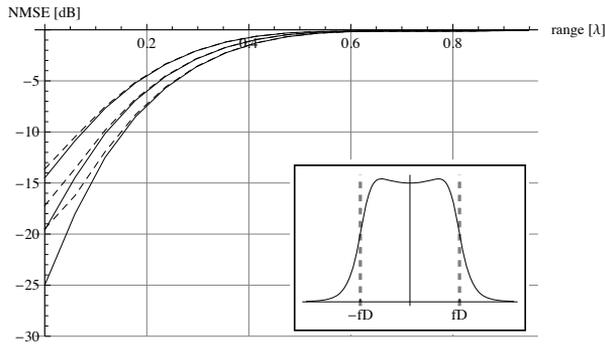


Figure 6.8: Prediction NMSE performance versus prediction range when the channel's Doppler spectrum is described by the inset picture, which is here given by a flat AR4 model. The maximum normalized Doppler frequency is ≈ 0.06 . The curves correspond, in pairs, to SNR levels of 6, 12, and 18 dB, respectively. In each pair, the dashed and solid curve shows the NMSE performance for frequency selective and flat fading channels, respectively.

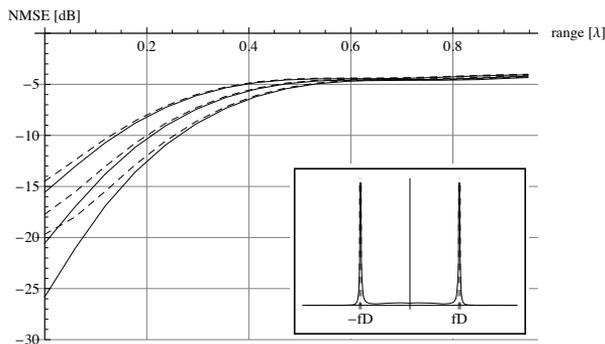


Figure 6.9: Prediction NMSE performance versus prediction range when the channel's Doppler spectrum is described by the inset picture, which is here given by a Jakes-like AR4 model. The maximum normalized Doppler frequency is ≈ 0.06 . The curves correspond, in pairs, to SNR levels of 6, 12, and 18 dB, respectively. In each pair, the dashed and solid curve shows the NMSE performance for frequency selective and flat fading channels, respectively.

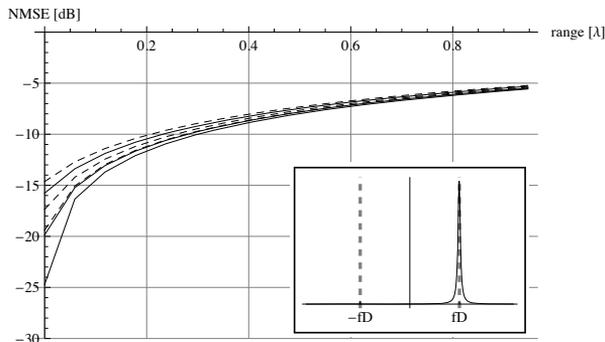


Figure 6.10: Prediction NMSE performance versus prediction range when the channel’s Doppler spectrum is described by the inset picture, which is here given by an AR1 model with its pole very close to the unit circle. The maximum normalized Doppler frequency is ≈ 0.06 . The curves correspond, in pairs, to SNR levels of 6, 12, and 18 dB, respectively. In each pair, the dashed and solid curve shows the NMSE performance for frequency selective and flat fading channels, respectively.

We also illustrate how the prediction performance is improved even further when the Doppler spectrum contains a single strong frequency component (Figure 6.10). This is the case when there is line-of-sight to the transmitter, or when a single strong reflector is present.

Studying Figures 6.8–6.10, it is clear that the SNR has less impact on prediction performance than the Doppler spectrum. Even less essential to prediction performance is the frequency selectivity of the channel; only at very small prediction ranges does it matter whether the channel is frequency selective or not. To conclude, the fading statistics is the most crucial component that dictates prediction performance. Accurate modelling of the fading behaviour is therefore very important.

It was shown in Section 6.4.2 that cyclic pilots can greatly improve prediction performance when the number of UEs u competing in a competition band exceeds the filter width w . When studying prediction performance as a function of prediction range, we could expect to see an improved prediction performance from using cyclic pilots also when $u \leq w$, but this turns out not to be the case; from studying Figures 6.8–6.10, we can see that when $u \leq w$, cyclic pilots have a very limited impact on prediction performance as the prediction range increases. Since here $u = w$, the prediction performance can never exceed that achieved for flat fading channels (with orthogonal pilots), even if the pilots are allowed to vary over time. As the prediction range increases, the gaps close between the curves for frequency selective channels and the corresponding curves for flat fading channel, implying that the best possible gain that could ever be achieved by using cyclic pilots is very lim-

ited for the case $u \leq w$. However, it should again be emphasized that the technique of using cyclic pilots is crucial for high UE loads, $u > w$.

6.6 Numerical complexity

For efficient link utilisation, several UEs will compete for the radio resource in a specific competition band. To cover a competition band that contains c predicted subchannels, $\lceil w_0/w \rceil$ KFs are run in parallel, where $\lceil x \rceil$ is x rounded to the nearest larger integer. Each of these filters has a numerical complexity given by (4.4.1). Assuming that one complex operation corresponds to four real operations, the solid lines in Figure 6.11 display the number of real operations required per filter update versus number of UEs for $w_0 = 160$ predicted subchannels, for designs with $w = 4$ or $w = 8$. We have then assumed that $u = 4$ UEs compete for the resources. For each of these UEs, the fading channel is modelled using subchannel modelling according to Chapter 4, with model order $k = 4$. The total number of states in the channels' state space model is $n = uwk$. The prediction range is $L = 2$ steps.

Expression (4.4.1) holds for general choices of pilot symbols, but the complexity may be decreased further by considering only dedicated pilots (as opposed to overlapping pilots). The measurement equation is then completely decoupled between different UEs, which makes the state estimation error covariance matrix $\mathbf{P}_{t|t}$ in the KF recursions block diagonal. This, in turn, means that we may run a separate KF for each UE without losing performance, which means that the complexity increases only linearly with the number of UEs. In that case, the complexity is easily computed by setting $n = wk$ instead of $n = uwk$ in the complexity expressions, and then multiplying the final result by u . The dashed lines of Figure 6.11 show the number of real operations required for one update for filter widths $w = 4$ and $w = 8$ when these decoupled KFs can be used.

While the numerical complexity is considerably lower when dedicated pilots are used than when the pilots are overlapping, we should point out, as we did in Section 6.4.1, that dedicated pilots may exhibit a substantial spread in NMSE values. When adaptive OFDMA is used, so that modulation formats are adapted to the instantaneous SNR in small resource blocks, it is preferable to have a constant NMSE over a resource block.

The WINNER baseline system would require a new prediction for each slot of duration 0.34 ms for vehicular UEs. To assess the feasibility of the required computational complexity, we here investigate the consequences of setting 10^{10} real operations per second as a target for feasibility for uplink predictors realised in the base station. This would correspond to a limit of $3.4 \cdot 10^6$ operations per update (0.345 ms). As is evident from Figure 6.11,

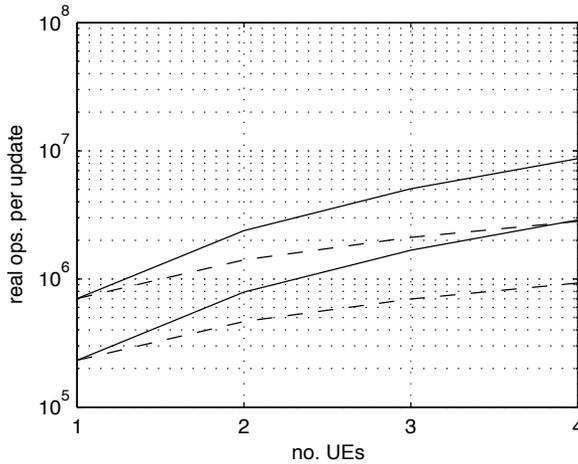


Figure 6.11: Complex arithmetic operations per update vs. number of UEs u for predicting a competition band containing 160 predicted subchannels, using either 40 KFs of filter width $w = 4$ (lower pair) or 20 KFs with width $w = 8$ (upper pair). Solid lines represent a general choice of pilots. Dashed lines represent the use of dedicated pilots and u decoupled KFs for each set of w subchannels.

using eight parallel subchannels would then be infeasible both for general and dedicated pilots, while four parallel subchannels is slightly above the limit for general pilots but falls well within our boundary for the decoupled case (dashed). The use of Kalman-based uplink prediction may therefore be feasible with a filter width of 4 under these assumptions.

Lower complexity targets would be realistic for predictors located in terminals. We return to that case in Section 8.7.

6.7 Discussion and conclusions

In this chapter, we investigated various choices of pilot design in the uplink of a multi-user multi-transmit-antenna OFDMA system dimensioned according to the WINNER radio concept. In this experiment, channels from u potential uplink users are predicted simultaneously by Kalman predictors that use w subchannels. It was stressed that if the pilots from different UEs are allowed to overlap, then the pilot overhead can be kept at a reasonable level while the frequency selectivity over the whole filter bandwidth may still be sampled for every UE.

It was found that prediction performance drops considerably when the number of UEs exceeds the number of simultaneously tracked subchannels

w . Expressions for the size of this performance drop was derived for a special case.

We found that the use of time varying pilots, here referred to as *cyclic pilots*, can remedy the performance restriction imposed by the filter width w . This is especially important in the downlink of CoMP systems, where many beams from multiple base stations have to be tracked simultaneously.

It should however be pointed out that overlapping pilots will cause logical correlations between the estimated/predicted channel coefficients, although there may be no physical correlation in the actual channel. If in a MIMO system the channel matrix \mathcal{H} has uncorrelated elements, then the estimation \mathcal{H}_{est} or the prediction \mathcal{H}_{pred} may still have correlated elements due to the logical coupling introduced by the overlapping pilots. If spatial multiplexing is to be used, the predicted channel capacity can be calculated from the eigenvalues of the matrix $\hat{\mathcal{H}}_{pred}\hat{\mathcal{H}}_{pred}^*$. However, if the elements of \mathcal{H}_{pred} are Gaussian but correlated, then to the best of the author's knowledge, no closed form expression for the joint probability density function for the eigenvalues of $\hat{\mathcal{H}}_{pred}\hat{\mathcal{H}}_{pred}^*$ exists.

When studying how prediction performance decrease with increased prediction range, it was found that the Doppler spectrum is the channel characteristic that mainly dictates prediction performance.

Although we have here investigated an OFDMA uplink, where many simultaneous UEs compete for the uplink resource, the problem formulation is trivially extendable to MIMO systems and to coordinated multi-point (CoMP) scenarios, where several base stations transmit simultaneously to one UE. However, in CoMP settings it is especially important to consider the case when the received signal strengths differ between the inputs. Building upon the framework described in this thesis, the CoMP setting has been considered in [86], and is under continued investigation at present.

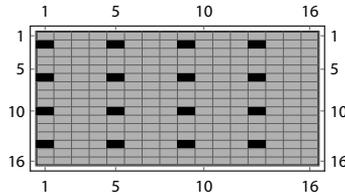


Figure 6.12: Pilot layout. Time runs horizontally and frequency runs vertically. Light gray indicates payload symbols. Dark gray indicates pilot symbols.

6.A Further studies

6.A.1 Channel estimate interpolation

When pilot symbol assisted modulation is used, as is the case here, the estimated channel at the pilot sub-locations inevitably needs to be interpolated (and possibly extrapolated) to payload sub-locations prior to channel equalization. This is typically done by e.g. fitting a quadratic surface over a region (in most cases a chunk) surrounding a few pilot sub-symbols. Compared to the NMSE obtained by the filter on pilot sub-locations, the interpolation process may lead to an increased NMSE. We here investigate whether this is the case.

We assume the WINNER II C2 NLOS channel with a flat Doppler spectrum (see Figure 5.2) at a maximum normalized Doppler frequency of 0.02. For example, the WINNER system parameters with a pilot spacing of 115.2 μs at a user terminal velocity of 50 km/h and a carrier frequency of 3.7 GHz gives rise to this Doppler frequency. Because of the frequency selectivity and the lack of spectral components in the Doppler spectrum, this choice of channel makes extrapolation reasonably “difficult”. subchannel spacing and OFDM symbol duration is 15 kHz and 28.8 μs , respectively. Pilot symbols are located at every 4:th subchannel and every 4:th OFDM symbol. See Figure 6.12.

The NMSE at payload sub-locations can be calculated by setting up an “augmented” state space by pretending that the payload sub-symbols in Figure 6.12 are pilots with value 0. The total filter bandwidth is then 16 and pilot sequences are time-varying with period 4. In setting up the augmented system, the Doppler frequency should be divided by 4, since the duration between “pilots” is a quarter of that in the original (pilot-spaced) system. By calculating NMSE values for the augmented system as given by an optimal Kalman filter (KF), we obtain the lowest NMSEs that can be achieved with any type of interpolation over measurements that were available to the filter.

Figure 6.13 shows NMSE performance for the filtered estimates (i.e. es-

imates using measurement up to the most recent time) for each of the 16 subchannels for a number of SNR values. For symmetry reasons, the curves are pairwise identical so that 9, not 16, curves are visible⁶. Clearly, they all have a period of 4. The thick, solid lines indicate the NMSE for the pilot symbol locations (which means that they are only defined for time indices 1,5,9, and 13). Again, because of symmetry, there are only two such lines visible. Note that the payload subchannel curves corresponding to subchannels 2,6,10, and 14 coincide with the pilot lines at every fourth time index, as should be.

The two uppermost curves, which correspond to subchannels that need to be *extrapolated* from pilots, have considerably higher NMSE than the rest. Extrapolation should therefore be avoided if possible, by using a pilot layout that have pilots located on its edges.

In the figure, dashed lines indicate the mean NMSE for the whole 16-by-16 grid of interpolated channel coefficients (upper line) and pilot symbols locations (lower line). The difference between the two ranges from a fraction of a decibel at low SNRs up to about 1.5 decibels for high SNRs.

Whether the mean NMSE is a valid performance measure is a matter of discussion. Channel encoding is a way of balancing out spots of high NMSEs and/or SNRs, so if a channel code is applied to the payload symbols in the grid, then the use of the mean NMSE as performance measure is justified. On the other hand, if no encoding is used, then the weakest link – the payload symbol with the highest NMSE – will dictate the overall performance. We see from the figure that the highest NMSE values, if we exclude the two highest curves (extrapolation), peak about one decibel up from the payload mean NMSE, and hence is a performance degradation from the pilot NMSE of between 1 and 2.5 dB.

However, if channel interpolation is carried out by fitting e.g. a quadratic surface over the estimated channel at pilot locations, then in effect one has *smoothed* the channel estimates. That is, future data is used for the estimation of channel coefficients with sub-locations early in the grid. Despite that this kind of smoothing is by no means optimal, one may expect that the overall smoothed estimation performance on the whole grid is actually better than the filtered estimates on pilot locations. To analyze whether this is true, we reiterate the experiment from Figure 6.13, but this time smoothing is used, so that all 16 pilot measurements are used when estimating each and every location in the grid. See Figure 6.14.

We observe that the overall NMSE is generally considerably better than the filtered estimates at pilot positions. Analogous to the case where extrapo-

⁶The asymmetric pilot sequence is the reason why 9 rather than 8 curves show in the plots; subchannels 1 and 15 have identical NMSEs, as do subchannels 2 and 14, and so on. subchannel 16, however, takes a unique position in terms of bad NMSE performance.

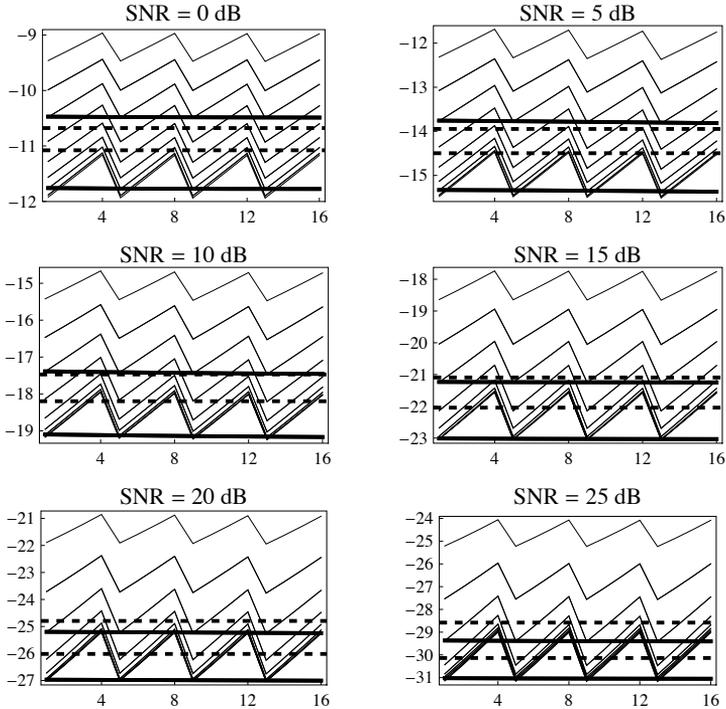


Figure 6.13: Filtering NMSE versus time index for payload subchannels (thin, solid) and pilot locations (thick, solid). Dashed lines indicate mean NMSE for payload and pilot positions.

lation is avoided by placing pilots at the edges of the pilot layout, we may here avoid the last three time indices, where smoothing performance coincide with filtering performance, by truncating the grid at time index 13 where the last pilots are located. If this is done, smoothing performance is always as good or better than pilot filtering performance. Even if sub-optimal smoothing is used, we therefore find that the filtering performance at the pilot positions should be a good representative for the final interpolated channel estimation performance.

6.A.2 Pilot design

The NMSE performance of a channel estimator/predictor will depend on the pilot matrix Φ_t in (6.2.4), but finding an optimal choice of Φ_t is difficult. The solution may vary depending on whether we look at channel estimation or channel prediction. In this section we will attempt to optimize the pilots for channel *estimation* when the channel is static and the filter is stationary. As

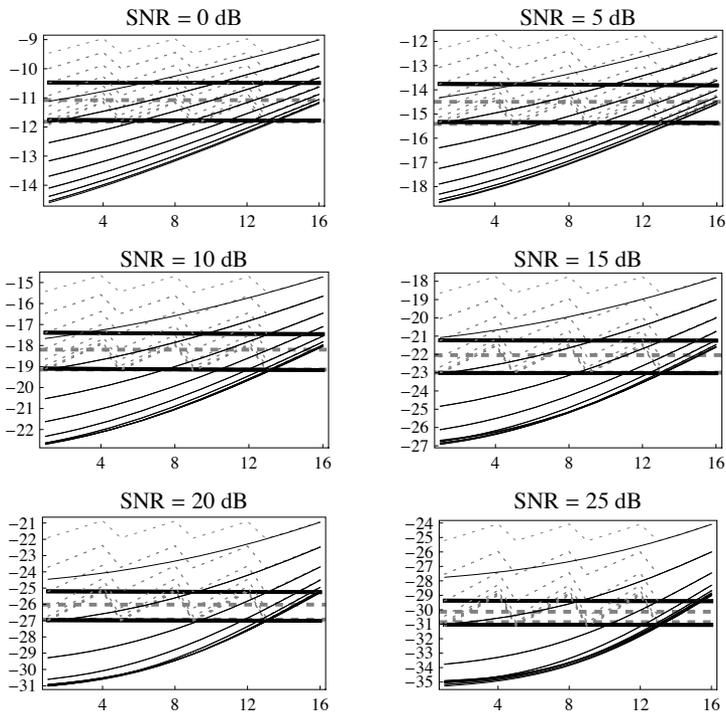


Figure 6.14: Smoothing NMSE versus time index for payload subchannels (thin, solid) and pilot locations (thick, solid). Thick dashed lines indicate mean NMSE for payload and pilot positions. For comparison, the result from Figure 6.13 has been plotted as thin dashed lines.

criterion for optimality we use the minimization of the total NMSE. Hence we seek to find (see Equation (4.3.14))

$$\arg_{\Phi_t} \min \text{tr}((\mathbf{H}\bar{\mathbf{P}}_f\mathbf{H}^*) \oslash \mathbf{R}_h), \quad (6.A.1)$$

where $\bar{\mathbf{P}}_f = \mathbf{P}_{t|t}$ (for all t) is the covariance matrix for the filtered state estimation error $\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}$, and \mathbf{R}_h is the covariance matrix for the channel coefficients \mathbf{h}_t . The regressor matrix \mathbf{H} is given by (6.2.1). From Table 4.1, we see that $\bar{\mathbf{P}}_f$ satisfies

$$\bar{\mathbf{P}}_f = \bar{\mathbf{P}}_p - \bar{\mathbf{P}}_p\mathbf{H}^*\Phi_t^*\bar{\mathbf{R}}_e^{-1}\Phi_t\mathbf{H}\bar{\mathbf{P}}_p, \quad (6.A.2)$$

with

$$\bar{\mathbf{P}}_p = \mathbf{F}\bar{\mathbf{P}}_f\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* \quad \text{and} \quad \bar{\mathbf{R}}_e = \Phi_t\mathbf{H}\bar{\mathbf{P}}_p\mathbf{H}^*\Phi_t^* + \mathbf{R}, \quad (6.A.3)$$

where $\bar{\mathbf{P}}_p$ is the error covariance for the one-step state predictions, and \mathbf{F} , \mathbf{G} , \mathbf{Q} , and \mathbf{R} are model matrices defined in Section 6.2. Unfortunately, it is not possible to find an explicit formula for $\bar{\mathbf{P}}_f$, which makes the optimization problem (6.A.1) very hard to solve, if not impossible. We therefore consider a special case defined by two assumptions:

1. We consider u independent UEs with *flat fading* channels.
2. We consider a filter without memory, so that $\bar{\mathbf{P}}_p = \Pi_0$, where $\Pi_0 = \|\mathbf{x}_t\|^2$ is the prior covariance for the state vector chosen in such a way that the model is stationary (see Theorem 4.2.1).

Assumption 1 means that, for each UE, all w elements in the vector of channel coefficients, \mathbf{h}_t , are identical. We can therefore rewrite the measurement equation (6.2.4) as

$$\mathbf{y}_t = \bar{\Psi}\bar{\mathbf{h}}_t + \mathbf{v}_t = \bar{\Psi}\bar{\mathbf{H}}\mathbf{x}_t + \mathbf{v}_t, \quad (6.A.4)$$

where $\bar{\mathbf{h}}_t$ is a w -vector whose i :th element is the (single) channel coefficient for UE i . Here, $\bar{\mathbf{H}} = \mathbf{S}\mathbf{H}$, where \mathbf{S} is a $u \times wu$ matrix with exactly one 1 in each row and column and all other elements 0. The matrix \mathbf{S} is constructed so that it picks out one of the w identical channel coefficients from each UE. The $w \times u$ matrix $\bar{\Psi}_t$ is defined in (6.4.1). Column i in $\bar{\Psi}$ is the pilot sequence used by UE i . Since we here consider a static model and a filter without memory, there is no point in having time varying pilots, which is why we have omitted the time index for $\bar{\Psi}$ in (6.A.4). We also assume normalized pilots, so that each pilot sequence of w symbols has energy w . This means that the diagonal elements of $\bar{\Psi}^*\bar{\Psi}$ are identically w .

Multiplying (6.A.2) from the left and the right by $\bar{\mathbf{H}}$ and $\bar{\mathbf{H}}^*$, respectively, we get

$$\bar{\mathbf{H}}\bar{\mathbf{P}}_f\bar{\mathbf{H}}^* = \bar{\mathbf{H}}\bar{\mathbf{P}}_p\bar{\mathbf{H}}^* - \bar{\mathbf{H}}\bar{\mathbf{P}}_p\bar{\mathbf{H}}^*\bar{\Psi}^*(\bar{\Psi}\mathbf{R}_h\bar{\Psi} + \mathbf{R})^{-1}\bar{\Psi}\bar{\mathbf{H}}\bar{\mathbf{P}}_p\bar{\mathbf{H}}^*. \quad (6.A.5)$$

The channel covariance matrix is given by $\mathbf{R}_h = \mathbf{H}\mathbf{H}_0\mathbf{H}^*$. Using Assumption 2, we find that

$$\bar{\mathbf{H}}\bar{\mathbf{P}}_f\bar{\mathbf{H}}^* = \mathbf{R}_h - \mathbf{R}_h\Psi^*(\Psi\mathbf{R}_h\Psi + \mathbf{R})^{-1}\Psi\mathbf{R}_h = (\mathbf{R}_h^{-1} + \Psi^*\Psi)^{-1}. \quad (6.A.6)$$

To obtain the last expression in (6.A.6), we have used the matrix inversion lemma (see Appendix 3.A.2) and the fact that $\mathbf{R} = \mathbf{I}$.

The optimization problem is then to find

$$\arg_{\Psi} \min \text{tr}[(\mathbf{R}_h^{-1} + \Psi^*\Psi)^{-1} \oslash \mathbf{R}_h] \quad (6.A.7)$$

under the constraint that

$$\text{diag}(\Psi^*\Psi) = [w \ w \ \dots \ w]^T, \quad (6.A.8)$$

where \oslash means element-wise division, and $\text{diag}(\mathbf{A})$ is a vector holding the diagonal elements of \mathbf{A} .

The optimization problem (6.A.7) is a system of polynomial equations in the unknown elements of Ψ and can be solved numerically. Carrying out the numerical optimization on a few example cases, we make the following empirical observations:

Observation 1: When the number of UEs u is less than or equal to the filter width w and all UEs have the same SNR γ and independent channels, the optimal choice of pilots under the assumptions in Section 6.A.2 is any set of orthogonal sequences. The filter estimation NMSE per UE, $\xi_{u \leq w}$, is then

$$\xi_{u \leq w} = (1 + w\gamma)^{-1}. \quad (6.A.9)$$

Observation 2: When the number of UEs u is $w + 1$, where w is the filter width and all UEs have the same SNR γ and independent channels, one optimal choice of pilot sequences under the assumptions in Section 6.A.2 is one where the sequences constitute the vertices of a w -simplex (with an arbitrary rotation in the complex w -space). The filter estimation NMSE per UE, $\xi_{u=w+1}$, is then

$$\xi_{u=w+1} = \frac{1 + \gamma}{1 + (w + 1)\gamma}. \quad (6.A.10)$$

Observation 3: Multiplying the UEs' respective pilot sequences (i.e. the columns of Ψ) with arbitrary complex numbers of unit modulus does not change the total NMSE. The set of all optimal solutions to (6.A.7) are given by simplex vertex sequences subjected to such multiplications.

We study a system of 5 UEs ($u = 5$) tracking four parallel pilot subchannels ($w = 4$). Each UE has a flat fading channel with a flat AR4 Doppler

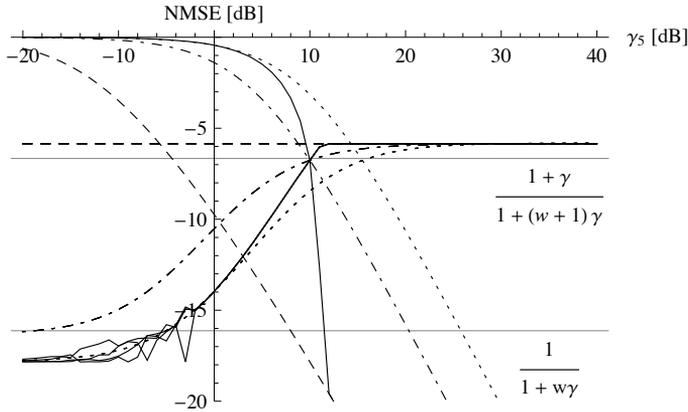


Figure 6.15: Five UEs sharing four subchannels, four UEs of which have SNR=10 dB and the fifth having variable SNR, here denoted γ_5 . Pilot sequences were optimized for an assumed SNR of UE 5= $\{0$ dB, 10 dB, 20 dB $\}$ (dotted,dash-dotted,dashed), and optimized for the true SNR for UE 5= γ_5 in each point(solid).

spectrum with maximum normalized Doppler frequency 0.06. The SNRs of four UEs are held constant at 10 dB while the SNR of the fifth UE is varied over a wide range. The pilot sequences are then optimized for a number of different scenarios and the NMSE performance of optimal filters operating on these pilot sequences are calculated.

Figure 6.15 illustrates the respective NMSEs of the five UEs for different pilot optimizations. We use pilot sequences calculated according to (6.A.7) for the true SNRs, 10 dB, for UEs 1 through 4, and for the SNR for the fifth UE set to 0 dB (case A), 10 dB (case B), and 20 dB (case C), respectively. For completeness, we also consider a fourth case, where we in each point calculate new pilot sequences for the correct SNRs of all five UEs (case D). For each of the four investigated scenarios, two curves are produced; one for UEs 1–4 (lower curve at the far left), and one for UE 5 (upper curve at the far left).

In case C (dashed lines), when the fifth UE is assumed to strongly dominate over the other UEs, the optimization algorithm minimizes the interference from UE 5 by allocating to it an exclusive dimension in the four-dimensional signal space. UEs 1–4 will be symmetrically distributed over the remaining three dimensions, so that their NMSE is given by (6.4.5) with $w = 3$ (straight dashed line). The fifth UE is undisturbed by other UEs. Its NMSE performance is therefore the same as that of a single UE having filter width 4. Its performance curve therefore coincides with the case $w = 4$ for flat fading in Figure 6.2.

The other extreme case is case A (dotted lines), in which the interference caused by UE 5 is assumed to be almost negligible. The optimization algorithm will then ignore UE 5 and create mutually orthogonal pilot sequences for UEs 1–4. As the SNR of UE 5 goes to zero, their NMSE curves will therefore asymptotically go to the level of performance for a single user at 10 dB (lower dotted curve at the left). This level is somewhat lower than (6.4.4) due to some performance gained from the channel’s time dynamics.

Case B falls in between cases A and C. At 10 dB, the point for which the pilot sequences have been optimized, the two curves cross at (approximately) the NMSE value given by (6.4.5) with $w = 4$.

Note the general behaviour of case D: the optimization scheme “sacrifices” the UE(s) having the worst SNR. To the left in the figure, the tendency is towards case A, so that the group of four UEs have decent NMSEs whereas UE 5 would be practically undetectable for the detector. As the SNR of UE 5 rises past 10 dB, the behaviour of case C is preferred so that all effort is put on producing a high NMSE for the upcoming promise, UE 5.

As the SNR of UE 5 goes to infinity, one would perhaps expect that the asymptotic performance of UEs 1–4 would be lower than what is indicated in Figure 6.15. Except for case C, when the pilot sequence used by UE 5 is practically orthogonal to other pilot sequences, the interference caused by UE 5 could be expected to completely destroy the performance of UEs 1–4. It is therefore interesting to note that by jointly estimating all five channels, the performance for UEs 1–4 does not go below that of case C, where, although one dimension in the signal space is lost, UE 5 does not cause any interference.

Note that when we go from a fully loaded system with NMSE (6.4.4) to an overloaded system with NMSE (6.4.5), the NMSE penalty is severe. The overall conclusion from the pilot optimization is therefore that UE overloading should be avoided. However, there may be situations when a poor NMSE performance is acceptable. For example, opportunistic greedy scheduling aims to allocate radio resources to the UE currently having the best channel. This may be possible to do even at comparably high NMSE levels. If the channels of the respective UEs are slowly fading so that channel prediction performance is virtually the same as channel estimation performance, then it may be worth the while to let one extra UE compete for the resources in the competition band, at the expense of increased NMSE levels.

The study conducted in Figure 6.15 was carried out on flat fading channels. When the UEs experience channels that are frequency selective, NMSE levels will depend on what pilot sequences the numerical optimization algorithm happens to find. We repeat our study of optimal pilot sequence performance, but here we use the frequency-selective WINNER II C2 NLOS channel for all five UEs. All other parameters are identical. See Figure 6.16,

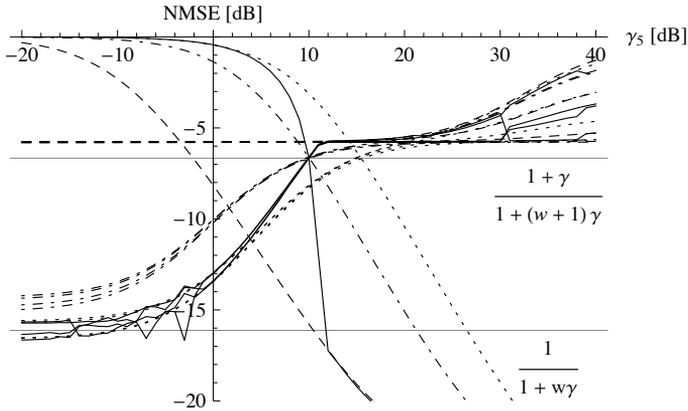


Figure 6.16: NMSEs of five UEs tracking four parallel subchannels, four UEs of which have SNR=10 dB and the fifth has variable SNR, here denoted γ_5 . Pilot sequences were optimized for $\gamma_5 = \{0 \text{ dB}, 10 \text{ dB}, 20 \text{ dB}\}$ (dotted, dash-dotted, dashed), and optimized in each point γ_5 (solid).

in which $u = 5$ UEs share $w = 4$ subchannels. The difference between Figures 6.15 and 6.16 is, except some discrepancies in actual NMSE levels, that the curves spread out in the latter, especially for high SNRs. The reason is that for frequency-selective channels of w subchannels, the cross-correlation between the channels depend on how the set of u pilot sequences happens to be rotated in w -space. For flat fading channels, this rotation is irrelevant.

Link adaptation for uncertain channel state information

7.1 Introduction

In this chapter we consider the problem of link adaptation in a MIMO-OFDM system with imperfect channel estimation and prediction. Depending on channel quality, the MIMO capability allows for the creation of one or many spatially independent links between the base station (BS) and the user equipment (UE), through spatial multiplexing or beamforming.

The basic principle for link adaptation is to adjust transmission parameters to compensate for or utilize fluctuations in the instantaneous link qualities, so as to produce as high a transmission rate as possible, while maintaining performance constraints such as bit error rates or packet error rates. Link adaptation is complicated by the fact that channel estimates are imperfect and that channel prediction is required to compensate for system delays when using the most advanced, high-performance strategies. This chapter discusses how to take into account uncertainties stemming from channel estimation and prediction when taking link adaptation decisions.

The system that we will consider has a given set of available coded modulation (CM) formats, each giving rise to certain spectral efficiencies and bit error probabilities, ranging from a CM format with few bits per symbol¹ and a low error probability (e.g. coded BPSK), to a format with many bits per symbol but comparably high error probability (e.g. uncoded 64QAM).

The transmission resource unit used in OFDM link adaptation and scheduling is the *resource block* (RB). The size of the RB is assumed to have been chosen in such a way that the channel fading is moderate within the block.

¹We shall here refer to a time-frequency symbol in an OFDM symbol as just a *symbol*.

A single CM format is used within an RB. If the CM format is a coded format, then the code operates within the block only. We will assume that K simultaneous RBs per user are considered for each link adaptation decision. The K blocks can be distributed in the frequency and space domains, but are assumed to be fairly narrowly localized in time.

To achieve good performance in terms of bit error rate (BER) versus signal-to-noise ratio (SNR), it is also common to apply a strong channel code, commonly denoted the *outer* code. Turbo codes and Low-Density Parity-Check (LDPC) codes are known to have near-capacity performance for Additive White Gaussian Noise (AWGN) channels and are therefore used in e.g. LTE [33] and the WINNER II system concept [74]. If the channel code is a block code, it should operate on fairly large codewords, since large block codes generally have better performance than small block codes. The number of resource blocks K assigned to a user should be large enough that at least one codeword of the channel code fits within the K RBs. When an outer code is used, we assume the use of *bit-interleaved coded modulation* (BICM) [87], so that the output bits from the encoder are interleaved before they are mapped onto symbols.

From channel quality measurements of the K RBs, the link adaptation algorithm should then determine transmission parameters such as transmission power, the CM formats for the respective blocks, and/or the rate of the outer code. Due to the inevitable delay in the control feedback loop, adaptation decisions will have to be based on uncertain channel predictions.

The ultimate purpose of the link adaptation is to maximize the throughput of error-free data, which we here refer to as just the *throughput*. Transmitted and decoded data packets of a predefined size are checked for errors using e.g. a *cyclic redundancy check* (CRC) code, which almost certainly will detect any error. The expected throughput for a packet of K resource blocks with SNRs given by a K -vector $\bar{\gamma}$ is then

$$\text{expected throughput} = (1 - P_{pa}(\bar{\gamma}, \mathbf{m}))r(\bar{\gamma}, \mathbf{m}), \quad (7.1.1)$$

where $P_{pa}(\bar{\gamma}, \mathbf{m})$ is the packet error probability and $r(\bar{\gamma}, \mathbf{m})$ is the transfer bit rate, given SNRs $\bar{\gamma}$ and transmission parameters \mathbf{m} . For simplicity, we above assume that erroneously received data packets are discarded and have to be entirely retransmitted. The transmission parameters should then be adapted so as to maximize the expected throughput (7.1.1). The dilemma is of course that changing a transmission parameter will increase the error probability at the same time as the rate increases, or vice versa. Also, since the SNRs may vary among the resource blocks, maximization of (7.1.1) is a hard multidimensional optimization problem.

We may also express the expected throughput in terms of the codeword error rate (CWER). The outer code, if used, produces codewords that ex-

hibit a code word error probability P_{cw} . Assuming independent and equal codeword error probabilities between the different codewords in a packet, we have:

$$\text{expected throughput} = (1 - P_{cw}(\bar{\gamma}, \mathbf{m}))^{n_{cw}} r(\bar{\gamma}, \mathbf{m}), \quad (7.1.2)$$

in which n_{cw} is the number of codewords per data packet. Analogously, we can write

$$\text{expected throughput} = (1 - P_b(\bar{\gamma}, \mathbf{m}))^{n_b} r(\bar{\gamma}, \mathbf{m}), \quad (7.1.3)$$

in terms of the post-decoder bit error probability P_b and the number of bits per packet n_b . Again, we then assume independent bit error probabilities. This is a valid assumption if BICM is used, so that the bits are scrambled over the transmission.

The link adaptation algorithm should determine the optimum values of the error probability and the data rate in (7.1.2) or (7.1.3), and then adjust the transmission parameters \mathbf{m} to achieve those values based on uncertain predictions of the SNRs $\bar{\gamma}$. An example of such an optimization is given in [88], where the throughput (7.1.3) is maximized for uncertain predictions, without using an outer code.

Here, we simplify the problem statement and assume that the target P_{cw} or P_b has been determined in such a way as to produce a high expected throughput, possibly taking constraints on the packet transmission delay statistics into account. Link adaptation then amounts to producing as high a rate $r(\bar{\gamma}, \mathbf{m})$ as possible, under the constraint of a fixed codeword error probability P_{cw} or a fixed bit error probability P_b . In this work, the transmission parameters \mathbf{m} to be adjusted are the local CM formats and the rate of the outer code.

7.1.1 Decision metric

To maintain a certain error rate, we need to find a metric ν that correlates with the error rate that we wish to constrain. We will here examine two different choices of ν . If we choose to have a fixed bit error rate, then ν is the average bit error rate P_b itself, averaged over the K RBs. If, on the other hand, we wish to have a fixed codeword error rate P_{cw} and we employ a near-capacity achieving channel decoder based on soft detection metrics, then the appropriate metric is the average mutual information per bit, \bar{I}_c , again averaged over the K RBs, as a number of works have shown [89],[41],[90]. We shall therefore consider $\nu = P_b$ or $\nu = \bar{I}_c$, depending on link adaptation strategy.

Link adaptation may be divided into three categories; optimization of the rate of the outer code, optimization of local CM formats, or joint optimization of both local CM formats and outer code rate. In this chapter we will study the two latter link adaptation strategies under the presence of channel estimation and prediction errors. Link adaptation decisions will be based on the metric ν , which in turn is formed from uncertain channel predictions.

Two time instants are central to the link adaptation problem in a channel prediction setting: the time of decision, which we here take to be $t = 0$ and which is the time when the link adaptation decision needs to be taken, and the time of detection, which we set to $t = L$ so that the prediction range is L time steps.

The fundamental problem of link adaptation with channel prediction is that at $t = 0$, we will be uncertain about what value ν will have at time $t = L$. We will here assume that ν is the arithmetic mean value of local metrics $\{\nu_k\}_{k=1}^K$, where ν_k is the local metric for resource block k . If ν_k is associated with bits, we then have

$$\nu = \frac{\sum_{k=1}^K m_k \nu_k}{\sum_{k=1}^K m_k}, \quad (7.1.4)$$

where m_k is the number of bits per symbol used in resource block k . The relation (7.1.4) holds for the two metrics \bar{P}_b and \bar{I}_c to be studied here, if ν_k is defined as either local bit error probability or local average mutual information per bit. Since we assume practically flat fading within an RB, ν_k will be a function of the transmission parameters and the SNR γ_k in that RB.

Each ν_k may have a broad distribution and the $\{\nu_k\}_{k=1}^K$ are often correlated, complicating the calculation of the distribution of ν . To simplify the problem, we group $\{\nu_k\}_{k=1}^K$ into N sets $\{\Omega_n\}_{n=1}^N$. A set Ω_n defines a small region in the parameter space for the pdf for a resource block's SNR, so that all $\gamma_k : k \in \Omega_n$ have approximately the same pdf. The link adaptation must then attribute the same transmission parameters to all RBs in a set Ω_n , so that all ν_k belonging to the same set Ω_n therefore have approximately the same distribution. The total metric ν can thus be written as

$$\nu = \frac{\sum_{n=1}^N m_{k(n)} |\Omega_n| \bar{\nu}_n}{\sum_{n=1}^N m_{k(n)} |\Omega_n|}, \quad (7.1.5)$$

where $|\Omega_n|$ is the cardinality for Ω_n , and $k(n)$ is the index for an arbitrary RB in Ω_n , so that $m_{k(n)}$ is the number of bits per symbol used by all RBs in the set Ω_n . The quantity $\bar{\nu}_n$ is defined as

$$\bar{\nu}_n = |\Omega_n|^{-1} \sum_{k \in \Omega_n} \nu_k. \quad (7.1.6)$$

If K is large enough that the channel decorrelates several times over the K blocks, this means that although an individual ν_k may have a broad distribution, $\bar{\nu}_n$ may be narrow owing to the central limit theorem if $|\Omega_n|$ is large. Instead of having K broad distributions to consider, we then have N fairly narrow distributions, making the link adaptation decision simpler. In Section 7.2.4, we shall see how letting K go to infinity in fact makes all uncertainty about the future value of ν vanish.

7.2 Decision metrics using channel prediction

At time $t = 0$ we want to infer a decision metric ν pertaining to time $t = L$. This decision metric ν is a mean value of values ν_k local to individual RBs. In flat-fading scenarios such as ideal OFDM, a local metric ν_k will be a function of the local SNR γ . To enable efficient link adaptation, we must therefore consider the distribution of the future $\gamma(t = L)$ at $t = 0$. This requires us to model the fading channel.

7.2.1 Channel model

High quality channel estimates, necessary for coherent detection, are here assumed to be produced by a Kalman filter based on a state space channel and measurement model,

$$\begin{aligned}\mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t, \\ \mathbf{h}_t &= \mathbf{H}\mathbf{x}_t, \\ \mathbf{y}_t &= \Phi_t\mathbf{h}_t + \mathbf{v}_t,\end{aligned}\tag{7.2.1}$$

where \mathbf{u}_t , \mathbf{v}_t , and \mathbf{x}_0 are zero-mean Gaussian, white and

$$\|\{\mathbf{u}_t, \mathbf{v}_t, \mathbf{x}_0\}\|^2 = \text{diag}(\mathbf{Q}, \mathbf{R}, \Pi_0).\tag{7.2.2}$$

The state covariance matrix $\Pi_t = \|\mathbf{x}_t\|^2$ obeys the recursion (see Theorem 3.4.1)

$$\Pi_{t+1} = \mathbf{F}\Pi_t\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*.\tag{7.2.3}$$

Here we will assume that Π_0 is assigned in such a way that the model becomes stationary, i.e. the state covariance has a constant value (see Section 4.2.4),

$$\Pi \equiv \Pi_t.\tag{7.2.4}$$

We furthermore assume pilot-assisted channel estimation and prediction, so that (7.2.1) is sampled on sub-locations containing known reference symbols (pilots). The vector \mathbf{h}_t is the modelled fading channel coefficients, and

Φ_t contains pilot symbols. The model (7.2.1) is general enough to represent a multiuser MIMO system as described in Chapter 4. If u simultaneous inputs are tracked, where each input is the signal from one transmitting antenna, the vector \mathbf{h}_t has length uw , so that w parallel pilot-bearing subchannels per input are tracked in parallel in the fading OFDM channel. A high value of w leads to good estimation performance but also high algorithmic complexity.

Optimal filtered estimates $\hat{\mathbf{x}}_{t|t}$ of the state vector \mathbf{x}_t , i.e. estimates based on noisy measurement up to the present time t , are recursively produced by a Kalman filter (KF),

$$\hat{\mathbf{x}}_{t|t} = (\mathbf{I} - \mathbf{K}_{f,t}\mathbf{H})\mathbf{F}\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{K}_{f,t}\mathbf{y}_t, \quad (7.2.5)$$

where $\mathbf{K}_{f,t}$ is the Kalman filter gain calculated from the KF recursions (see Section 4.3.1). Optimal channel estimations of the w parallel channel coefficients are then calculated as

$$\hat{\mathbf{h}}_{t|t} = \mathbf{H}\hat{\mathbf{x}}_{t|t} \quad (7.2.6)$$

and optimal L -step predictions are calculated as

$$\hat{\mathbf{h}}_{t+L|t} = \mathbf{H}\mathbf{F}^L\hat{\mathbf{x}}_{t|t}. \quad (7.2.7)$$

From (3.4.47) it follows that the prior distribution for L -step predictions, which is here the same as the frequency distribution over time, is

$$p(\hat{\mathbf{h}}_{t+L|t}|I) = \mathcal{CN}(\hat{\mathbf{h}}_{t+L|t}; \mathbf{0}, \mathbf{H}\Sigma_{t+L|t}\mathbf{H}^*). \quad (7.2.8)$$

where $\Sigma_{t+L|t}$ is the covariance matrix for the state prediction, see Definition 3.4.1. Further, for a predicted vector of channel coefficients we have the pdf

$$p(\mathbf{h}_{t+L}|\mathbf{Y}_t, I) = \mathcal{CN}(\mathbf{h}_{t+L}; \hat{\mathbf{h}}_{t+L|t}, \mathbf{H}\mathbf{P}_{t+L|t}\mathbf{H}^*), \quad (7.2.9)$$

where $\mathbf{P}_{t+L|t}$ is the covariance matrix for the state prediction error. In the above, following [48], we denote by a symbol \mathbf{Y}_t all measurements $\{\mathbf{y}_0, \dots, \mathbf{y}_t\}$ up to time t . From (3.4.31) it follows that

$$\mathbf{H}\mathbf{P}\mathbf{H}^* = \mathbf{H}\Sigma_{t+L|t}\mathbf{H}^* + \mathbf{H}\mathbf{P}_{t+L|t}\mathbf{H}^*, \quad (7.2.10)$$

where $\mathbf{H}\mathbf{P}\mathbf{H}^*$ is the constant covariance matrix for the channel coefficients \mathbf{h}_t . Hence there is a balance between the power $\mathbf{H}\Sigma_{t+L|t}\mathbf{H}^*$ for the predictions and the variance $\mathbf{H}\mathbf{P}_{t+L|t}\mathbf{H}^*$ of the prediction errors. That is to say, for a channel with constant power, when the predictions are uncertain, then $\mathbf{H}\mathbf{P}_{t+L|t}\mathbf{H}^*$ has large diagonal elements, while $\mathbf{H}\Sigma_{t+L|t}\mathbf{H}^*$ has small diagonal elements, i.e. the predictions will almost always be close to zero. Conversely, when the predictions are very accurate, then the error covariance $\mathbf{H}\mathbf{P}_{t+L|t}\mathbf{H}^*$

has small diagonal elements while $\mathbf{H}\Sigma_{t+L|t}\mathbf{H}^*$ has large elements. The predictions will then be virtually the same as the true channel.

The KF produces filtered channel estimates for pilot sub-locations. To detect a payload symbol at sub-location $\{f', t\}$, the channel estimates at pilot sub-locations in its vicinity need to be intra- and extrapolated to this sub-location, to produce a (sub-optimal) channel estimate $\hat{h}_{f',t|t}$. For simplicity, we assume that the inter-/extrapolation consists of choosing $\hat{h}_{f',t|t} = \hat{h}_{f,t|t}$ as channel estimate, where $\hat{h}_{f,t|t}$ is the optimal estimate of the true channel coefficient $h_{f,t}$ for the nearest pilot sub-location $\{f, t\}$. Thus, $\hat{h}_{f,t|t}$ is an element in $\hat{\mathbf{h}}_{t|t}$ and corresponds to a location slightly displaced from the sub-location $\{f', t\}$ (to simplify notation, we let the displacement be a frequency-wise displacement only). In what follows, we assume that the error caused by the inter-/extrapolation can be neglected.

A received symbol y_L at $t = L$ can be expressed as

$$y_L = s\hat{h}_{L|L} + s\tilde{h}_{L|L} + v, \quad (7.2.11)$$

where $\hat{h}_{L|L}$ and $\tilde{h}_{L|L} = h_{L|L} - \hat{h}_{L|L}$ are the filtered channel estimate and the filtered channel estimation error, respectively. The unknown payload symbol with variance S is denoted s , and $s\tilde{h}_{L|L}$ is the residual unknown error from the imperfect channel estimation. The thermal noise and interference is denoted by v and is assumed to have known variance σ_v^2 . For brevity, we have excluded the subchannel index, so that we define

$$h_{L|L} \triangleq h_{f,L|L} \quad \text{and} \quad \hat{h}_{L|L} \triangleq \hat{h}_{f,L|L}. \quad (7.2.12)$$

Note that $\hat{h}_{L|L}$ and $\tilde{h}_{L|L}$ are associated with the nearest pilot sub-location $\{f, t\}$, while s and v are associated with the actual payload sub-location $\{f', t\}$, so that v and $\tilde{h}_{L|L}$ are uncorrelated.

At the moment of decision, $t = 0$, the channel h_L is unknown. At this point, two aspects of the signal (7.2.11) need to be considered: the pdf of the magnitude of the “effective” channel $\hat{h}_{L|L}$, and the power of the effective noise $s\tilde{h}_{L|L} + v$. In subsequent sections we will discuss these two aspects and show that at the moment of detection, $t = L$, the effective SNR is

$$\gamma = \frac{S|\hat{h}_{L|L}|^2}{\sigma_v^2 + S\sigma_f^2}, \quad (7.2.13)$$

where σ_f^2 is the error variance for the filtered estimates which can be calculated from the KF. Note that we will infer the power of the *future filtered estimate* $\hat{h}_{L|L}$, since this is what ultimately will dictate the effective SNR, and not the power of the true future channel h_L .

7.2.2 The distribution of a future filtered estimate

The scalar channel estimate $\hat{h}_{L|L}$ in (7.2.13) is an element in the vector $\hat{\mathbf{h}}_{L|L}$ in (7.2.6). From measurements up to time $t = 0$ we therefore wish to calculate the pdf of the future filtered estimate $\hat{\mathbf{h}}_{L|L}$. First we study the future filtered *state* estimate and observe that it may be written (by recursive application of (3.4.40))

$$\hat{\mathbf{x}}_{L|L} = \mathbf{F}^L \hat{\mathbf{x}}_{0|0} + \sum_{t=1}^L \mathbf{F}^{L-t} \mathbf{K}_{f,t} \mathbf{e}_t. \quad (7.2.14)$$

If measurements up to and including $t = 0$ is available, then $\hat{\mathbf{x}}_{0|0}$ can be constructed. The sum in (7.2.14) is however unknown at $t = 0$. Since the innovations $\{\mathbf{e}_t\}$ are Gaussian, white, and zero mean, it is clear that the sought distribution is Gaussian,

$$p(\hat{\mathbf{x}}_{L|L} | \mathbf{Y}_0, I) = \mathcal{CN}(\hat{\mathbf{x}}_{L|L}; \hat{\mathbf{x}}_{L|0}, \Xi_{L|0}), \quad (7.2.15)$$

with mean value $\hat{\mathbf{x}}_{L|0} = \mathbf{F}^L \mathbf{x}_{0|0}$ and some covariance matrix $\Xi_{L|0}$. To find the covariance $\Xi_{L|0}$, we first make the following definition.

Definition 7.2.1 (Variance of a future filtered state estimate). *The covariance of a future filtered state estimate is defined as*

$$\Xi_{t|t_0} \triangleq E\{\hat{\mathbf{x}}_{t|t} \hat{\mathbf{x}}_{t|t}^* | \mathbf{Y}_{t_0}, I\}, \quad t > t_0. \quad (7.2.16)$$

A recursion formula for the future filtered state estimate can readily be found as stated by the following theorem.

Theorem 7.2.1 (Recursion for $\Xi_{t|t_0}$). *Given the model (7.2.1) and $\mathbf{K}_{f,t}$ and $\mathbf{R}_{e,t}$ given by (3.4.33) and (3.4.34) in the KF recursions in Chapter 3, it holds that*

$$\Xi_{t+1|t_0} = \mathbf{F} \Xi_{t|t_0} \mathbf{F}^* + \mathbf{K}_{f,t+1} \mathbf{R}_{e,t+1} \mathbf{K}_{f,t+1}^*, \quad \Xi_{t_0|t_0} = 0. \quad (7.2.17)$$

Proof. Since the innovations $\{\mathbf{e}_t\}$ are uncorrelated with earlier state estimates, taking the variance of $\hat{\mathbf{x}}_{t+1|t+1} = \mathbf{F} \hat{\mathbf{x}}_{t|t} + \mathbf{K}_{f,t+1} \mathbf{e}_{t+1}$ (see (3.4.40)) gives the recursions. This, in combination with the fact that the filtered state estimation vector can be constructed without error when $t = t_0$ so that the variance then is zero, proves the result. \square

Using this result, the future filtered state covariance can be expressed in terms of state estimation error covariance matrices:

Theorem 7.2.2 ($\Xi_{t|t_0}$ in terms of state estimation error covariance matrices). *Consider the model (7.2.1) and $\mathbf{P}_{t|t_0}$ and $\mathbf{P}_{t|t}$ obtained as in (3.4.46) and (3.4.41). Then*

$$\Xi_{t|t_0} = \mathbf{P}_{t|t_0} - \mathbf{P}_{t|t}. \quad (7.2.18)$$

Proof. Using the recursions (3.4.41) and (3.4.46),

$$\mathbf{P}_{t+1|t+1} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^* - \mathbf{K}_{f,t+1}\mathbf{R}_{e,t+1}\mathbf{K}_{f,t+1}^*, \quad (7.2.19)$$

$$\mathbf{P}_{t+1|t_0} = \mathbf{F}\mathbf{P}_{t|t_0}\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*, \quad (7.2.20)$$

and using (7.2.17), we find that

$$\Xi_{t+1|t_0} + \mathbf{P}_{t+1|t+1} = \mathbf{F}(\Xi_{t|t_0} + \mathbf{P}_{t|t})\mathbf{F}^* + \mathbf{G}\mathbf{Q}\mathbf{G}^*. \quad (7.2.21)$$

Comparing (7.2.20) and (7.2.21), we see that the recursion in $\Xi_{t|t_0} + \mathbf{P}_{t|t}$ is the same as that for $\mathbf{P}_{t|t_0}$ and that they are initialized with the same value, $\Xi_{t_0|t_0} + \mathbf{P}_{t_0|t_0} = \mathbf{P}_{t_0|t_0}$. It must therefore hold that

$$\Xi_{t|t_0} + \mathbf{P}_{t|t} = \mathbf{P}_{t|t_0}, \quad (7.2.22)$$

which gives the sought result. \square

Returning now to Equation (7.2.15) and using $\hat{\mathbf{h}}_{L|L} = \mathbf{H}\hat{\mathbf{x}}_{L|L}$, we find that a future filtered estimate $\hat{h}_{L|L}$ of a channel coefficient h_L has distribution

$$p(\hat{h}_{L|L}|\mathbf{Y}_0, I) = \mathcal{CN}(\hat{h}_{L|L}; \hat{h}_{L|0}, \sigma_{pf}^2), \quad (7.2.23)$$

where the mean value $\hat{h}_{L|0}$ is an element in $\mathbf{H}\hat{\mathbf{x}}_{L|0}$, and the variance σ_{pf}^2 is the corresponding diagonal element in $\mathbf{H}\Xi_{L|0}\mathbf{H}^*$. The squared magnitude of $\hat{h}_{L|L}$, which we define as

$$\hat{z}_{L|L} \triangleq |\hat{h}_{L|L}|^2, \quad (7.2.24)$$

can be shown to have the distribution (see Appendix 4.A)

$$p(\hat{z}_{L|L}|\mathbf{Y}_0, I) = \chi^2(\hat{z}_{L|L}; |\hat{h}_{L|0}|^2, \sigma_{pf}^2), \quad (7.2.25)$$

where $\chi^2(\cdot)$ is the non-central χ^2 -distribution.

A number of previous works [4], [63], [64] use a different expression than a non-central χ^2 -distribution for the pdf of a predicted channel. However, this alternative expression can be shown to be identical to a non-central χ^2 -distribution, although the variance used there differs from that used in (7.2.25). See Appendix 4.B.

The diagonal elements of $\mathbf{H}\Xi_{L|0}\mathbf{H}^*$ constitute the variances of future filtered estimates, which play an important role in the channel prediction procedure, but what is the nature of $\Xi_{L|0}$? Is it preferable to have a $\Xi_{L|0}$ with large or small eigenvalues? To acquire a better understanding of the matrix $\Xi_{L|0}$, study Figure 7.1. Theorem 7.2.2 for the case of a static channel model and a stationary filter is here illustrated in the form of a triangle in which the sides and corners represent various extreme conditions.

As seen in (3.4.31), state estimation error covariance matrices can at most take the value Π , where Π is the stationary state covariance, see Definition 3.4.1. The left side of Figure 7.1 represents an extreme case in the sense that prediction here is as bad as it could possibly be; this is the case when the prediction range is very long, or the UE velocity is high, so that no relevant information about the future channel is available to the predictor. Another extreme situation, but now at the good end of the scale, is indicated by the uppermost side. Here, channel estimation at the time of detection is error-free. Going from a point on the left side to a point on the upper side, the state estimation error covariances $\mathbf{P}_{L|0}$ and $\mathbf{P}_{L|L}$ inevitably decrease. $\Xi_{L|0}$, on the other hand, may increase, decrease, or stay constant along the path, depending on start and stop locations. Therefore, $\Xi_{L|0}$ is not a quality measure on its own standard. We need either $\mathbf{P}_{L|0}$ or $\mathbf{P}_{L|L}$ as “reference” to be able to tell whether the situation is good or bad.

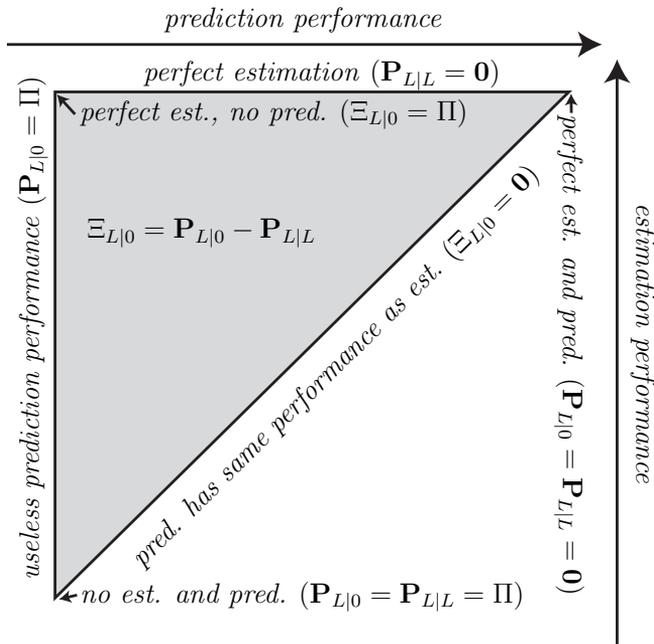


Figure 7.1: The relationship between the three covariances $\Xi_{L|0}$, $\mathbf{P}_{L|0}$, and $\mathbf{P}_{L|L}$ for a static channel model with constant state covariance Π and a stationary filter.

In Table 7.1, we summarize definitions of useful variances. From (3.4.31) we know that $\Pi_L = \Sigma_{L|0} + \mathbf{P}_{L|0}$. By (7.2.22) it therefore holds for a stationary

Table 7.1: Definitions of variances.

symbol	description	diagonal element of...
σ_h^2	Channel power	$\mathbf{H}\mathbf{I}\mathbf{H}^*$
σ_f^2	Error variance of a filtered estimate	$\mathbf{H}\mathbf{P}_{L L}\mathbf{H}^*$
σ_{pf}^2	Variance of a future filtered estimate	$\mathbf{H}\mathbf{\Xi}_{L 0}\mathbf{H}^*$
σ_p^2	Variance of a prediction	$\mathbf{H}\mathbf{\Sigma}_{L 0}\mathbf{H}^*$

filter ($\Pi_t \equiv \Pi$) that

$$\sigma_h^2 = \sigma_f^2 + \sigma_{pf}^2 + \sigma_p^2. \quad (7.2.26)$$

7.2.3 The effective noise at detection

In (7.2.11), we saw that the total effective noise at $t = L$ is $s\tilde{h}_{L|L} + v$, where s is the transmitted symbol, $\tilde{h}_{L|L} = h_L - \hat{h}_{L|L}$ is the channel estimation error of the nearest pilot sub-location, and v is noise with variance σ_v^2 . We assume approximate flat fading within the resource block, so that the optimal channel estimate for the nearest pilot sub-location can be used as estimate for the current payload sub-location.

Since by assumption v is white over both frequency and time and $\hat{h}_{L|L}$ is the channel estimate for a sub-location slightly displaced from the current one, $s\tilde{h}_{L|L}$ and v will be uncorrelated. Assuming that s is unknown but has prior variance S , the total effective noise variance is then $S\sigma_f^2 + \sigma_v^2$, where σ_f^2 is the error variance of the filtered estimates, see Table 7.1.

We overestimate the impact of the total noise by modelling it as Gaussian noise, which has the highest entropy for a given variance. However, we expect the improvement that would result from a more careful modelling to be negligible.

The reason why we cannot use the optimal estimate in the current sub-location is that there is no pilot symbol present at that sub-location to aid the pilot-assisted channel estimation. Iterative channel estimation (ICE) can be used to remedy this. The unknown payload symbol is then detected in a primary phase, based on a nearby channel estimate. Then, in a second phase, this detected symbol is used as pilot to produce a channel estimate $\hat{h}_{f',L|L}$ for the current sub-location $\{f', t\}$. We can now use $\hat{h}_{f',L|L}$ instead of the estimate $\hat{h}_{f,L|L}$ for the pilot sub-location $\{f, t\}$ for payload detection, but since the noise v lies in the subspace spanned by $\hat{h}_{f',L|L}$, this has the consequence that $\hat{h}_{f',L|L}$ and v are correlated. The total noise variance will then not sum to $S\sigma_f^2 + \sigma_v^2$. See Appendix 7.A for further details on this matter.

Above, we assume that the unknown payload symbol has variance S .

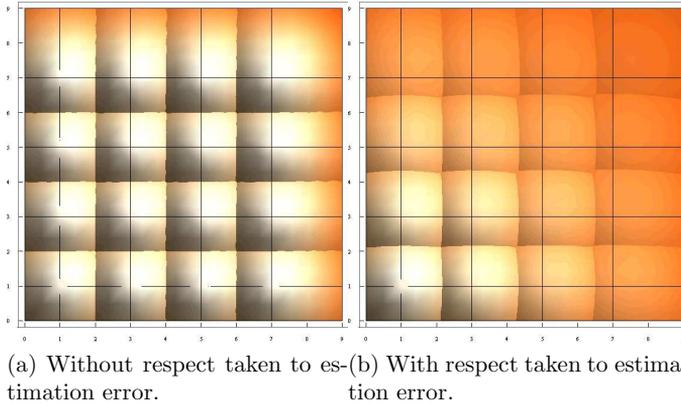


Figure 7.2: Decision boundaries for a 64QAM constellation. Only the first quadrant is illustrated. The lines intersect at the positions of the error-free symbols. The decision boundaries are given by the intersections of circular symmetric complex Gaussian functions with appropriate variances centred at these positions. The SNR without estimation error is 10 dB and the mean estimation error variance here equals the noise variance, so that $\sigma_v^2 = S\sigma_f^2 = 0.1S$.

However, depending on the value of the symbol s , the term $\hat{s}h_{L|L}$ will vary considerably in magnitude. For example, the difference in energy between the innermost and outermost symbols in a square 64QAM constellation is an astonishing 17 dB [71]! The detector could take this knowledge into account when considering each of the hypothetically transmitted symbols. When hard decisions on transmitted symbols are used, this will lead to a displacement of the ordinary decision boundaries (which are perpendicular bisectors to the lines joining any two symbols). See Figure 7.2. When soft decisions are used, the soft metric is calculated by summing over sets of hypothetically transmitted symbols. The variance of $\hat{s}h_{L|L} + v$, which now differs between the terms in this summation, can in this case be varied appropriately without imposing any additional complexity on the detector.

Assuming that ICE is not used and that we do not take the symbol-dependent power variation of $\hat{s}h_{L|L}$ into consideration, it is convenient to define an instantaneous SNR scaling factor as follows, *cf.* (7.2.13):

$$\gamma_0 \triangleq \frac{S}{\sigma_v^2 + S\sigma_f^2}. \quad (7.2.27)$$

The instantaneous effective SNR of a channel power z_L is then $\gamma_0 \hat{z}_{L|L}$, see (7.2.24), and its distribution based on measurements up to $t = 0$ is (see

Appendix 4.A.2)

$$p(\gamma_0 \hat{z}_{L|L} | \mathbf{Y}_0, I) = \chi^2(\gamma_0 \hat{z}_{L|L}; \gamma_0 \hat{z}_{L|0}, \gamma_0 \sigma_{pf}^2). \quad (7.2.28)$$

Also, we deduce from (7.2.8) that the frequency distribution for predictions $\gamma_0 \hat{z}_{L|0}$ is

$$p(\gamma_0 \hat{z}_{L|0} | I) = \chi^2(\gamma_0 \hat{z}_{L|0}; 0, \gamma_0 \sigma_p^2). \quad (7.2.29)$$

Since (7.2.29) has mean value zero, $p(\gamma_0 \hat{z}_{L|0} | I)$ is an exponential distribution.

7.2.4 Predicted error rate

Turning now to specific decision metrics, we begin by studying the decision metric $\nu = \bar{P}_b$, where \bar{P}_b is the average bit error rate over the K resource blocks assigned to a user. The link adaptation problem is to associate modulation formats with predicted effective SNR levels so as to maintain ν at a certain prescribed level, while maximizing the transmission data rate.

Rate maximization under a constraint on the error rate has been considered before, e.g. in [32] and in [64], although the former type of results considered neither channel estimation errors nor channel prediction errors, and the latter considered channel prediction errors only. In these previous works, the decision metric is completely known already at the time of decision. This may at first seem a bit surprising, because one would expect ν to feature at least some small degree of uncertainty stemming from the prediction. Below, we will see how $\nu = \bar{P}_b$ constitutes a special case in which, at $t = 0$, no uncertainty as to the future value of ν exists.

The metric $\nu = \bar{P}_b$ is appropriate to use in situations where no outer channel code is used, or when the outer channel decoder operates on hard bit decisions. We consider detected bits before a prospective decoder and we want to maximize the transmission rate while maintaining a given bit error rate in the long run. Therefore, the data set over which we want ν to take a specific value is infinitely large. Here, we may therefore say that the number of resource blocks K assigned to a user goes to infinity.

As stated in (7.1.4), we take ν to be a mean value of K local metrics ν_k . As further described in (7.1.5), it is convenient to group the K resource blocks into N sets $\{\Omega_n\}_{n=1}^N$ and express the total metric ν as

$$\nu = \frac{\sum_{n=1}^N m_{k(n)} |\Omega_n| \bar{\nu}_n}{\sum_{n=1}^N m_{k(n)} |\Omega_n|} \quad (7.2.30)$$

where m_k is the number of bits per symbols used in resource block k , and $k(n)$ is an arbitrary RB in Ω_n .

A set Ω_n defines a region in the parameter space for the effective SNR. From (7.2.28), we know that the effective SNR is parameterized by the variance σ_{pf}^2 , and by the expected effective SNR $\gamma_0 \hat{z}_{L|0}$, which we in this section for brevity define as

$$\hat{\gamma} \triangleq \gamma_0 \hat{z}_{L|0}. \quad (7.2.31)$$

However, assuming that the same channel model is used over the whole bandwidth and spatial domain, and assuming that the same set of pilots are used for all RBs, the variance σ_{pf}^2 will be nearly constant. Therefore, the parameter space for the effective SNR is here one-dimensional so that a set Ω_n corresponds to a small range of $\hat{\gamma}$ -values.

As $K \rightarrow \infty$, we also let the number of sets $N \rightarrow \infty$ as well as the cardinality for each set $|\Omega_n| \rightarrow \infty$. In this limit, there is a direct correspondence between a predicted effective SNR $\hat{\gamma}$ and a set Ω_n .

With the assumption of infinitely large sets, it follows that $\bar{\nu}_n$, i.e. the arithmetic mean value (7.1.6) of the local metrics $\{\nu_k\}_{k=1}^K$ in a set, is

$$\bar{\nu}_n = \int_0^\infty \nu_k p(\nu_k | \mathbf{Y}_0, I) d\nu_k, \quad k \in \Omega_n \quad (7.2.32)$$

without uncertainty, if the channel decorrelates several times over the set.

For an RB with index k , assume that we choose to use modulation format j in this RB. We then assign the local metric $\nu_k = P_{b,j}(\gamma_0 \hat{z}_{L|L})$, where $P_{b,j}(\gamma_0 \hat{z}_{L|L})$ is the bit error probability for modulation format j when the effective SNR in the approximately flat fading RB is $\gamma_0 \hat{z}_{L|L}$. Since there is a one-to-one relationship between $P_{b,j}$ and $\gamma_0 \hat{z}_{L|L}$, we change variables (see Section 3.2.2), and use (7.2.28) to find that

$$\begin{aligned} p(\nu_k | \mathbf{Y}_0, I) d\nu_k &= p(\gamma_0 \hat{z}_{L|L} | \mathbf{Y}_0, I) d(\gamma_0 \hat{z}_{L|L}) \\ &= \chi^2(\gamma_0 \hat{z}_{L|L}; \hat{\gamma}, \gamma_0 \sigma_{pf}^2) d(\gamma_0 \hat{z}_{L|L}). \end{aligned} \quad (7.2.33)$$

Hence we have that

$$\begin{aligned} \bar{\nu}_n &= \int_0^\infty \nu_k p(\nu_k | \mathbf{Y}_0, I) d\nu_k \\ &= \int_0^\infty P_{b,j}(\gamma_0 \hat{z}_{L|L}) \chi^2(\gamma_0 \hat{z}_{L|L}; \hat{\gamma}, \gamma_0 \sigma_{pf}^2) d(\gamma_0 \hat{z}_{L|L}). \end{aligned} \quad (7.2.34)$$

Note that, since $|\Omega_n| = \infty$, we do not need to know the pdf:s for the individual ν_k 's; it suffices to know their mean value, $\bar{\nu}_n$, which is a function of $\hat{\gamma}$ and the modulation format j . In what follows, we will denote

$$P_{p,j}(\hat{\gamma}, \gamma_0 \sigma_{pf}^2) \triangleq \int_0^\infty P_{b,j}(\gamma_0 \hat{z}_{L|L}) \chi^2(\gamma_0 \hat{z}_{L|L}; \hat{\gamma}, \gamma_0 \sigma_{pf}^2) d(\gamma_0 \hat{z}_{L|L}) \quad (7.2.35)$$

the predicted bit error rate.

Turning now to the factors $m_{k(n)}|\Omega_n|/\sum_{r=1}^N m_{k(r)}|\Omega_r|$ in (7.1.5), the probability that the predicted effective SNR lies in a small interval around $\hat{\gamma}$ is $|\Omega_n|/K$, assuming that Ω_n is associated with the effective SNR $\hat{\gamma}$. As $K \rightarrow \infty$, this quantity goes to the limit

$$|\Omega_n|/K \rightarrow p(\hat{\gamma}|I)d(\hat{\gamma}). \quad (7.2.36)$$

Inserting (7.2.35) and (7.2.36) into (7.2.30), we find

$$\nu = \frac{\int m_{k(\hat{\gamma})} P_{p,j(\hat{\gamma})}(\hat{\gamma}, \gamma_0 \sigma_{pf}^2) p(\hat{\gamma}|I) d(\hat{\gamma})}{\int m_{k(\hat{\gamma})} p(\hat{\gamma}|I) d(\hat{\gamma})}, \quad (7.2.37)$$

where $j(\hat{\gamma})$ and $m_{k(\hat{\gamma})}$ are the modulation format and the number of bits per symbol, respectively, to be used for predicted effective SNR $\hat{\gamma}$. Observe that (7.2.37) is completely known at the moment of decision $t = 0$. There is no uncertainty involved. However, we have to know the long run frequency distribution $p(\hat{\gamma}|I)$, see (7.2.29), in order to solve the link adaptation problem here.

The link adaptation problem is to associate modulation formats with predicted effective SNR levels so as to maintain ν at a certain prescribed level, while maximizing the transmission data rate. Since there is an infinite number of SNR levels, the link adaptation problem is here a infinite-dimensional problem. However, it is reasonable to assume that high-level modulation formats should be used for high predicted effective SNRs and vice versa, so that one only needs to determine the SNR *switching levels* for the modulation formats. As a function of J switching levels $\{b_0, \dots, b_{J-1}\}$ for the predicted SNR $\gamma_0 \hat{z}_{L|0}$, where J is the number of available CM formats, we can therefore write

$$\nu = \frac{\sum_{j=0}^{J-1} \bar{m}_j \int_{b_j}^{b_{j+1}} P_{p,j}(\gamma_0 \hat{z}_{L|0}, \gamma_0 \sigma_{pf}^2) p(\gamma_0 \hat{z}_{L|0}|I) d(\gamma_0 \hat{z}_{L|0})}{\sum_{j=0}^{J-1} \bar{m}_j \int_{b_j}^{b_{j+1}} p(\gamma_0 \hat{z}_{L|0}|I) d(\gamma_0 \hat{z}_{L|0})}. \quad (7.2.38)$$

where \bar{m}_j and $P_{p,j}(\cdot)$ are the number of bits per symbol and the predicted BER, respectively, for CM format j , and where $b_J = \infty$. Note that, since $K = \infty$, no uncertainty as to the value of ν exists at $t = 0$. If, on the other hand, one would want to restrict the bit error rate over a *finite* number K of resource blocks, then ν *would* have a distribution, requiring the use of decision theory to solve the link adaptation problem.

In Section 7.3, we will optimize switching levels $\{b_0, \dots, b_{J-1}\}$ in a few case studies.

A special case: perfect estimation

In [64], rate optimization is studied in the presence of channel prediction errors, assuming perfect channel estimation at the receiver. The decision metric was then marginalized over the pdf of the actual future channel h_L instead to the future filtered estimate $\hat{h}_{L|L}$ as we do here. Our result is identical to [64] in that case since under the assumption of perfect channel estimation, marginalization over h_L is the same as marginalization over $\hat{h}_{L|L}$. This is clear from Theorem 7.2.2, since perfect channel estimation means that $\mathbf{P}_{L|L} = \mathbf{0}$.

We close this section by studying some characteristics of the predicted bit error rate $P_p(\gamma_0 \hat{z}_{L|0})$.

Characteristics of predicted BER curves

In [91], it is shown that the bit error rate for square MQAM, with $M = 2^k$ and k is an even integer, can be fairly well approximated by

$$P_{b,k}(\gamma_0 \hat{z}_{L|L}) \approx \alpha_1 \exp(-\alpha_2 \gamma_0 \hat{z}_{L|L}), \quad (7.2.39)$$

with $\alpha_1 = 0.2$ and $\alpha_2 = 1.6/(2^k - 1)$. This approximation is good for high SNRs but clearly not so for low SNRs, since $P_{b,k}$ goes to α_1 but should go to 0.5 as $\gamma_0 \hat{z}_{L|L}$ goes to zero. However, with $\alpha_1 = 0.5$, we may envision a hypothetical modulation format for which (7.2.39) provides a good fit over the whole range of SNRs, and use this hypothetical format to study predicted bit error rates.

To see how the bit error rate performance is affected by prediction errors, we will use (7.2.39) as the exact bit error rate probability for AWGN channels for this hypothetical format, even though it does not exactly follow the behaviour of any square MQAM format.

Using (7.2.39) in the expression for the predicted bit error rate (7.2.35), it turns out that the predicted bit error rate is here given by

$$\begin{aligned} P_{p,k}(\gamma_0 \hat{z}_{L|0}) &\approx \int_0^\infty P_{b,k}(\gamma_0 \hat{z}_{L|L}) \chi^2(\gamma_0 \hat{z}_{L|L}; \gamma_0 \hat{z}_{L|0}, \gamma_0 \sigma_{pf}^2) d(\gamma_0 \hat{z}_{L|L}) \\ &= \beta_1 \exp(-\beta_2 \gamma_0 \hat{z}_{L|0}), \end{aligned} \quad (7.2.40)$$

with $\beta_1 = \alpha_1 c(\sigma_{pf}^2)$ and $\beta_2 = \alpha_2 c(\sigma_{pf}^2)$, where the function $c(\sigma_{pf}^2)$ is given by

$$c(\sigma_{pf}^2) = (1 + \alpha_2 \sigma_{pf}^2)^{-1}. \quad (7.2.41)$$

Conveniently, (7.2.39) and (7.2.40) have the same form, and their respective coefficients differ only by a factor $c(\sigma_{pf}^2)$ which is independent of the prediction $\hat{z}_{L|0}$.

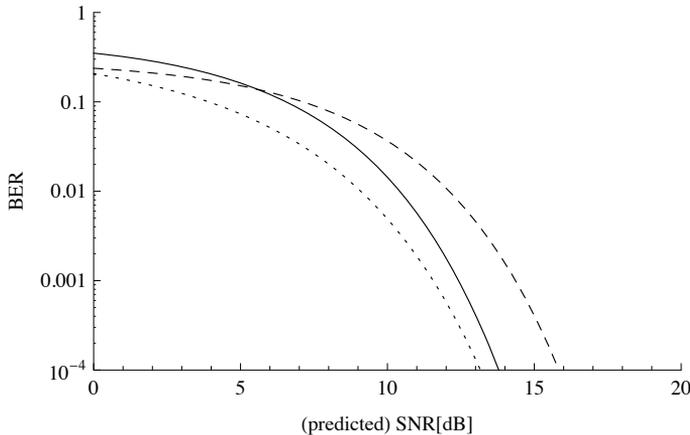


Figure 7.3: Approximative bit error rate performances for σ_{pf}^2 equal to 0 (solid) and 3 dB (dashed) for an approximation of QPSK. The dotted curve marks the exact BER curve for QPSK.

We now study the impact from varying the variance σ_{pf}^2 when α_1 and α_2 are fixed. We set $\alpha_1 = 0.5$ and $\alpha_2 = 1.6/(2^2 - 1)$, so that (7.2.39) is lower bounded by the bit error rate performance for QPSK, and examine how the predicted BER is affected when $\sigma_{pf}^2 = 0$ and $\sigma_{pf}^2 = 3$ dB, respectively. See Figure 7.3. Two mechanisms are at work. One is that a non-zero σ_{pf}^2 leads to $\beta_2 < \alpha_2$, so that (7.2.40) decreases more slowly than (7.2.39) when the SNR increases. This is to be expected, since the uncertainty on the SNR should have some negative impact on performance. The other effect is that a non-zero σ_{pf}^2 also leads to $\beta_1 < \alpha_1$, so that (7.2.40) is actually *lower* than (7.2.39) for low SNRs. This is perhaps a bit more surprising. How can it be that, at the lower end of the curve, it actually seems beneficial to be at a high level of uncertainty? Since a high σ_{pf}^2 may stem from ignorance on our part about the channel state, it would then seem that we could achieve a better error rate performance simply by discarding information! Surely this cannot be true.

Indeed it is not true. For an optimal filter, as was demonstrated in (7.2.26), it holds that the total power of the tracked signal, σ_h^2 , is divided between the error variance for the estimates, σ_f^2 , the variance of the future filtered estimates, σ_{pf}^2 , and the power of the prediction, σ_p^2 ,

$$\sigma_h^2 = \sigma_f^2 + \sigma_{pf}^2 + \sigma_p^2. \quad (7.2.42)$$

Therefore, when varying only one of the three variances, we are at the same time varying the channel power σ_h^2 . This is the reason why the error rate may decrease with an increased NMSE – the increased prediction NMSE

must come from an increased (true) signal power.

Although (7.2.39) may not be an appropriate approximation for MQAM modulation here, the qualitative behaviour of predicted bit error rate performance still holds; when prediction NMSE is increased, the bit error rate curve is shifted upwards towards higher SNRs at the same time as it goes down for low SNRs.

7.2.5 Mutual information for predicted channels

The second metric to be studied here is $\nu = \bar{I}_c$, where \bar{I}_c is the average mutual information (MI) per bit over the K resource blocks assigned to a user. As first noted in [89] and later demonstrated in a number of works (see e.g. [41],[90]), the error rate performance of efficient channel codes operating on soft decisions from the detector is determined by the mutual information between transmitted bits and the soft decisions, regardless of the modulation type used. This means that constraints on e.g. the codeword error probability can be translated to a constraint on the MI per bit.

The conditional mutual information for the channel between a transmitted bit and a continuous soft detector output, given that the soft output is z and the effective SNR is γ , is a function of z and γ that is here denoted $I_i(z, \gamma)$ and that is defined by²

$$\begin{aligned} I_i(z, \gamma) &= \sum_{b_i \in \{0,1\}} P(b_i) \log \frac{p(z, b_i | \gamma)}{p(z | \gamma) P(b_i)} \\ &= \sum_{b_i \in \{0,1\}} P(b_i) \log \frac{p(z | b_i, \gamma)}{\sum_{b'_i \in \{0,1\}} p(z | b'_i, \gamma) P(b'_i)}, \end{aligned} \quad (7.2.43)$$

where the index i indicates which bit in a symbols' bit pattern that is addressed, so that e.g. for 16-QAM, $i \in \{1, 2, 3, 4\}$. Note that $I_i(z, \gamma)$ is a function of both the SNR γ and the soft output z . The MI will generally be different for different bits in a symbol's bit pattern, so that, for an M -ary alphabet, $\log_2 M$ different channels per symbol need to be considered (see Figure 7.4).

For a symbol with M bits in a flat fading resource block we then have

²Mutual information conditioned on the SNR γ but unconditional on the channel's input is in the literature often denoted $I(X, Y | \gamma)$, where X and Y are "stochastic variables". Conditional mutual information, conditioned on that b was input to the channel and z was received, may be written $I(X = b, Y = z | \gamma)$. Here, we are interested in the mutual information conditioned only on that z was received, $I(X, Y = z | \gamma)$.

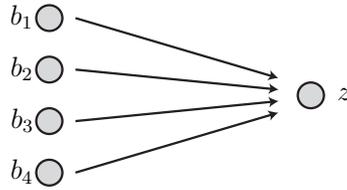


Figure 7.4: Channels from individual bits in a 16-ary symbol to the soft detector output z .

the average mutual information per bit

$$\nu_k = \bar{I}_c \triangleq (\log_2 M)^{-1} \sum_{i=1}^{\log_2 M} I_i(z, \gamma). \quad (7.2.44)$$

Since at $t = 0$, our state of knowledge will be the same for all symbols within an RB, the expression (7.2.44) can be used as local decision metric for a resource block.

The soft output z can be any of a number of different measures. The most common choice is to let z be the so called L-value, which is the logarithm of the odds for b_i taking either value, e.g.

$$z = \log \frac{P(b_i = 1|y, I)}{P(b_i = 0|y, I)}, \quad (7.2.45)$$

where y is the received signal. For this choice of z , exact and approximate expressions for $p(z|b_i, \gamma)$ can be found in [92] for AWGN channels when M-QAM symbol constellations are used³.

Alternatively, one may also simply let $z = y$, so that the mutual information is calculated directly from the received noisy signal y . Then $p(z|b_i, \gamma) = \sum_{s \in \chi_b^i} p(z|s, \gamma)P(s)$, where s is a hypothetical transmitted symbol and χ_b^i is the set of symbols in the symbol constellation whose i :th bit is b .

Below, we outline the link adaptation procedure for any metric ν that can be expressed as a sum of local metrics $\{\nu_k\}_{k=1}^K$, which is especially true for \bar{I}_c :

1. Calculate the pdf:s of the local average MI per bit $\{\nu_k\}_{k=1}^K$ through a change of variables:

$$p(\nu_k)|d\nu_k| = p(z|\gamma)p(\gamma) |dz d\gamma|, \quad (7.2.46)$$

³When the soft output is simplified to the so called max-log metric, expressions for $p(z|b_i, \gamma)$ can be found in [93] for Nakagami fading channels, but since at $t = L$ the receiver will effectively experience an AWGN channel, we have no use for such results here.

where $p(z|\gamma)$ is acquired as discussed above, and $p(\gamma)$ is the pdf for the future effective SNR, $p(\gamma_0 \hat{z}_{L|L} | \mathbf{Y}_0, I)$, see (7.2.28). Note that the pdf:s will depend on the local modulation format chosen.

2. Given an initial set of modulation formats $\{m_0, \dots, m_{J-1}\}$, calculate the pdf of the total metric $\nu = \frac{\sum_{k=1}^K m_k \nu_k}{\sum_{k=1}^K m_k}$.
3. Use decision theory to decide whether the pdf for ν with sufficient certainty guarantees that the condition on ν will be fulfilled when $t = L$.
4. Vary the transmission parameters, i.e. the local modulation formats, to vary the shape of $p(\nu)$, and the outer code rate, which changes the relation between ν and the prescribed CWER, so that the data rate is maximized.

The above procedure is impractical due to a number of reasons: i) the pdf:s of local MI per bit $\{\nu_k\}_{k=1}^K$ are very complicated to obtain, ii) the K RBs will often exhibit strong correlation, making the calculation of the pdf of ν difficult to carry out, iii) the pdf for ν may be irregular, making it hard to take a decision from it, and iv) solving the control problem, i.e. deciding which transmission parameters to use in order to maximize rate under a constraint on the pdf of ν , is extremely hard due to the complicated coupling from transmission parameters to the pdf of ν .

Steps 1 and 2 in the procedure outlined above can probably be replaced by sampling; by drawing a large number of samples from the distributions $p(z|\gamma)p(\gamma)$ for the respective RBs, a set of representative samples from ν can readily be calculated, so that an approximative distribution for ν is obtained. The question still remains, however, as to how the transmission parameters should be varied appropriately.

In Section 7.4, we shall attempt to use a simplified metric for link adaptation in a soft decoder scenario.

7.3 Adaptation of local CM formats

The optimization of local modulation formats has been solved for a number of different assumptions. In [32] (and references therein), rate optimization with perfect channel estimation and prediction is considered. In [64],[88], the corresponding problem for perfect channel estimation but imperfect channel prediction is presented. Here, we consider rate optimization when neither channel estimation nor channel prediction are error-free.

We attempt to maximize the rate

$$\sum_{j=0}^{J-1} \bar{m}_j \int_{b_j}^{b_{j+1}} p(\gamma_0 \hat{z}_{L|0}) d(\gamma_0 \hat{z}_{L|0}) \quad (7.3.1)$$

under the constraint that the average bit error rate (7.2.38) has a given value, which we here set to 10^{-3} . The $\{\bar{m}_j\}$ are the numbers of bits per symbol for the respective modulation formats, as in (7.2.38). Here we use BPSK, QPSK, 16QAM, and 64QAM. We thus optimize the rate limits of uncoded adaptive square MQAM under a bit error rate constraint. See Appendix 7.B for further details.

We know from (7.2.26) that the total channel power is divided between the estimation error variance σ_f^2 , the variance of the future filtered estimates σ_{pf}^2 , and the power of the predictions σ_p^2 . For a given SNR, i.e. a given channel power σ_h^2 , two degrees of freedom therefore exist. In Figure 7.5, we investigate three cases: perfect estimation and prediction (solid curve), error-free estimation and a prediction NMSE of 0.1 (dashed curve), and finally, a prediction NMSE of 0.1 with estimation error at 50% of the noise power (dotted curve). An NMSE of 0.1 is comparably low, while an estimation error of 50 % is quite reasonable. Considering the performance drop from the solid curve to the dashed curve, it is therefore evident from the figure that the prediction error dominates the total decrease in spectral efficiency. The impact from the estimation error is largely determined by the power of the total noise that σ_f^2 contributes to, and not so much by the decrease in σ_p^2 and/or the decrease in σ_{pf}^2 . For example, here we would anticipate the dashed and the dotted curves to be separated by $1.5 = 1.76$ dB, which is close to the truth.

To see how spectral efficiency is affected by a varying prediction range, we restudy Examples 4.2.1 and 4.2.2 from Chapter 4. In these examples, a two-tap channel is modelled, in which each tap is characterized by an AR2 model described by equations (4.2.17)–(4.2.19). The stabilizing solutions to the discrete algebraic Ricatti equation (DARE) for filtered estimates errors and prediction errors are calculated with (3.4.41) and (3.4.46), from which we can calculate the stationary $\Xi_{L|0}$ with Theorem 7.2.2. The triplet $\{\sigma_f^2, \sigma_{pf}^2, \sigma_p^2\}$ is then calculated according to the table at the end of Section 7.2.2.

The solid lines in Figure 7.6 illustrate the three variances for the velocity 72 km/h. The SNR is 10 dB as in Example 4.2.2. As seen from the lower line in Figure 7.7, which illustrates the spectral efficiency for a range of prediction horizons when uncoded adaptive MQAM transmission is used, the performance deterioration is quite dramatic in the beginning; the spectral efficiency falls to half its initial value in about 0.5 ms, which would here correspond to 0.1 wavelengths.

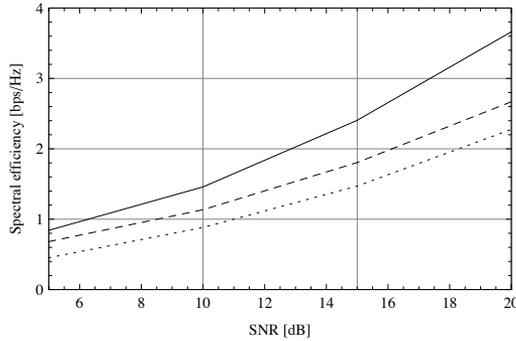


Figure 7.5: Spectral efficiency for different distributions of the total channel power. Solid: $\sigma_p^2 = \sigma_h^2$ (perfect estimation and prediction), dashed: $\sigma_p^2 = 0.9\sigma_h^2$ and $\sigma_{pf}^2 = 0.1\sigma_h^2$ (error-free estimation and a prediction NMSE of 0.1), and dotted: $\sigma_p^2 = 0.9\sigma_h^2 - 0.25$, $\sigma_{pf}^2 = 0.1\sigma_h^2 - 0.25$, and $\sigma_f^2 = 0.5$ (prediction NMSE=0.1 and estimation error at 50% of the noise power)

We also study link adaptation performance at pedestrian speed, which we here set to $v = 7.2$ km/h. The dashed lines in Figure 7.6 illustrate the three parameters $\{\sigma_f^2, \sigma_{pf}^2, \sigma_p^2\}$ for this velocity, and the spectral efficiency is given by the upper line in Figure 7.7. At a tenth of the previous speed, we see that the spectral efficiency decline is slower due to fact that the prediction ranges are shorter (in number of wavelengths), and also that the overall level is higher, since the estimation error is now lower.

It should however be pointed out that the Doppler spectrum in the studied example is rather broad due to the low model order ($=2$). This makes the channel inherently very hard to predict. With a narrower Doppler spectrum, e.g. resembling the Jakes Doppler spectrum or a unipolar line-of-sight spectrum, the deterioration in spectral efficiency with an increasing prediction horizon would be significantly slower.

7.4 Joint adaptation of local CM formats and outer code rate

The Mutual Information based Adaptive Modulation and Coding (MI-ACM) algorithm was proposed in the WINNER project[94] and was first suggested in [41]. The outer channel code is here derived from the so called *mother code*, which is designed to have a low code rate (1/2 or less). By puncturing the mother code, we may obtain an outer code with a desired rate. The

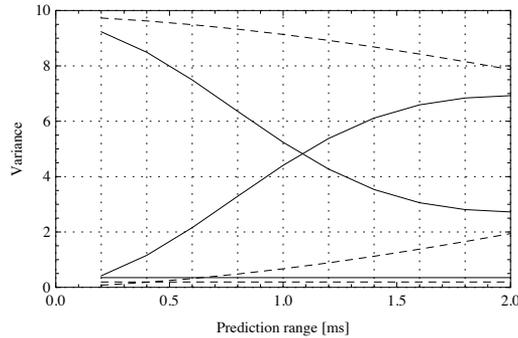


Figure 7.6: Prediction power σ_p^2 (upper curves at the lower end of the scale), future filtered estimate variance σ_{pf}^2 (middle curves at the lower end of the scale), and estimation error variance σ_f^2 (lower, horizontal curves) for Example 4.2.2 from Chapter 4. Solid curves correspond to $v = 72$ km/h, dashed to $v = 7.2$ km/h. The carrier frequency was $f_c = 3$ GHz, the pilot sampling period was $t_p = 0.2$ ms, and the SNR was 10 dB.

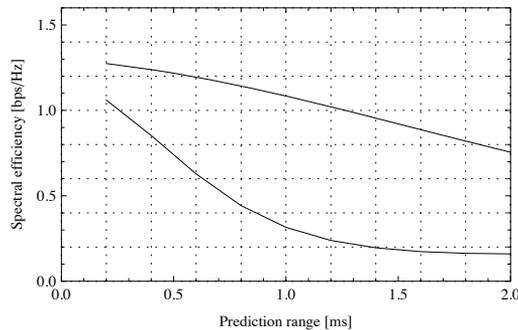


Figure 7.7: Spectral efficiency as a function of prediction range for the studied example. The upper curve corresponds to $v = 7.2$ km/h. The lower curve corresponds to $v = 72$ km/h. The carrier frequency was $f_c = 3$ GHz, the pilot sampling period was $t_p = 0.2$ ms, and the SNR was 10 dB.

algorithm aims to adapt transmission parameters for local resource blocks in conjunction with the rate of the outer code. This is generally a very difficult problem, but as we shall see, the proposed algorithm possesses a number of properties that will pave the way forward.

The code used in [41] is a Rate-Compatible Punctured Block-circulant Low Density Parity Check (RCP-BLDPC) code. LDPC codes have very good performance and can be efficiently implemented in hardware, which is the reason why they were chosen as *forward error correction* (FEC) scheme in the WINNER framework. Moreover, rate-compatible punctured codes keep a “priority list” that describes in which order bits should be punctured. This means that low priority bits are always punctured first, regardless of which code rate is desired. With an hybrid automatic repeat request (HARQ) retransmission scheme, extra parity bits can therefore be sent in the retransmissions if needed and be added to earlier codewords, hence decreasing the code rate and increasing probability for correct detection.

In a Block-LDPC code, the parity check matrix is built from quadratic blocks of size Z . These blocks can be either of two types:

$$\mathbf{0}_{Z \times Z} \quad \text{or} \quad \text{circ}([\mathbf{0}_{1,s}, 1, \mathbf{0}_{1,Z-s-1}]), \quad (7.4.1)$$

where s is the number of steps in a circular shift of an identity matrix of size Z . By conventionally letting $s = -1$ represent the zero matrix, a parity check matrix that would otherwise be very large can now be represented by a much smaller “base matrix”, in which each entry $s \in [-1, \dots, Z - 1]$ represents a $Z \times Z$ -block in the sparse parity check matrix. In the present investigation, the base matrix has dimensions 24×48 and $Z = 48$, so that for a rate $1/2$ code, a codeword is $48^2 = 2304$ bits long. By puncturing elements in the base matrix in the order given by the priority list, one can then produce codes of code rates from $24/48$ to $24/24$. For AWGN channels, the codeword error rates versus SNR for a range of code rates are illustrated in Figure 7.8.

The objective of the algorithm is to determine the modulation formats $\{m_k\}_{k=1}^K$ to be assigned to a set of K RBs. After that, as many codewords as possible should be loaded onto these RBs. How many codewords that can fit depends on the code rate R , which also needs to be determined by the algorithm. The modulation formats and the code rate should be jointly optimized to maximize spectral efficiency, i.e.

$$\eta = R \sum_{k=1}^K m_k \quad (7.4.2)$$

should be maximized. The maximization must be done under some constraint, as discussed in Section 7.1. Here, the constraint is that the codeword error rate $P_{cw}=0.01$.

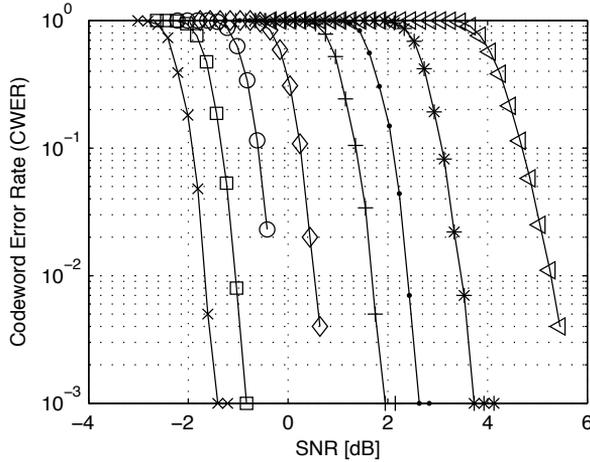


Figure 7.8: Codeword error rates versus SNR for the used RCP-BLDPC code, punctured to rates (from left to right) 0.50, 0.55, 0.60, 0.66, 0.75, 0.80, 0.86, and 0.92. An AWGN channel was used, and the codeword size at rate 0.50 is 2304 bits.

7.4.1 Perfect channel state information

The optimization of (7.4.2) under the constraint $P_{cw} = 0.01$ seems very difficult to carry out as it stands. As discussed in Section 7.1.1, a metric ν must be chosen, that correlates with the codeword error rate. This metric will be a function of the modulation formats $\{m_1, \dots, m_K\}$ as well as the code rate R , making (7.4.2) a multidimensional optimization problem. However, by imposing a number of assumptions on the metric ν , the rate optimization problem can be considerably simplified. The first assumption is as follows:

Assumption 7.1. *The codeword error rate should be a function of the metric ν and the code rate R of the outer code alone.*

It has been shown ([89],[41],[90]) that when bit-interleaved coded modulation (BICM) is used, the codeword error rate can be expressed as a function of the code rate R and the *average mutual information per bit* (MI per bit) \bar{I}_c ,

$$P_{cw} = P_{cw}(R, \bar{I}_c), \quad (7.4.3)$$

even when the channel is fading and multiple modulation formats are used within one codeword, meaning that $\nu = \bar{I}_c$ fulfills Assumption 7.1. The MI per bit is expressed as

$$\bar{I}_c = \frac{\sum_{k=1}^K J_m(m_k, \gamma_k)}{\sum_{k=1}^K m_k}, \quad (7.4.4)$$

where $J_m(m, \gamma)$ is the mutual information per symbol for bit-interleaved channels for constellation size 2^m and SNR γ [41]. As will be evident presently, we will not be needing an explicit expression for $J_m(m, \gamma)$. From (7.4.4), it follows that $\nu = \bar{I}_c$ fulfills the second assumption:

Assumption 7.2. *The metric ν should be an arithmetic mean value of local metrics $\{\nu_k\}_{k=1}^K$.*

Since the SNR may vary within one RB it is necessary to use some sort of mean SNR value. The mean value used here and in [41] is

$$\gamma = \beta f^{-1} \left(n_s^{-1} \sum_{s=1}^{n_s} f(\gamma_s) \right) + (1 - \beta) \min_s \gamma_s, \quad f(x) = \log_2(1 + x), \quad (7.4.5)$$

where s and n_s are the index for, and number of, symbols in an RB, respectively. That is, AWGN capacity is used for averaging and RBs with high SNR variations may be penalized by appropriately choosing the value β .

Choosing the specific ‘‘contour’’ $P_{cw}=0.01$ in (7.4.3), we construct the implicit function $R_{0.01}(\bar{I}_c)$, which for a given \bar{I}_c tells which rate is needed to achieve $P_{cw}=0.01$. Inserting $R = R_{0.01}(\bar{I}_c)$ into (7.4.2), we get

$$\eta = R_{0.01}(\bar{I}_c) \sum_{k=1}^K m_k, \quad (7.4.6)$$

but this in itself does not make the optimization problem any easier. To simplify the problem, we need the third assumption:

Assumption 7.3. *For the required condition, e.g. $P_{cw} = 0.01$, there should be an approximate linear relationship between ν and the code rate R of the outer code.*

It turns out that $R_{0.01}$ is indeed approximately linear in \bar{I}_c , which means that by (7.4.4),

$$R_{0.01}(\bar{I}_c) \approx \frac{\sum_{k=1}^K m_k R_{0.01}(J_m(m_k, \gamma_k)/m_k)}{\sum_{k=1}^K m_k}. \quad (7.4.7)$$

The factors $R_{0.01}(J_m(m_k, \gamma_k)/m_k)$ here appear as ‘‘local’’ code rates, i.e. they have the same form as the total code rate $R_{0.01}(\bar{I}_c)$, but they depend each only on the SNR for a single RB. The spectral efficiency can now be written

$$\eta = \sum_{k=1}^K m_k R_{local}(m_k, \gamma_k), \quad R_{local}(m_k, \gamma_k) = R_{0.01}(J_m(m_k, \gamma_k)/m_k), \quad (7.4.8)$$

which means that the total spectral efficiency is maximized when the local efficiencies on each RB are maximized.

Since there is only a discrete number of modulation formats available, local optimization over an RB is performed by choosing the modulation format m that gives the highest product $mR_{local}(m, \gamma)$. Simulations of all the modulation formats over a wide range of SNR values show that it always pays off to choose the highest modulation format, i.e. if there at a certain SNR is a choice between selecting a low modulation format in combination with a high rate and a higher modulation format at a lower rate, then the latter will always give the highest spectral efficiency.

Local optimization is therefore carried out by first selecting the modulation format m that can meet the P_{cw} constraint at the present SNR γ for a code with the lowest rate $1/2$, and then looking up the local code rate $R_{local}(m, \gamma)$ from simulated curves for AWGN channels.

The total code rate $R = R_{0.01}(\bar{I}_c)$ can then be calculated from

$$R = \frac{\sum_{k=1}^K m_k R_{local}(m_k, \gamma_k)}{\sum_{k=1}^K m_k}. \quad (7.4.9)$$

From the K local format decisions and the total code rate R , the number of codewords that can be interleaved, properly punctured, and loaded onto the K RBs can now be determined.

We evaluate the algorithm on a synthetic channel described by the WINNER B1 NLOS channel model (see Section 2.3.2). The velocity was 10 km/h and the Doppler spectrum had broad peaks at $\pm f_D$, where f_D is the maximum Doppler frequency. Figure 7.9 shows the result. The algorithm manages to fulfill the constraint $P_{cw} = 0.01$ for nearly all SNRs while maintaining a high spectral efficiency, which is in accordance with the results in [41].

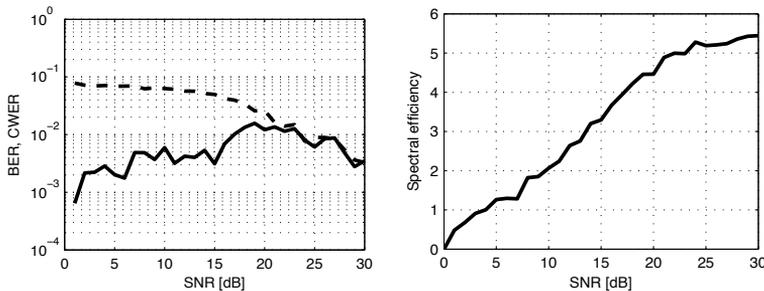


Figure 7.9: Codeword error rate (solid line, left figure) and bit error rate before decoder (dashed line, left figure), and spectral efficiency (right figure) for a known synthetic channel.

We now reiterate the same experiment on a measured channel. Details regarding the measurements will be presented in Section 8.2. See Figure 7.10. Here, the codeword error rate increases far beyond the acceptable

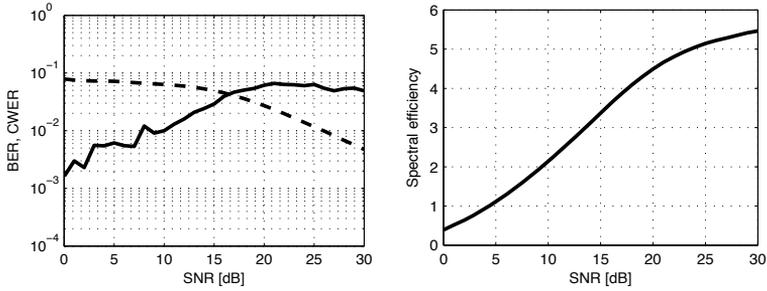


Figure 7.10: Codeword error rate (solid line, left figure) and bit error rate before decoder (dashed line, left figure), and spectral efficiency (right figure) for a known measured channel.

level for high SNRs. There can be a number of reasons for this, one being the possibility that the fading within the RBs is severe. To diagnose, we conduct the same experiment again, but now the measured channel has been preprocessed so that the fading within RBs is flat. See Figure 7.11. This

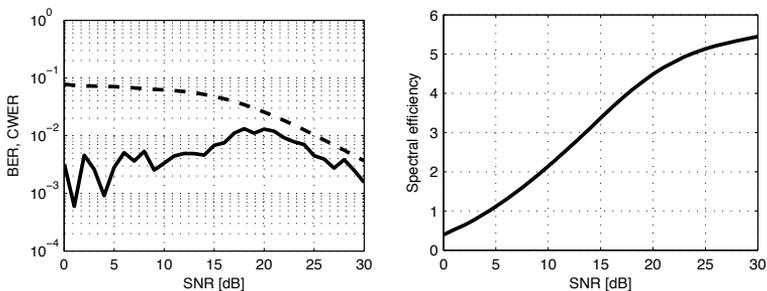


Figure 7.11: Codeword error rate (solid line, left figure) and bit error rate before decoder (dashed line, left figure), and spectral efficiency (right figure) for a known measured and block fading channel.

remedies the problem, which makes it clear that the factor β in (7.4.5) needs to be set appropriately to penalize fluctuations within the blocks.

7.4.2 Imperfect channel state information

The algorithm described in the previous section chooses the transmission parameters $\{m_k\}$ and R based on exact channel state information. But how do we choose transmission parameters when the SNRs $\{\gamma_k\}$ are uncertain? As we saw in Section 7.2.5, the theoretically correct procedure is impractical. The MI-ACM algorithm based on the MI per bit suggested in the previous section simplifies the link adaptation problem so that it reduces to solving many local optimizations instead of one global optimization. Unfortunately this simplification does not carry across to the case with imperfect channel predictions, because the local metrics are then impaired with uncertainties so that local optimizations cannot easily be carried out. We will therefore attempt to use an alternative to the mutual information metric $\nu = \bar{I}_c$ here.

As demonstrated in Section 7.4.1, the average MI per bit obeys a number of assumptions that simplify the link adaptation problem considerably. We here seek an alternative metric ν that adheres to the same set of properties. The metric we propose to use here is

$$\nu = 1 - 2\bar{P}_c, \quad (7.4.10)$$

where \bar{P}_c is the hard bit error rate (before the decoder). For the metric (7.4.10), Assumption 7.2 is clearly met if the local metrics are defined as the hard bit error rates for the respective RBs. It is also reasonable to think that Assumption 7.1 will be met; since BICM is used, the interleaver scrambles the bits so that it after hard detection will be more or less indistinguishable from an AWGN channel. The channel is then characterized by one parameter alone, for example the bit error rate, or an affine mapping thereof. It should be noted that because the channel decoder used here operates on soft decisions, not all characteristics of the fading are lost in the interleaving. We expect, however, that in the present context the channel will be adequately characterised by the hard bit error rate alone.

Assumption 7.3 is investigated in Figure 7.12. Here we have calculated bit error rate curves for all four modulation formats considered, as well as for a large number of code rates. Noting for which SNR value each curve intersects the P_{cw} constraint and converting this SNR value to a corresponding ν value, Figure 7.12 is produced. The contour $P_{cw} = 0.01$ is plotted for all four modulation types. As is clear from the figure, the ν - R correspondence is more or less independent of the modulation format, which is desirable. The relation is approximately affine, but linearity is required to simplify link adaptation. The metric $\nu = 1 - 4P_c$ would have been a better choice in this respect, but in fact it does not matter; it suffices that there *exists* an affine mapping from P_c to R that satisfies the three conditions for us to use any affine mapping as metric ν . The reason we choose $\nu = 1 - 2P_c$ is that then,

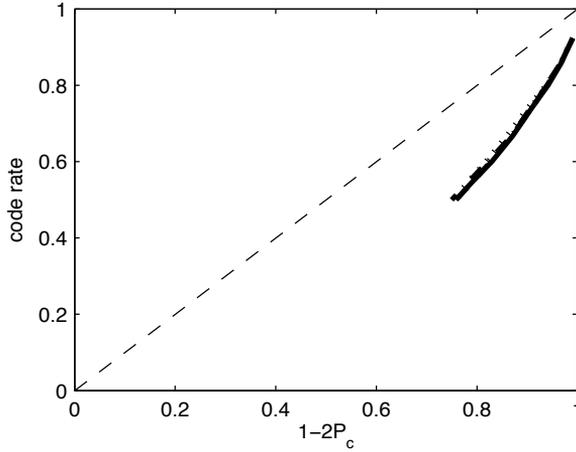


Figure 7.12: code rate R versus the metric $\nu = 1 - 2\bar{P}_c$ required to fulfil the constraint $P_{cw} = 0.01$. Results are plotted for all four modulation formats, but the curves more or less overlap.

$\nu \in [0, 1]$, which also holds for $\nu = \bar{I}_c$. Although the curve in Figure 7.12 is not quite straight, the deviation from a straight line is approximately as that for the \bar{I}_c - R curve in [41].

Expressions for predicted bit error rates over infinitely many resource blocks ($K = \infty$) were derived in Section 7.2.4, with details described in Appendix 7.B. In order to reuse these results, we assume that the number of resource blocks K allotted to the current user is very large.

We conduct an experiment with this choice of decision metric on a synthetic channel and prediction range $1988\mu\text{s}$. The channel has a power delay profile according to the WINNER II B1 NLOS channel model (see Section 2.3.2). The pilots are spaced $284\mu\text{s}$ in time and 60 kHz apart in frequency. The Doppler spectrum has broad peaks at $\pm f_D$, where f_D is the “maximum” Doppler frequency, as described in Section 2.3.1. The velocity is $v = 10\text{ km/h}$, and the carrier frequency is 2.66 GHz.

A pilot-assisted Kalman filter produces channel predictions and estimates. For each RB, at time $t = 0$ a prediction $\gamma_0 \hat{z}_{L|0}$ of the SNR is produced. Local bit error rates, one per modulation format, are then calculated from the appropriate predicted bit error rate expressions, and converted to $1 - 2P_c$ -values. From these values, the local code rates can then be determined from Figure 7.12. The combination of modulation format and local code rate that gives the highest spectral efficiency is then determined for the respective RB. Finally, the global code rate R can be determined from the local code rates, see (7.4.9). As in the previous experiment, the number of codewords that

can be interleaved, properly punctured, and loaded onto the K RBs can now be determined from the K local format decisions and the total code rate R .

The results are depicted in Figure 7.13. The constraint $P_{cw} = 0.01$ is almost met, while preserving the spectral efficiency on a high level. This indicates that the proposed algorithm is functional in the low velocity scenario investigated here. However, we expect that at higher velocities, a more proper implementation of the procedure outlined on page 173 should be used in conjunction with a metric based on mutual information.

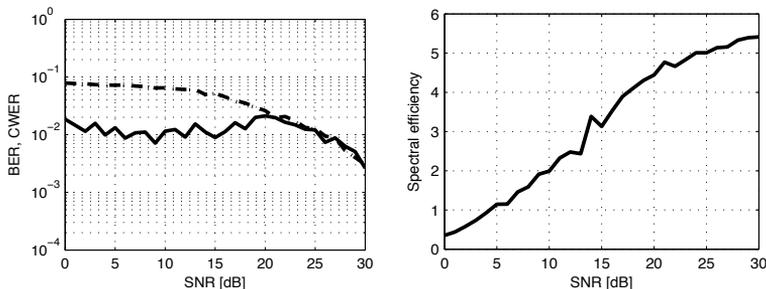


Figure 7.13: Codeword error rate (solid line, left figure) and bit error rate before decoder (dashed line, left figure), and spectral efficiency (right figure) for a predicted synthetic channel.

7.5 Summary

We have in this chapter considered link adaptation when the system is impeded by imperfect channel estimation and prediction. It was emphasized that the object of prediction should be an *effective* SNR, characterized by future filtered estimates of channel coefficients rather than the actual coefficients themselves, and by a noise level dictated not only by thermal noise and interference, but also by the channel estimation error variance. An expression was derived for the pdf of the predicted effective SNR. Although we have in this chapter only considered single-input links, this expression is valid for general multi-input channel models.

Link adaptation amounts to maximizing e.g. data transfer rate. The maximization is subject to one or many constraints on system parameters. In order to fulfil these constraint(s), a decision metric that correlates well with the constraint(s) needs to be established. Two types of link adaptation were considered: adaptation of local modulation formats, in which we used the average bit error rate after hard detection as decision metric, and joint

adaptation of local modulation formats and the rate of an outer channel code. Here we used a decision metric based on average mutual information per bit.

In the former type of adaptation, we showed that uncertainties, which normally are associated with channel prediction, vanish when we let the number of resources over which we adapt go to infinity. Link adaptation performance was studied in a few example systems. However, we stressed that when the adaptation is to be carried out on only a finite number of resources at a time, then uncertainty is an unavoidable and integral part of the decision problem, so that decision theory necessarily has to be used to solve the link adaptation problem. For the latter type of adaptation, we outlined a procedure for how to take channel prediction and estimation uncertainty into account. Although correct in principle, we stressed that the proposed algorithm is difficult to implement. We studied link adaptation performance for the idealized case of perfect channel prediction, and proposed and evaluated a simplified scheme for how to take prediction uncertainty into account. This scheme proved to work for low velocities.

7.6 Acknowledgements

The author would like to express his gratitude towards Annea Barkefors, who performed the main part of the experimental work in Section 7.4.

7.A Signal/noise correlation for decision feedback channel estimation

In Section 7.2.3 it was shown that channel estimation errors give rise to a total noise that has variance $S\sigma_f^2 + \sigma_v^2$, where S is the average symbol energy, σ_f^2 is the channel estimation error variance, and σ_v^2 is the thermal noise and interference. This result rests on the fact that the channel estimation is based on a nearby pilot rather than on the payload symbol at hand, which makes the additive noise and the filtered channel estimate uncorrelated.

When iterated channel estimation is used, detected symbols are fed back into the channel estimator. When symbols are correctly detected, they act as regular pilots. To analyze the magnitude of the total effective noise in this case, we consider the state-space system (7.2.1), where, in the ideal case, Φ_t is diagonal with known diagonal elements. The Kalman filter produces optimal filtered state estimates $\hat{\mathbf{x}}_{t|t}$ so that

$$\mathbf{y}_t = \Phi_t \mathbf{H} \hat{\mathbf{x}}_{t|t} + \Phi_t \mathbf{H} \tilde{\mathbf{x}}_{t|t} + \mathbf{v}_t, \quad (7.A.1)$$

where $\hat{\mathbf{x}}_{t|t}$ is the filtered estimate of the state vector, and $\tilde{\mathbf{x}}_{t|t}$ is the estimation error. The two last terms, $\Phi_t \mathbf{H} \tilde{\mathbf{x}}_{t|t} + \mathbf{v}_t$, are inherently unknown and therefore constitute the total noise. From the relation $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_t + \mathbf{K}_{f,t}(\mathbf{J}_t \tilde{\mathbf{x}}_t + \mathbf{v}_t)$, where $\hat{\mathbf{x}}_t$ and $\tilde{\mathbf{x}}_t$ are the one-step state predictions and the one-step state error, and $\mathbf{K}_{f,t}$ is the Kalman filter gain, it follows that $\langle \hat{\mathbf{x}}_{t|t}, \mathbf{v}_t \rangle = \mathbf{K}_{f,t} \mathbf{R}$, where $\mathbf{R} = \|\mathbf{v}_t\|^2$, since $\langle \hat{\mathbf{x}}_t, \mathbf{v}_t \rangle = \mathbf{0}$ and $\langle \tilde{\mathbf{x}}_t, \mathbf{v}_t \rangle = \mathbf{0}$. It also holds that $\langle \tilde{\mathbf{x}}_{t|t}, \mathbf{v}_t \rangle = \langle \mathbf{x}_t - \hat{\mathbf{x}}_{t|t}, \mathbf{v}_t \rangle = -\langle \hat{\mathbf{x}}_{t|t}, \mathbf{v}_t \rangle$. It is now easy to verify that the terms in (7.A.1) are correlated according to

$$\left\| \begin{bmatrix} \Phi_t \mathbf{H} \hat{\mathbf{x}}_{t|t} \\ \Phi_t \mathbf{H} \tilde{\mathbf{x}}_{t|t} \\ \mathbf{v}_t \end{bmatrix} \right\|^2 = \begin{pmatrix} \Phi_t \mathbf{H} \Sigma_{t|t} \mathbf{H}^* \Phi_t^* & \mathbf{0} & \Phi_t \mathbf{H} \mathbf{K}_{f,t} \mathbf{R} \\ \mathbf{0} & \Phi_t \mathbf{H} \mathbf{P}_{t|t} \mathbf{H}^* \Phi_t^* & -\Phi_t \mathbf{H} \mathbf{K}_{f,t} \mathbf{R} \\ \mathbf{R} \mathbf{K}_{f,t}^* \mathbf{H}^* \Phi_t^* & -\mathbf{R} \mathbf{K}_{f,t}^* \mathbf{H}^* \Phi_t^* & \mathbf{R} \end{pmatrix}. \quad (7.A.2)$$

The total noise at detection has variance given by the diagonal elements of the sum of the four matrices in the lower right corner of (7.A.2),

$$\Phi_t \mathbf{H} \mathbf{P}_{t|t} \mathbf{H}^* \Phi_t^* - \mathbf{R} \mathbf{K}_{f,t}^* \mathbf{H}^* \Phi_t^* - \Phi_t \mathbf{H} \mathbf{K}_{f,t} \mathbf{R} + \mathbf{R}, \quad (7.A.3)$$

and not the diagonal elements of just

$$\Phi_t \mathbf{H} \mathbf{P}_{t|t} \mathbf{H}^* \Phi_t^* + \mathbf{R} \quad (7.A.4)$$

as would be the case when ICE is not used. The situation is further complicated by the fact that when ICE is used, the utility signal $\Phi_t \mathbf{H} \hat{\mathbf{x}}_{t|t}$ is correlated with \mathbf{v}_t .

7.B Rate limit optimization

We wish to maximize the data rate in an uncoded adaptive modulation system based on uncertain channel predictions. Based on past measurements \mathbf{Y}_0 , at time $t = L$ the channel coefficients' effective SNR $\gamma_0 \hat{z}_{L|L}$ has the distribution (7.2.28),

$$p(\gamma_0 \hat{z}_{L|L} | \mathbf{Y}_0, I) = \chi^2(\gamma_0 \hat{z}_{L|L}; \gamma_0 \hat{z}_{L|0}, \gamma_0 \sigma_p^2), \quad (7.B.1)$$

and the predicted SNR values $\gamma_0 \hat{z}_{L|0}$ are frequency distributed according to (7.2.29),

$$p(\gamma_0 \hat{z}_{L|0} | \gamma_0 \sigma_p^2, I) = \chi^2(\gamma_0 \hat{z}_{L|0}; 0, \gamma_0 \sigma_p^2), \quad (7.B.2)$$

where we for clarity have conditioned on $\gamma_0 \sigma_p^2$, although $\gamma_0 \sigma_p^2$ is given by the model, i.e. the prior information I . Above, $\gamma_0 = S/(S\sigma_f^2 + 1)$, where S is the average symbols energy and we have assumed that the thermal noise and interference has unit variance. The three variances σ_f^2 , σ_{pf}^2 , and σ_p^2 are the error variance for the filtered estimates, the variance for the future filtered estimate, and the variance of the channel predictions, respectively. They are obtained from the Kalman filter.

Data rate maximization is carried out by optimizing the levels of the predicted SNR at which the modulation formats are switched. Here we assume that the modulation formats are BPSK, QPSK, 16QAM, and 64QAM. Defining

$$f(S, b_0, b_1, b_2, b_3) = \sum_{j=0}^{J-1} \bar{m}_j \int_{b_j}^{b_{j+1}} p(\gamma_0 \hat{z}_{L|0} | \gamma_0 \sigma_p^2 I) d(\gamma_0 \hat{z}_{L|0}), \quad (7.B.3)$$

we want to maximize the rate $f(S, b_0, b_1, b_2, b_3)$ with respect to the integration limits $\{b_0, b_1, b_2, b_3\}$ and the transmitted energy S subject to a fixed value P_{tgt} on the bit error rate (7.2.38). This constraint can be expressed as

$$g(b_0, b_1, b_2, b_3) = \sum_{j=0}^{J-1} \bar{m}_j \int_{b_j}^{b_{j+1}} (P_{p,j}(\gamma_0 \hat{z}_{L|0}, \gamma_0 \sigma_{pf}^2) - P_{tgt}) p(\gamma_0 \hat{z}_{L|0} | \gamma_0 \sigma_p^2, I) d(\gamma_0 \hat{z}_{L|0}) = 0, \quad (7.B.4)$$

where

$$P_{p,j}(\gamma_0 \hat{z}_{L|0}, \gamma_0 \sigma_{pf}^2) = \int_0^\infty P_{b,j}(\gamma_0 \hat{z}_{L|L}) \chi^2(\gamma_0 \hat{z}_{L|L}; \gamma_0 \hat{z}_{L|0}, \gamma_0 \sigma_{pf}^2) d(\gamma_0 \hat{z}_{L|L}), \quad (7.B.5)$$

and \bar{m}_j is the number of bits per symbol for modulation format j . Here, $\bar{m}_j \in \{1, 2, 4, 6\}$. For the bit error rate expressions $P_{b,j}(\gamma)$ for the respective modulation formats, we here use [71]

$$P_{b,j}(\gamma) = \begin{cases} 0.5 \operatorname{erfc}(\sqrt{\gamma}) & j = 1, \\ 0.5 \operatorname{erfc}(\sqrt{\gamma/2}) & j = 2, \\ \frac{1 - \left(1 - \left(1 - \frac{1}{\sqrt{2\bar{m}_j}}\right) \operatorname{erfc}\left(\sqrt{\frac{3}{2}} \sqrt{\frac{\gamma}{2\bar{m}_j - 1}}\right)\right)^2}{\bar{m}_j} & j \in \{3, 4\}, \gamma > \gamma_c. \end{cases} \quad (7.B.6)$$

For $j \in \{3, 4\}$ we use the above expression only above a certain cutoff SNR γ_c , since they show poor accuracy for low SNRs. We here need to marginalize bit error probability expression over χ^2 -distributions that may have considerable support at low SNRs, making it crucial to use bit error rate expressions that are accurate for any SNR. For SNRs below the cutoff value γ_c , which we set to 15 dB for 16QAM and 22 dB for 64QAM, we use numerical approximations based on tables acquired by simulation. When $\gamma < -30$ dB, we set $P_{p,j}(\gamma) = 0.5$.

According to the method of Lagrange multipliers, we solve the set of equations $\nabla_{b_0, b_1, b_2, b_3, \lambda}(f + \lambda g) = 0$:

$$P_{p,0}(b_0, \gamma_0 \sigma_{pf}^2) = P_{tgt} - \lambda^{-1}, \quad (7.B.7)$$

$$\begin{aligned} (\bar{m}_{j-1} - \bar{m}_j)^{-1} (\bar{m}_{j-1} P_{p,j-1}(b_j, \gamma_0 \sigma_{pf}^2) - \bar{m}_j P_{p,j}(b_j, \gamma_0 \sigma_{pf}^2)) \\ = P_{tgt} - \lambda^{-1}, j = 1, \dots, J-1, \end{aligned} \quad (7.B.8)$$

$$g(b_0, b_1, b_2, b_3) = 0. \quad (7.B.9)$$

Since the channel is in outage when the predicted SNR is below b_0 , the transmitted energy S must be adjusted with respect to this lower switching level. Therefore we first optimize b_0 and S jointly:

$$\begin{cases} \int_0^\infty P_{b,0}(\gamma) \chi^2\left(\gamma; b_0, \frac{S}{S\sigma_f^2 + 1} \sigma_{pf}^2\right) d\gamma = P_{tgt} - \lambda^{-1}, \\ S \int_{b_0}^\infty \exp\left(-\hat{\gamma} / \frac{S\sigma_p^2}{S\sigma_f^2 + 1}\right) / \frac{S\sigma_p^2}{S\sigma_f^2 + 1} = 1, \end{cases} \quad (7.B.10)$$

where the integral in the second equation is the fraction of time when the system is not in outage. Once the solution to this system of equations has been found, we can form $\gamma_0 = S/(S\sigma_f^2 + 1)$. After that, equations (7.B.8) and (7.B.9) are solved.

Studies on measured channels

8.1 Introduction

In this chapter, we conduct a number of studies on measurements of wideband fading MIMO radio channels at 2.66 GHz. We estimate model parameters from noisy and noise-free measurements, and study how the acquired channel models can be used in a Kalman filter for channel estimation and prediction. The aim is to see to what extent channel model parameters can be estimated in a realistic, noisy setting, since in a real system, good estimates of channel model parameters are required for efficient channel estimation/equalization and prediction. The chapter is organized as follows.

In Section 8.2, details regarding the recording of the channels are given. Section 8.3 is concerned with the modelling of a channel's time dynamics. In Chapter 4, it was suggested that a fading channel coefficient be modelled by an autoregressive (AR) model of fixed order. The following aspects of the AR-model must be considered: What model order should be used? Over how long a time period can static model parameters be used? What methods should be used to acquire the AR model parameters? How do we handle noisy measurements when it comes to estimating AR parameters? These questions and more are treated in Section 8.3.

In Section 8.4, we complement the results in Section 8.3 by also studying estimation and prediction of parallel subchannels and simultaneous inputs from multiple transmit antennas. We study how performance depends on SNR. When multiple inputs are tracked, we also consider whether it suffices to estimate AR parameters from one of the inputs and then use the same model for all inputs, or whether individual parameter estimation has to be used for each input.

In this thesis we use Kalman filters to track fading radio channels in a number of scenarios. When models with static parameters are used to

represent the fading characteristics, the Kalman filter will settle to a periodic state, drastically reducing the numerical complexity compared to the non-stationary case, as was shown in Chapter 4. The rate at which the filter converges to periodicity will largely impact the energy consumption of the filter, since the updating of large covariance matrices can be turned off once periodicity has been reached. In Section 8.5, we study the convergence rate for filters operating on some typical channel models.

Section 8.7 concludes the Chapter.

8.2 Measurements

The measurements that we will study here were conducted by Ericsson Research in November 2008 in the Stockholm suburb Kista. The measurements consist of up to 480 seconds long channel data series of a fading 8-by-4 MIMO-OFDM channel. We denote such a series a *route*. A route is divided into at most 90000 slots of 5.33 ms each. The slots, in turn, are divided into 8 frames of 667 μ s each. For each transmitting antenna element, measurements are taken only during one frame in each slot. The frame consists of 9 OFDM symbols, and an OFDM symbol has 66.7 μ s for data and 7.41 μ s for the cyclic prefix, the latter of which is removed after reception. The cyclic prefix admits a difference in propagation path lengths of just above 2 km. The OFDM symbol duration of 66.7 μ s gives a subchannel bandwidth of 15 kHz, which is the same as in the LTE standard.

Eight roof-mounted transmitting antenna elements (AE) and four receiving AEs mounted on the roof of the measuring vehicle were used. Short channel snapshots were taken once per slot (5.33 ms). A snapshot is 1296 samples long (excluding cyclic prefix) and sampled at 19.44 MHz in the baseband, hence representing 66.7 μ s. To avoid inter-carrier interference, each of the eight channels sampled at each receiving AE used a dedicated frame in the slot.

The stationary transmitting equipment and the receiving equipment in the measurement vehicle were initially synchronized by rubidium clocks to high accuracy. Training symbols known to the receiver were transmitted from the transmitting AEs. Apart from being spread out in time, the transmissions were also separated in frequency, so that each transmitting AE used only every eighth subchannel for transmission, i.e. the channels were sampled once per 120 kHz. After the channel snapshots were collected from the mobile receiving AEs, the cyclic prefixes were removed and the remaining parts Fourier transformed and element-wise divided by the known sequence of symbols, yielding the complex channel coefficients that we here consider to be noiseless.

The recorded complex baseband channel coefficients $\{h_{f,t}\}$, where t is the time index and f is the frequency index, then consists of 32 matrices (one per transmitter–receiver antenna pair), each of dimension $1296/8=162$ pilot subchannels and a very large number of OFDM symbols.

The measured channels discussed here have been studied perviously, with respect to direction-of-arrival and polarization diversity [95], and the potential benefits of cooperative MIMO [96]. In this chapter, we study the time-dynamical properties of the measured channels, as well as the prospect of tracking several channels at the same time.

Because of the sparse sampling, the raw channels will be interpolated in this study, both in time and in frequency, by *upsampling factors* of the user's choice. Interpolation by 2D-DFT, then zero padding, then 2D-IDFT is used, meaning that the interpolated channel matches the original exactly at the sampling points, and that no frequency information is added.

An important aspect to consider when dealing with modelling of measured channels is how often the channel model needs to be updated. In this chapter we will use channel models that are held constant over short periods of time. The length of such a period is denoted an *evaluation segment* and we talk about *segmentation* of the measured channel. Each evaluation segment is preceded by a *training segment*. Here we will let training segments and evaluation segments have the same lengths. The segmentation will be on the order of seconds or fractions of a second.

8.3 Time dynamics

In Chapter 4, it was suggested that a fading channel coefficient h_t be modelled by an autoregressive (AR) model

$$h_t + a_1 h_{t-1} + \dots + a_k h_{t-k} = u_t, \quad (8.3.1)$$

where k is the model order, $\{a_1, \dots, a_k\}$ are the AR parameters, and u_t is the process noise of variance σ_u^2 that excites the process. It was pointed out that the model order impacts the numerical complexity of the channel estimator/predictor. This fact should be considered when choosing the model order. Furthermore, two approaches for choosing the AR parameters were suggested, that were both motivated by the presence of only vague prior information about the fading characteristics of the channel. One produces a flat Doppler spectrum, based on the knowledge that an upper limit to the Doppler spectrum exists and is known. The other produces a Doppler spectrum resembling the Jakes' spectrum, based on the knowledge that there is no line-of-sight between transmitter and receiver, and that one or several

independent scatterers are distributed at unknown angles around the mobile unit.

When actual channel measurements, albeit noisy, are available, channel models based on more cogent prior information may be used. Assuming first that noise-free channel measurements are available, we define the vector of AR parameters

$$\mathbf{a} = [a_1, \dots, a_k]^T, \quad (8.3.2)$$

and the set of all past channel coefficients

$$\mathbf{h}_t = \{h_{t-1}, \dots, h_0\}. \quad (8.3.3)$$

Equipped with the model (8.3.1) and assuming that the model order k and the process noise variance σ_u^2 are known, the correct Bayesian procedure to infer a channel coefficient h_t from past coefficients \mathbf{h}_t would be to assign a prior distribution to \mathbf{a} and then marginalize over the unknown parameters in \mathbf{a} :

$$p(h_t|\mathbf{h}_t, I) = \int p(h_t|\mathbf{h}_t, \mathbf{a}, I)p(\mathbf{a}|\mathbf{h}_t, I)d\mathbf{a}. \quad (8.3.4)$$

If both distributions in the integral are Gaussian, then a closed form expression for $p(h_t|\mathbf{h}_t, I)$ may easily be found. The first distribution follows directly from (8.3.1):

$$p(h_t|\mathbf{h}_t, \mathbf{a}, I) = \mathcal{CN}(h_t; \mu_h, \sigma_u^2) \quad (8.3.5)$$

with $\mu_h = -[h_{t-1} \dots h_{t-k}]\mathbf{a}_t$. The second, the AR parameter distribution $p(\mathbf{a}|\mathbf{h}_t, I)$, may be a prior distribution $p(\mathbf{a}|I)$ given by some principle for deriving prior distributions (see Section 3.2.3). Alternatively, if the AR parameters \mathbf{a}_t are time-varying and correlated with \mathbf{h}_t , then its pdf should be propagated as time evolves. For example, assuming that a state-space model well represents the dynamics of \mathbf{a}_t ,

$$\begin{aligned} \mathbf{a}_{t+1} &= \mathbf{F}\mathbf{a}_t + \mathbf{G}\mathbf{w}_t, \\ h_t &= -[h_{t-1} \dots h_{t-k}]\mathbf{a}_t + u_t, \end{aligned} \quad (8.3.6)$$

for some matrices \mathbf{F} and \mathbf{G} , and some variance for the process noise \mathbf{w}_t , the distribution $p(\mathbf{a}_t|\mathbf{h}_t, I)$ is calculated by means of the Kalman filter recursions.

This is to say: if noise-free channel measurements \mathbf{h}_t would be available, then we may marginalize over the unknown AR parameters, hence conducting optimal inferences about h_t without making unsupported assumptions about \mathbf{a} . However, if noise-free measurements are *not* present, then the expected value $-[h_{t-1} \dots h_{t-k}]\mathbf{a}_t$ of h_t cannot be formed, neither in the marginal distribution $p(h_t|\mathbf{h}_t, \mathbf{a}, I)$ nor in the state-space model for \mathbf{a}_t . In this case, which is the case we are facing in real scenarios, marginalization over \mathbf{a} is computationally prohibitive. It is then necessary to fixate the AR parameters by

choosing a point estimate $\hat{\mathbf{a}}$ of \mathbf{a} , so that (8.3.1) is linear in h_t . A state-space model of the fading radio channel(s) can then be constructed according to the framework outlined in Chapter 4. Then, the channel h_t may be estimated and/or predicted with a Kalman filter. It is therefore necessary to have efficient methods for estimating model parameters.

This section is devoted to various aspects of AR parameter estimation, such as model order selection, parameter estimation, estimation based on noisy data, subsampling, and model stationarity. We conduct a number of experiments on a measured channel route of about 8 minutes length. In this section we are only interested in the time dynamics of the channel. Therefore, we study a single subchannel. In the various experiments that will be conducted, we use the following baseline setup as a standard case with which we compare when evaluating alternative settings:

- SNR $\gamma=10$ dB
- upsampling factor (in time): 16, giving a pilot spacing of ~ 0.33 ms
- segmentation $N=1600$ samples ≈ 0.5 seconds
- model order $k=4$
- number of subchannels used for AR parameter estimation: $c = 1$
- number of subchannels used in the evaluation process: $w = 1$
- prediction range $L=4$ samples ≈ 1.33 ms
- velocity: maximum 30 km/h
- noise-free channel measurements are used in the training segment

We should note the following: using a point estimate of \mathbf{a} instead of the full pdf in (8.3.1) means that information about \mathbf{a} is discarded. Desideratum IIIb in Section 3.2.1 is therefore violated. We are forced to do this for computational reasons. Since optimality has then already been abandoned, seeking an optimal strategy for finding model order and AR parameters is futile and is by no means what we seek to do here. The experiments below are merely empirical studies of specific cases. The conclusions that we draw, although based on a large data set, may or may not have validity outside the specific situations investigated.

When conducting an experiment, we loop over all channel measurements in the entire measurement route (for one subchannel). The route is segmented into evaluation segments. Each evaluation segment is preceded by an equally long training segment which is used for AR parameter estimation. Before the

parameter estimation takes place, the evaluation segments are rescaled to an SNR of the experimenter's choice. Each training segment is also rescaled, with the same scaling factor as its corresponding evaluation segment, so that also all training segments over the entire route have approximately the same SNR.

Once the AR parameters have been estimated, they are used to construct a state-space model representation of the fading channel. This model is then used by a Kalman filter to calculate filtered channel estimates and predictions for the evaluation segment.

In the experiments, we set the noise variance to 0 dB, so that the SNR γ is also the channel power. We will calculate channel estimates and predictions based on acquired AR models. The quality of these estimations and predictions are evaluated with respect to normalized mean square error (NMSE), that is

$$\frac{N^{-1} \sum_{t=0}^{N-1} |h_t - \hat{h}_{t|t}|^2}{\gamma}, \quad (8.3.7)$$

when evaluating estimating performance, where $\hat{h}_{t|t}$ is a filtered estimate, or

$$\frac{(N - L)^{-1} \sum_{t=L}^{N-1} |h_t - \hat{h}_{t|t-L}|^2}{\gamma}, \quad (8.3.8)$$

when evaluating prediction performance, where $\hat{h}_{t|t-L}$ is a prediction. In the equations above, N is the length of the evaluation segment, and $t = 0$ here marks the beginning of an evaluation segment. The parameter γ is the chosen SNR. Alternatively, some SNR estimate $\hat{\gamma}$ may be used in place of γ in (8.3.7) and in (8.3.8).

We will also compare the experimental NMSE values with theoretic NMSE values that would be expected if the used model is an accurate description of the channel. Details of how to calculate theoretic NMSEs can be found in Section 4.3.

NMSE performance is plotted versus measured SNR, which we here define as the mean value of the squared envelope of the *training* segment. Since it is the evaluation segment that is scaled to have an exact SNR of γ , and the training segment is scaled by the same scaling factor, the measured SNR will differ from the SNR target due to the variability of the fading channel. It will be approximately, but not exactly, γ , as will be clear from the figures.

When plotting NMSE results, we also plot performance boundaries for channel estimation and prediction. For channel estimation, the boundary is $\text{NMSE} = (\gamma + 2)^{-1}$. It was shown in Section 5.3.4 that this boundary guaranties that the estimation error decreases the total perceived SNR by at most 3 dB. For channel prediction, we use a performance boundary of -10 dB.

In Section 6.5.1, it was argued that prediction NMSE below this level will have a very limited impact on link adaptation and scheduling performance.

8.3.1 Parameter estimation method

The first issue to consider is how to estimate the AR parameters \mathbf{a} in (8.3.1) from N channel measurements, where N is the length of a training segment. Here we evaluate parametric methods and set up an overdetermined system of equations

$$\mathbf{A}\mathbf{a} = \mathbf{b}. \quad (8.3.9)$$

A point estimate $\hat{\mathbf{a}}$ is then found by least squares by solving for $\hat{\mathbf{a}}$ in

$$(\mathbf{A}^* \mathbf{A})\hat{\mathbf{a}} = \mathbf{A}^* \mathbf{b}, \quad (8.3.10)$$

where $(\cdot)^*$ indicates conjugate transposition. We first assume that noise-free channel measurements are available. If, for example, $N = 100$ and the model order is $k = 4$, we may construct

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ h_0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 \\ \hline h_2 & h_1 & h_0 & 0 \\ \hline h_3 & h_2 & h_1 & h_0 \\ h_4 & h_3 & h_2 & h_1 \\ & & \vdots & \\ & & & \\ \hline h_{98} & h_{97} & h_{96} & h_{95} \\ \hline h_{99} & h_{98} & h_{97} & h_{96} \\ 0 & h_{99} & h_{98} & h_{97} \\ 0 & 0 & h_{99} & h_{98} \\ 0 & 0 & 0 & h_{99} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} h_0 \\ h_1 \\ h_2 \\ \hline h_3 \\ \hline h_4 \\ h_5 \\ \vdots \\ \hline h_{99} \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (8.3.11)$$

leading to the *autocorrelation* or the *Yule-Walker* method if the part above and below the dashed lines are included, and the *covariance* method if only the parts in between the dashed lines are incorporated.

If the autocorrelation method is used, then $\mathbf{A}^* \mathbf{A}$ becomes an estimate of the channel's autocorrelation matrix,

$$(\mathbf{A}^* \mathbf{A})[i, j] = \hat{r}(i - j), \quad (8.3.12)$$

where

$$\hat{r}(\tau) = \begin{cases} \sum_{l=0}^{N-1-\tau} h_l^* h_{l+\tau}, & \tau \geq 0 \\ \hat{r}^*(-\tau), & \tau < 0, \end{cases} \quad (8.3.13)$$

is a “biased” estimate of the autocorrelation for h_t . Since $\mathbf{A}^*\mathbf{A}$ is then Toeplitz, the system (8.3.9) can be solved with the Levinson recursions.

Note however, that the autocorrelation method applies windowing to the data, explicitly assuming that $h_t = 0$ when $t < 0$ and $t \geq N$. This assumption is not only unsupported, but plainly wrong. But again, we must point out that using a point estimate of the channel model parameters in the channel estimation/prediction algorithm is by necessity non-optimal. The viability of an AR parameter point estimation method can only be measured through experimentation.

The covariance method can be extended to the *modified covariance* method by adding another $N - k$ rows to the original $N - k$ rows:

$$\mathbf{A} = \begin{pmatrix} h_3 & h_2 & h_1 & h_0 \\ h_4 & h_3 & h_2 & h_1 \\ & & \vdots & \\ -\frac{h_{98}}{h_1^*} & -\frac{h_{97}}{h_2^*} & -\frac{h_{96}}{h_3^*} & -\frac{h_{95}}{h_4^*} \\ h_2^* & h_3^* & h_4^* & h_5^* \\ & & \vdots & \\ h_{96}^* & h_{97}^* & h_{98}^* & h_{99}^* \end{pmatrix}, \quad \mathbf{b} = - \begin{pmatrix} h_4 \\ h_5 \\ \vdots \\ -\frac{h_{99}}{h_0^*} \\ h_1^* \\ \vdots \\ h_{95}^* \end{pmatrix}. \quad (8.3.14)$$

While the covariance method minimizes the squared sum of the forward one-step prediction errors, the modified covariance minimizes the total squared sum of both the forward and the backward prediction errors [53].

We evaluate the autocorrelation method, the covariance method, and the modified covariance method on the baseline example, see Figure 8.1. Both estimation and prediction performance are plotted for the respective methods. Black dots indicate experimental results, and grey rings indicate theoretical values, calculated as described in Section 4.3. Each dot/ring corresponds to one evaluation segment. As can be seen from the figure, the autocorrelation method performs slightly better than the other methods, both in terms of estimation and in terms of prediction. Also, the agreement between theoretical and experimental values is best for the autocorrelation method. There is virtually no difference in performance between the covariance method and the modified covariance method.

We also study spectra produced by each method and compare them with the periodogram for the evaluation segment. See Figure 8.2, which illustrates spectra obtained with the respective methods, together with the periodogram, for an arbitrary segment of the measurement route. Notably, the two covariance methods will generally produce spectra that better conform to the periodogram of the signal than does the autocorrelation method (this is especially evident when studying an animation over several segments).

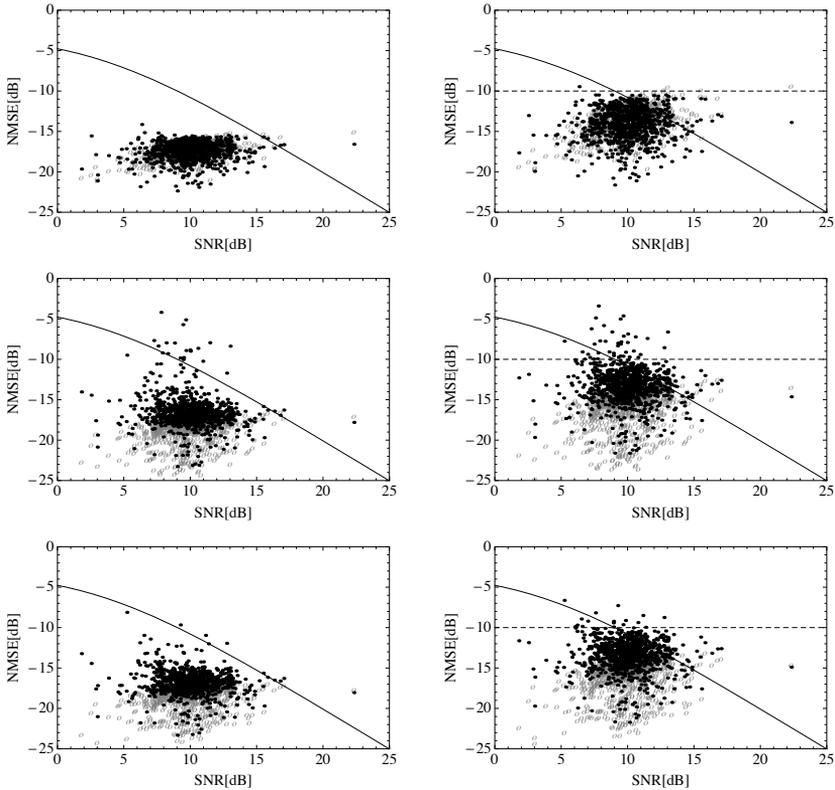


Figure 8.1: NMSE performance for channel estimation (left figures) and prediction (right figures) versus measured SNR, with simulation settings as given by the baseline setup. Upper figures illustrate results for the autocorrelation method. Middle figures illustrate results for the covariance method. Lower figures illustrate results for the modified covariance method. The solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Black dots indicate experimental results, whereas grey rings indicate theoretic results.

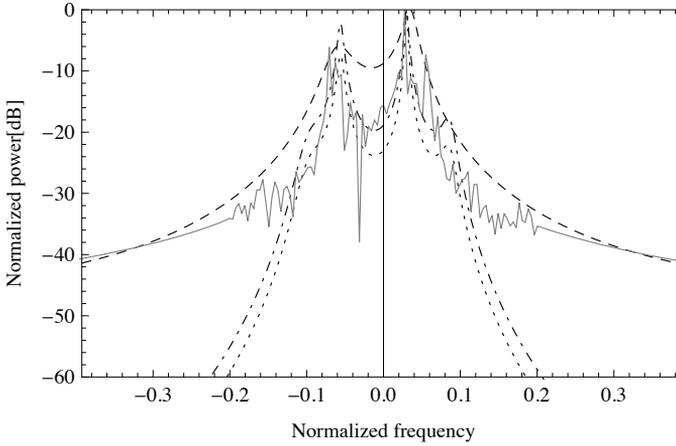


Figure 8.2: Spectra for an arbitrary evaluation segment for the baseline setup. The solid grey curve is the periodogram for the segment. Non-solid curves show spectra for the respective methods: the autocorrelation method (dashed curve), the covariance method (dash-dotted curve), and the modified covariance method (dotted curve). All curves have been normalized to have a maximum value of 0 dB.

In terms of pole placement, the covariance methods place poles closer to the unit circle, which generates more distinct spectral peaks in the spectrum. Although this may produce a truthful spectrum, it seems as though the resonant poles yield an overmodelling of the fading, which produces bad NMSE results and gives a poor agreement between theoretic and experimental NMSE values.

8.3.2 Model order

The model order k of the AR model (8.3.9) may be estimated from previous measurements. Common methods for model order selection is Minimum Description Length (MDL) [10] and the Akaike Information Criterion (AIC). In this work, we are primarily interested in finding a model order that is able to well represent the fading environment in all sorts of settings. The baseline example uses a model order of 4. Here we reiterate the baseline experiment but with model order $k = 6$. See Figure 8.3, which displays NMSE results given by the autocorrelation method. The performance improvement is minute compared with the baseline. The reason for this can be discerned in Figure 8.4, which shows spectra for all three methods for model order 6. While the spectra for the covariance methods do have extra peaks due to the new poles in the model, these do not contribute to new detail and increased resolution in the spectra. Furthermore, the autocorrelation method shows

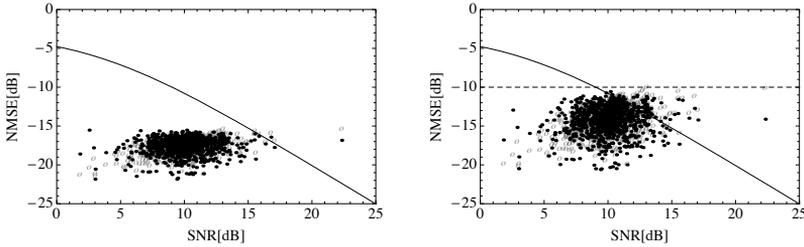


Figure 8.3: NMSE performance for channel estimation (left figure) and prediction (right figure) versus measured SNR, with simulation settings as given by the baseline setup, except that *the model order is here $k = 6$* . The autocorrelation method was used to acquire AR parameters. The solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Black dots indicate experimental results, whereas grey rings indicate theoretic results.

hardly any difference at all between the spectra for $k = 4$ and $k = 6$. We therefore conclude that, save for rare exceptions, a model order of 4 should be sufficient to capture significant details of the fading behaviour in the channel examined here.

8.3.3 Subsampling

The parameter estimation methods described in Section 8.3.1 find AR parameters that produce small *one-step* estimation errors. But in the prediction case, we are here interested in having a small *L-step* prediction error, where L is the prediction range. Using a model that works well for one-step prediction does not guarantee good L -step performance. This is especially evident when the channel is slowly fading. To make our point, we change the setting of the oversampling factor in the baseline example from 16 to 64. This has the effect of reducing the velocity from a top speed of 30 km/h, to a maximum value of about 7.5 km/h (pedestrian speed). To make a fair comparison, we also increase the prediction range from 4 to 16. Figure 8.5 shows NMSE results for this case. Since the prediction range is the same as in the baseline example in terms of wavelengths, while at the same time more measurements per wavelength are acquired here, one may expect the NMSE performance to be better in this experiment. This is indeed the case for channel estimation, as can be seen by comparing Figure 8.5 to the upper figures in Figure 8.1. Channel prediction, on the other hand, performs *worse* than the baseline experiment. The reason for this is that the parameter estima-

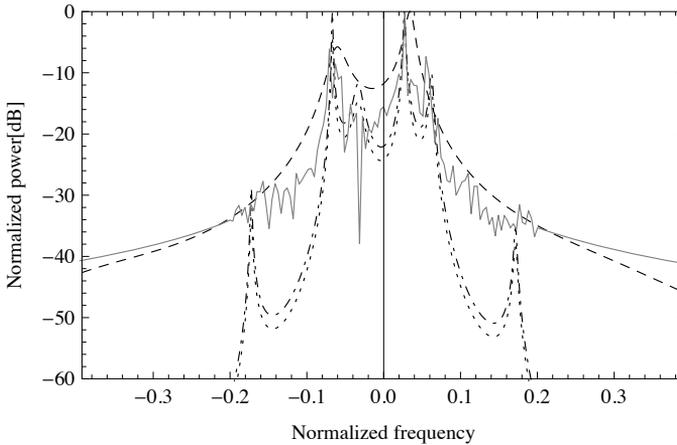


Figure 8.4: Spectra for an arbitrary evaluation segment for the baseline setup, except that *the model order is here $k = 6$* . The solid grey curve is the periodogram for the segment. Non-solid curves show spectra for the respective methods: the autocorrelation method (dashed curve), the covariance method (dash-dotted curve), and the modified covariance method (dotted curve). All curves have been normalized to have a maximum value of 0 dB.

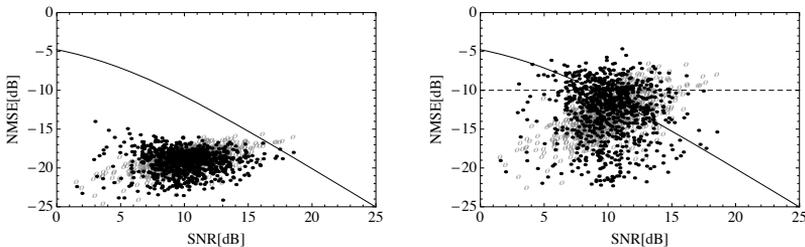


Figure 8.5: NMSE performance for channel estimation (left figure) and prediction (right figure) versus measured SNR, with simulation settings as given by the baseline setup, except that *the oversampling factor is here 64 and the prediction range is 16*. The autocorrelation method was used to acquire AR parameters. The solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Black dots indicate experimental results, whereas grey rings indicate theoretic results.

tion methods based on the one-step predictions cannot resolve details in the channel's spectrum, which is now very narrow due to the high oversampling factor.

To remedy this, an AR model based on the sub-sampled channel may be used:

$$h_t + \alpha_1 h_{t-s} + \alpha_2 h_{t-2s} \dots + \alpha_k h_{t-ks} = w_t, \quad (8.3.15)$$

where s is the *subsampling factor* that dictates the separation between channel measurements in the model and w_t is a white noise. The AR parameters $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_k\}$ are found as before, by setting up an overdetermined system of equations (8.3.9), with $\boldsymbol{\alpha}$ in place of \mathbf{a} , and then solve with least squares with respect to $\boldsymbol{\alpha}$. See Appendix 8.A for an example.

Since the channel estimator/predictor filter will be operating on the original, non-sampled, measurements and not on the sub-sampled data, the acquired AR model must be converted to the faster sampling domain. The poles $\{\pi_1, \dots, \pi_k\}$ of the sub-sampled model (which can be calculated from the AR parameters $\boldsymbol{\alpha}$) are here converted to poles $\{p_1, \dots, p_k\}$ for the full-sampled model through the transformation

$$p_i = \pi_i^{-s}, \quad 1 \leq i \leq k, \quad (8.3.16)$$

where s is the subsampling factor. As described in Channel 4, channel estimates and predictions can be produced by a Kalman filter based on a multiple-input multiple-subchannel state space model. The state space model is constructed so that the state transition matrix \mathbf{F} is diagonal with the poles $\{p_i\}$ of the channel model(s) along its diagonal. Channel predictions $\hat{\mathbf{h}}_{t+L|t}$ are calculated through $\hat{\mathbf{h}}_{t+L|t} = \mathbf{H}\mathbf{F}^L\hat{\mathbf{x}}_{t|t}$, where $\hat{\mathbf{x}}_{t|t}$ is the filtered state estimate and \mathbf{H} is a matrix in the state space model. By choosing the poles as in (8.3.16), \mathbf{F}^L becomes diagonal with the subsampled poles $\{\pi_i\}$ along its diagonal. We expect that the L -step prediction NMSE should then be low, since the subsampled poles were adjusted to produce a small prediction error for precisely this prediction horizon.

We now set the sub-sampling factor to $s = 16$ so that it matches the prediction range L , and reiterate the experiment, see Figure 8.6. Evidently, the prediction NMSE performance is raised considerably as compared to the experiment where sub-sampling was not used. We conclude that sub-sampling is important in slow fading scenarios. When the fading is faster, i.e. in vehicular velocity scenarios with the investigated carrier frequency, we have not seen any pronounced advantage with using sub-sampling. The importance of sub-sampling has also been pointed out in [4] and in [6].

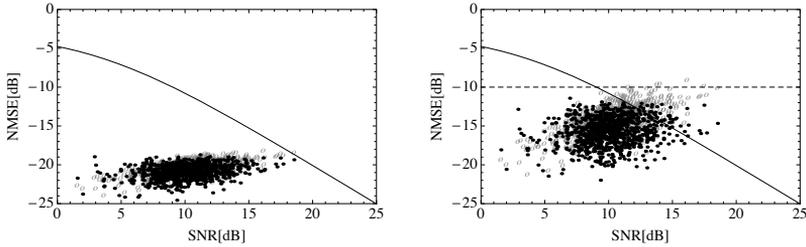


Figure 8.6: NMSE performance for channel estimation (left figure) and prediction (right figure) versus measured SNR, with simulation settings as given by the baseline setup, except that *the oversampling factor is here 64 and the prediction range is 16*. The autocorrelation method was used to acquire AR parameters. *In the parameter acquisition, the regressors were subsampled by a factor 16*. The solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Black dots indicate experimental results, whereas grey rings indicate theoretic results.

8.3.4 Parameter estimation in noise

So far, we have made the idealizing assumption that noise-free channel measurements are available in the training segments during which AR parameter estimation takes place. We here consider AR parameter estimation in the baseline example when complex-valued white noise of unit variance is added to the channel measurements. Denote by $\tilde{\mathbf{A}}$ the matrix \mathbf{A} in (8.3.11), (8.3.14), or (8.A.1), but with the noise-free channel coefficients replaced by noisy measurements. Then we may calculate

$$\mathbf{A}^* \mathbf{A} \approx \tilde{\mathbf{A}}^* \tilde{\mathbf{A}} - n \mathbf{I}, \quad (8.3.17)$$

where n is the number of non-zero elements in each column in \mathbf{A} , which depends on the size N of the training segment, the model order k , and the oversampling factor s . For the autocorrelation method, $n = N$; for the covariance method, $n = N - ks$; and for the modified covariance method, $n = 2(N - ks)$. Equation (8.3.10) can now be approximated as

$$(\tilde{\mathbf{A}}^* \tilde{\mathbf{A}} - n \mathbf{I}) \hat{\mathbf{a}} = \tilde{\mathbf{A}}^* \mathbf{b}, \quad (8.3.18)$$

from which we find estimates $\hat{\mathbf{a}}$ of the AR parameters¹. Note that it holds that $\mathbf{A}^* \mathbf{b} \approx \tilde{\mathbf{A}}^* \mathbf{b}$ because of the whiteness of the noise. Once $\hat{\mathbf{a}}$ has been

¹Approximately the same solution is obtained by low-pass filtering the noisy measurements with a filter with cut-off frequency at the maximum Doppler frequency and using (8.3.10)

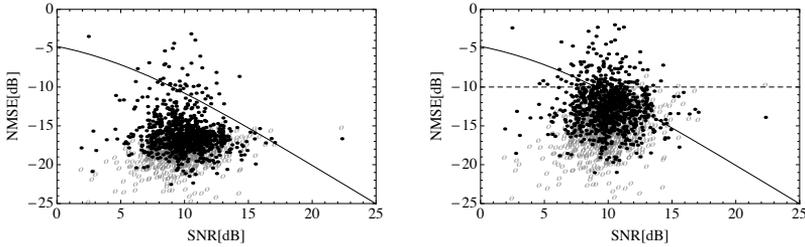


Figure 8.7: NMSE performance for channel estimation (left figure) and prediction (right figure) versus measured SNR, with simulation settings as given by the baseline setup, except that *noisy regressors are used during the AR parameter acquisition*. The autocorrelation method was used to acquire AR parameters. The solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Black dots indicate experimental results, whereas grey rings indicate theoretic results.

found, the corresponding poles can be calculated. Unstable poles are reflected in the unit circle.

Figure 8.7 displays NMSE performances for the autocorrelation method when noisy regressors are used for the parameter estimation. As expected, performance is considerably impeded by the noise, compared with the upper figures in Figure 8.1. When the regressors are noisy, the spectra for the respective methods typically feature random bumps, as can be seen in Figure 8.8.

8.3.5 Estimation on multiple subchannels

The quality of AR parameter estimation can be improved upon by taking into account several parallel subchannels in the estimation process. Denote by \mathbf{A}_i and \mathbf{b}_i the parameters in an equation system (8.3.9) corresponding to a subchannel with index i . In order to take c different subchannels into account, we construct an augmented system

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_c \end{pmatrix}, \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_c \end{pmatrix}, \quad (8.3.19)$$

which is then used in (8.3.9) to find AR parameters. It is desirable to use well-separated subchannels, so that they fade independently.

We study the performance for multi-subchannel AR parameter estimation on the baseline case with noisy regressors as in Section 8.3.4. Since the size of

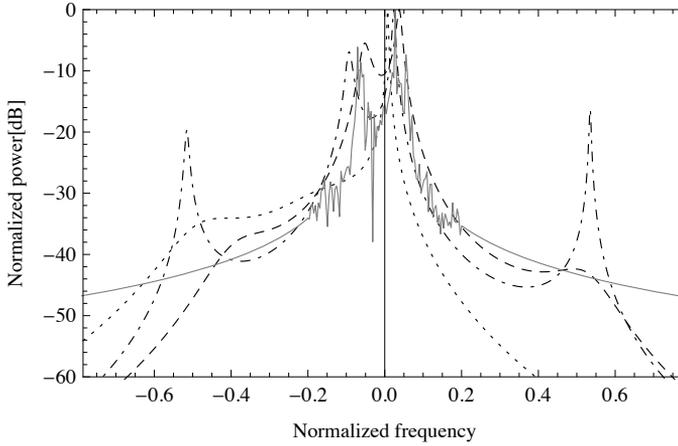


Figure 8.8: Spectra for an arbitrary evaluation segment for the baseline setup, except that *noisy channel measurements were used in the AR parameter acquisition*. The solid grey curve is the periodogram for the segment. Non-solid curves show spectra for the respective methods: the autocorrelation method (dashed curve), the covariance method (dash-dotted curve), and the modified covariance method (dotted curve). All curves have been normalized to have a maximum value of 0 dB.

the equation system has here increased by a factor c , the noise compensation factor n in (8.3.17) must be multiplied by c . As can be seen by comparing Figure 8.9 with 8.7, multi-subchannel estimation has a favourable impact on AR parameter estimation performance when noisy regressors are used.

8.3.6 Segmentation

The segmentation of the channel determines how often the channel model is updated. The segments must be short enough that the channel can be considered more or less stationary over the duration of a segment, otherwise the channel model will be outdated at the end of the segment. In the experiments above we have used a segmentation of 1600 samples, corresponding to training and evaluation segments with durations of about half a second each. From the experiments above, it would seem that this segmentation is short enough to capture the changing nature of the fading environment. But can the segments be made longer, lowering the updating frequency for the channel model? Figure 8.10 illustrates NMSE performances when the segments are 6400 samples long (about 2 seconds). The experimental results still comply with the criteria that we have set up for channel estimation and prediction performance, but some points start to fall above the performance boundaries. This indicates that while a segmentation of 0.5 seconds seem to

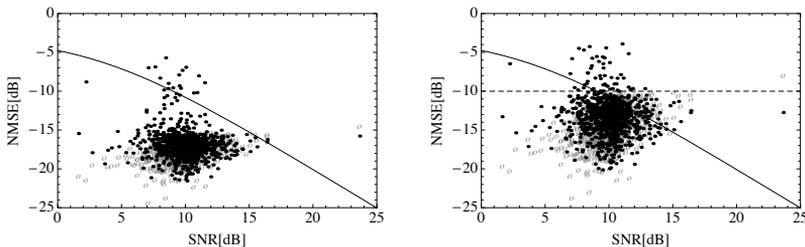


Figure 8.9: NMSE performance for channel estimation (left figure) and prediction (right figure) versus measured SNR, with simulation settings as given by the baseline setup, except that $c = 8$ well-separated subchannels are used for AR parameter acquisition. Noisy channel measurements are used. The autocorrelation method was used to acquire AR parameters. The solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Black dots indicate experimental results, whereas grey rings indicate theoretic results.

be well on the safe side, segments of 2 seconds are sometimes too long for the channel to be considered stationary over a segment's duration, when travelling at pedestrian or low vehicular velocities in the suburban environment examined in the evaluation.

8.4 MIMO channels

In Section 8.3 we studied various properties of the time dynamics of a measured channel. A single subchannel was studied. In this section we estimate and predict multiple parallel subchannels as well as simultaneous inputs from multiple transmitting antennas.

As previously in this thesis, we use a linear discrete-time state space model to model a vector \mathbf{h}_t of fading MIMO-OFDM channel coefficients:

$$\begin{aligned}\mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t, \\ \mathbf{h}_t &= \mathbf{H}\mathbf{x}_t, \\ \mathbf{y}_t &= \Phi_t\mathbf{h}_t + \mathbf{v}_t,\end{aligned}\tag{8.4.1}$$

where the process noise \mathbf{u}_t , the measurement noise \mathbf{v}_t , and the initial state \mathbf{x}_0 are zero-mean Gaussian, white, and

$$\|[\mathbf{u}_t^T, \mathbf{v}_t^T, \mathbf{x}_0^T]^T\|^2 = \text{diag}(\mathbf{Q}, \mathbf{R}, \Pi_0),\tag{8.4.2}$$

where $\mathbf{Q} > 0$, $\mathbf{R} > 0$, and $\Pi > 0$ as described in detail in Chapter 4. The uw -vector \mathbf{h}_t holds w fading time-frequency channel coefficients from

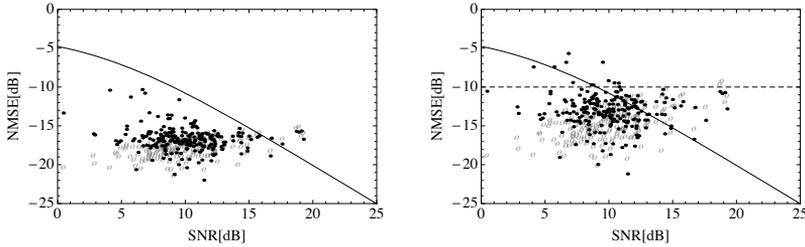


Figure 8.10: NMSE performance for channel estimation (left figure) and prediction (right figure) versus measured SNR, with simulation settings as given by the baseline setup, except that *training segments and evaluation segments are here about 2 seconds long*. The autocorrelation method was used to acquire AR parameters. The solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Black dots indicate experimental results, whereas grey rings indicate theoretic results.

u transmitting antennas. These channel coefficients are “observed” in noise through the $w \times uw$ pilot matrix Φ_t . It is assumed that the model only concerns those sub-locations that hold training symbols (pilots) so that Φ_t is known. Depending on the pilot patterns used, Φ_t potentially constructs a superposition of channel coefficients from several transmitting antennas in the measurement vector \mathbf{y}_t of dimension w .

From the observations \mathbf{y}_t , optimal inferences about the present channel \mathbf{h}_t (estimation, for coherent detection) and future channel \mathbf{h}_{t+L} (L -step prediction, e.g. for link adaptation and scheduling) can be made by using the Kalman filter.

8.4.1 Frequency correlation

In order to model multiple parallel subchannels, the frequency correlation between these subchannels has to be incorporated into the model. Exactly how this is done depends on whether impulse response domain modelling or subchannel domain modelling is used, see Chapter 4. In the former case, the frequency domain properties of the channel model is represented by a pulse-shaping matrix \mathbf{D} and by a diagonal power delay profile matrix \mathbf{R}_γ . In this case it is necessary to estimate the number of significant taps in the impulse response, as well as their respective delays and gains. This is typically done by using subspace methods. See for example [9]. However, we will not consider impulse response domain modelling here.

subchannel domain modelling, on the other hand, requires only an esti-

mate of the subchannel covariance matrix $\mathbf{R}_h = \|\mathbf{h}_t\|^2$. To find this matrix, we start by estimating the channel autocorrelation in the frequency direction. Based on n noisy pilot sub-symbol measurements from a particular OFDM symbol,

$$y_f = s_f h_f + v_f, \quad f = 0, \dots, n-1, \quad (8.4.3)$$

which need to be exclusively allotted to the channel that are to be modelled, we construct the sample autocovariance in the frequency direction²

$$r_{f,y}(\tau) = \begin{cases} \sum_{f=0}^{n-1-\tau} (s_{f+\tau}^{-1} y_{f+\tau}) (s_f^{-*} y_f^*), & \tau \geq 0, \\ r_{f,y}^*(-\tau), & \tau < 0. \end{cases} \quad (8.4.4)$$

Above, s_f , h_f , and v_f are a known pilot sub-symbol, the unknown channel coefficient, and complex-valued white noise of unit variance, respectively, for a pilot-bearing subchannel with index f . We then compensate for the influence of white noise:

$$r_{f,h}(\tau) = r_{f,y}(\tau) - \sigma_w^2 \delta(\tau), \quad (8.4.5)$$

where $\sigma_w^2 = n\sigma_v^2/|s_t|^2$ is the noise variance. We assume here that the variance σ_v^2 of v_t is known to be $\sigma_v^2 = 1$.

However, if the channel has a large coherence bandwidth, then the overall scale of $r_{f,h}(\tau)$ will depend greatly on the fading level that the present OFDM symbol happens to have. Therefore, $r_{f,h}(\tau)$ will generally have the wrong scaling. Since for an accurate frequency-domain autocorrelation $r_f(\tau)$ it should hold that $r_f(0) = \gamma$, where γ is the channel's SNR, it is appropriate to apply the following rescaling:

$$r_f(\tau) = (\hat{\gamma}/r_{f,h}(0)) \times r_{f,h}(\tau), \quad (8.4.6)$$

where $\hat{\gamma}$ is a good estimate of the SNR. We turn to the topic of how to obtain $\hat{\gamma}$ in Section 8.4.2 below.

The toeplitz subchannel covariance matrix can now be constructed from the rescaled sample autocovariances (8.4.6):

$$\mathbf{R}_h = \begin{pmatrix} r_f(0) & r_f^*(1) & \cdots & r_f^*(w-1) \\ r_f(1) & r_f(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_f^*(1) \\ r_f(w-1) & \cdots & r_f(1) & r_f(0) \end{pmatrix}. \quad (8.4.7)$$

As described in Chapter 4, \mathbf{R}_h can be used in the construction of the channel model.

²Autocovariance estimation comes in many flavours. It is not uncommon to use so called unbiased estimation, $r_{unb.}^{(f)}(\tau) = (n+\tau)^{-1}r^{(f)}(\tau)$. However, the τ -dependent scaling may result in an estimate that does not have the properties of an autocorrelation, e.g. that $r(0)$ should have the largest absolute value. This may eventually result in a negative definite autocovariance matrix which gives an unstable filter.

8.4.2 SNR

To measure the SNR, we need to measure both the noise power and the channel power. The noise power is conveniently measured on “empty” resources, i.e. sub-locations where neither pilots nor payload is transmitted. Let Γ_v be a set of such empty sub-locations. An estimate $\hat{\sigma}_v^2$ of the noise power σ_v^2 can then be calculated from

$$\hat{\sigma}_v^2 = |\Gamma_v|^{-1} \sum_{i \in \Gamma_v} |v_i|^2, \quad (8.4.8)$$

where v_i is the noise measurement at sub-location i . The thermal noise power can be assumed to be fairly constant over long periods of time, but due to the time-varying nature of interference, the noise power estimate has to be updated on a time scale of seconds.

Channel power estimation amounts to taking the mean of the squares of a large number of samples. Similar to the noise power estimation procedure, let Γ_h be a large set of pilot sub-locations that are well separated both in time and in frequency. Construct

$$\hat{p} = |\Gamma_h|^{-1} \sum_{i \in \Gamma_h} |y_i|^2, \quad (8.4.9)$$

where $y_i = s_i h_i + v_i$ is the noisy measurement of sub-location i bearing pilot sub-symbol s_i .

Let the number of samples be $|\Gamma_h| = n$. If the measurements are taken so sparsely that the $\{y_i\}$ are i.i.d. Gaussian with zero mean and identical variance σ_y^2 , then \hat{p} has a χ^2 -distribution with n degrees of freedom. This distribution has mean value σ_y^2 and variance σ_y^4/n [97]. When the number of samples n is large (say, $n > 50$), then \hat{p} is approximately Gaussian with the aforementioned moments. Since it must hold that $\sigma_y^2 = |s|^2 \sigma_h^2 + \sigma_v^2$, where σ_v^2 is the known noise variance and we assume that the pilots have constant modulus $|s|$, we can now solve for σ_h^2 .

However, if the channel samples are correlated, the power estimate may exhibit bias and the convergence will generally be slow. Since the channel taps in an OFDM channel are usually strongly correlated, a large portion of the system bandwidth needs to be sampled over a fair amount of time in order to get enough “diversity” to make the variance of the power estimate small. Loosely speaking, we can only get $(\Delta f_c \Delta t_c)/(f_{bw} t_{meas.})$ degrees of freedom, where Δf_c is the coherence bandwidth of the channel (the reciprocal of the multipath spread), Δt_c is the coherence time of the channel (the reciprocal of the Doppler spread), f_{bw} is the total system bandwidth, and $t_{meas.}$ is the time over which the power is estimated. The time $t_{meas.}$ is typically in the order of a second.

by at most 3 dB. For channel prediction, we use the performance boundary -10 dB, motivated in Section 6.5.1. Here, all curves satisfy the performance boundaries when the SNR is higher than 5 dB.

In the second case, separated model parameter acquisition is still used, but we consider tracking of parallel inputs with overlapping pilots as given by the pilot matrix (8.4.11). In both locations, tracking simultaneous inputs causes interference between the inputs due to the frequency selectivity of the channels, which increases the NMSE. See Figure 8.11b. Prediction performance is still adequate, but the channel estimation performance curves fall off, so that for high SNRs, channel estimation performance is inadequate.

In the third case we only estimate channel model parameters for one of the three inputs, and then we use the same model for all three inputs. Since we here consider inputs coming from the same antenna array, it is expected that all inputs are adequately characterized by the same model. Only having to acquire one channel model saves resources, since model parameter acquisition requires the use of exclusive pilots. As seen in Figure 8.11c, the performance curves are very nearly the same as those in Figure 8.11b, implying that when parallel inputs stemming from the same antenna array are tracked, it is sufficient to acquire only one channel model.

In coordinated multipoint (CoMP) settings, separate channel models need to be acquired for the individual input streams, since they originate from different base stations. This case has been studied in [86].

8.5 Filter convergence

As seen in Section 4.4, the Kalman filter recursions are computationally cumbersome. Although special matrix structures can be used here, alleviating the arithmetical pressure considerably compared to the case where a general state space model is used, the complexity is still in the order of n^2 , where $n = kxu$ is the total number of states and u , x , and k are the number of transmitting antennas, the number of modelled channel coefficients per transmitting antenna, and the model order used for each channel coefficient, respectively (typically, we set $x = w$, where w is the filter width). In terms of operations per second, modern processors can manage this computational labour, but power consumption is a problem. The sheer intensity would quickly drain the batteries in mobile terminals and might make base station's power consumption unacceptably high. It is therefore vital to reduce the number of Kalman filter updates to a minimum.

The Kalman filter permits the matrices in the state space model to be time variant. However, since scatterers surrounding mobile terminals are relatively static compared to typical channel sampling periods (fractions of mil-

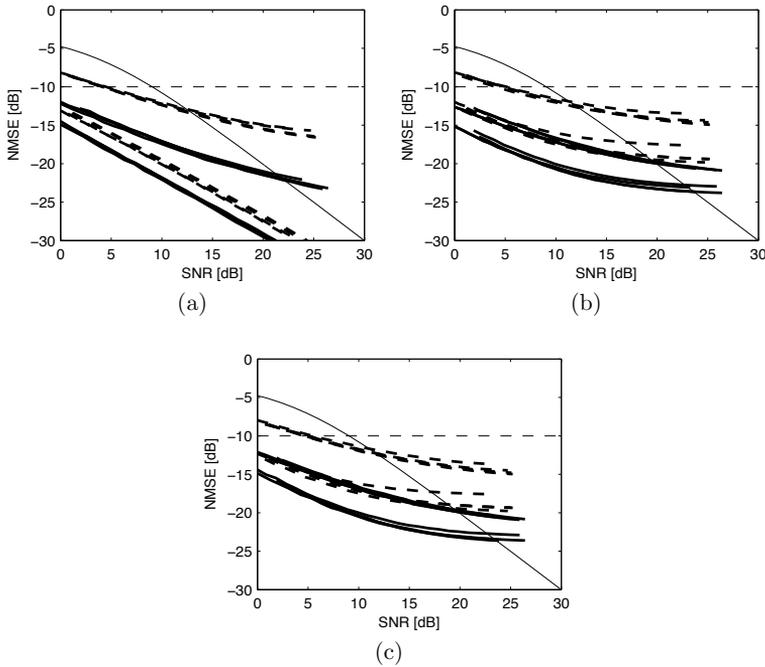


Figure 8.11: Channel estimation and prediction NMSE results for two distinct locations: one advantageous location (solid curves), and one somewhat less advantageous location (dashed curves). Channel estimation NMSE performance (lower lines) and prediction NMSE performance (upper lines) are plotted versus SNR. Three simultaneous channels were tracked. Their respective curves nearly overlap. The thin solid curve is the performance criterion limit (5.3.10) for channel estimation, given in Section 5.3.4. The thin dashed line is the performance criterion limit -10 dB for channel prediction, motivated in Section 6.5.1. Three cases for channel model parameter acquisition and channel estimation/prediction are investigated: completely separated model and channel estimation/prediction (a), separated model estimation but simultaneous estimation/prediction (b), and simultaneous estimation/prediction with the same model used for all channels (c).

liseconds), the channel model parameters will generally be constant for many channel sample periods. This is indicated in (8.4.1) by the time-invariability of all matrices but Φ_t (the pilot matrix Φ_t is allowed to vary on a short time scale).

When the model matrices are constant or periodic, the error covariance matrices computed by the Kalman filter recursions will quickly settle to a stationary or periodic state. When this happens, the error covariance matrices will not have to be updated again until the model changes.

A central question arising when assessing the feasibility of using Kalman filters for channel estimation and prediction is then: how quickly does the Kalman filter converge compared to the length over which a static model can be used to represent a fading channel?

To examine the convergence rate of the Kalman filter, we study two cases. The first is a multi-input system of four transmitting antennas (possibly four different users), each having an SNR of 12 dB. The filter width w is set to 4 and the four pilot patterns are orthogonal and periodic with a period of 4. The system parameters are set according to the WINNER FDD mode (see Section 6.2 for details) and the WINNER II C2 NLOS channel model is used (see Section 2.3.2). A Doppler spectrum resembling the Jakes' Doppler spectrum is used. The normalized Doppler frequency is set to 0.06, which for example corresponds to having pilot OFDM symbols at every 12:th OFDM symbol (345.6 μ s) when the carrier frequency is 3.7 GHz and the velocity is about 50 km/h. The estimation performance is expressed in terms of NMSE for the filtered channel estimates, but the actual numerical values are not interesting here. Instead we study the convergence rate versus number of iterations (recursion updates). As is clear from Figure 8.12, the MIMO filter has converged after about 60 iterations.

We also consider a SISO system where the pilot pattern is varied over time with a period of 12. The system parameters are set as in the above example, but we here consider the input from a single transmitting antenna. For most of the time, empty pilots (i.e. the pilot symbols have no energy) are transmitted. Only at two occasions are non-zero pilots transmitted: at subchannel 1 at time 10, and at subchannel 4 at time 12. The original intention of this experiment was to investigate how pilots should be distributed among payload symbols in a certain kind of block allocation scheme (see Chapter 5), but again, here we are only interested in convergence rates. Because of the asymmetric pilot placement, as is seen in Figure 8.13, the NMSE curves differ slightly between the four subchannels and they have a jagged appearance. At time $t = 10$, when the first pilot arrives, the NMSE drops down to about -12 dB. The NMSE is further improved slightly when the next pilot arrives at $t = 12$. After that, the NMSE increases until the first pilot in period two appears. After about three periods, the SISO filter with time varying pilots

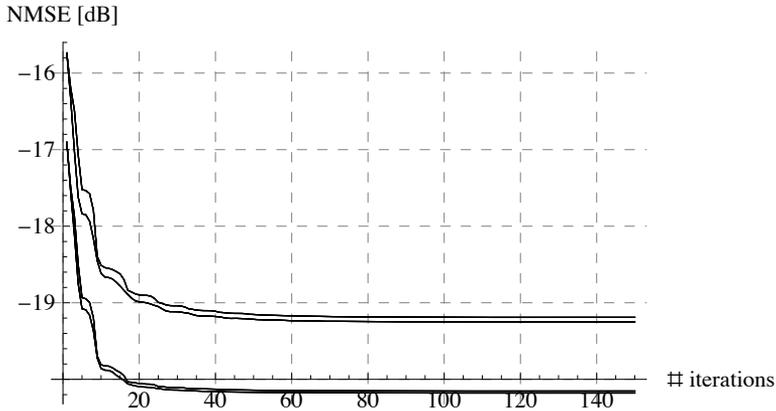


Figure 8.12: Channel estimation NMSE performance of a Kalman filter with width $w = 4$ operating on four simultaneous input streams using orthogonal and cyclic pilot patterns. Upper curves show performance results for the outer subchannels for the respective inputs. Lower curves show results for inner subchannels. Because of the way the pilots were chosen, the amount of energy that leaks from other inputs will differ slightly between inputs, which is why multiple curves are present.

settles to a periodic state.

In the above examples of multi-channel tracking and irregular pilot patterns, the Kalman filter converged in only a few tens of iterations. The convergence was this fast despite the fact that the Doppler spectrum featured sharp spectral components that causes the filter to converge more slowly than it would with a flat Doppler spectrum. Adding some margin, we therefore draw the conclusion that one can safely assume the Kalman filter to converge, also in fairly complicated cases, in at most a couple of hundred iterations.

8.6 CoMP complexity case study

In Section 8.3.6 we saw that it is in many cases sufficient to update the channel model twice per second. If the filter converges in a hundred iterations, this means that the computationally demanding state estimation error covariance matrices in the Kalman filter need to be updated 200 times per second and filter band. If the channel models and pilot configuration is the same between different filter bands, the covariance matrices need only be updated for one filter band.

To illustrate the resulting impact on the numerical complexity of Kalman predictors operating on multiple inputs in a CoMP setting, we consider a system with a total bandwidth of 20 MHz, divided into 1296 subchannels at

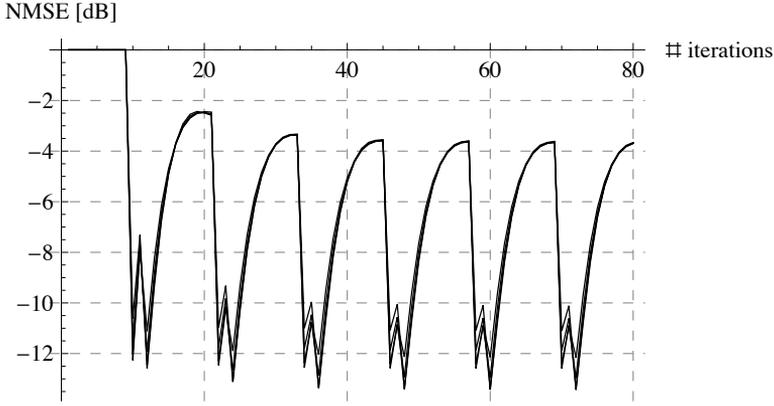


Figure 8.13: Channel estimation NMSE performance of a Kalman filter with width $w = 4$ operating on a single input. An asymmetric pilot pattern with period 12 was used. Upper and lower curves show the result for outer and inner subchannels, respectively.

15 kHz bandwidth each. Every third subchannel holds pilots, so that a total of 432 subchannels are to be tracked. Assuming a filter width of $w = 4$, this means that 108 Kalman filters need to operate in parallel. We assume that u simultaneous inputs should be tracked, but that exclusive pilots are used so that separate Kalman filters can be run on each input, as discussed in Section 6.6. Guided by the results obtained in Section 8.3.6, we here assume that new channel models need to be acquired once per second. The same channel model can be used over the entire bandwidth but different models need to be employed for the different inputs. The spacing between pilots, t_p , is here set to 0.25 ms, so that each filter has to be updated 4000 times per second. It is further assumed that a Kalman filter on average converges in 60 iterations. We set the model order to $k = 4$ so that the number of states is $n = wk$, and the prediction range to $L = 8$ steps. The resulting number of complex arithmetic operations required per second versus the number of tracked inputs u is presented in Table 8.1. For the complexity calculations we have used equation (4.4.1). We have also used the fact that a multiplication $\mathbf{M}_1\mathbf{M}_2$ of two matrices $\mathbf{M}_1 \in \mathbb{C}^{n_1 \times n_2}$ and $\mathbf{M}_2 \in \mathbb{C}^{n_2 \times n_3}$ requires $n_1n_2n_3$ complex operations. If the rows of \mathbf{M}_1 and/or the columns of \mathbf{M}_2 contain only c non-zero elements each, then the complexity is n_1cn_3 . Further, if the product is known to be Hermitian, then only half of the arithmetic operations need to be carried out.

The required number of complex arithmetic operations per second as a function of the number of simultaneously tracked inputs u is displayed in Figure 8.14. The figure shows that in the investigated case, the complexity

Table 8.1: Number of arithmetic operation per second required to cover the entire 20 MHz bandwidth, tracking u inputs at once. The calculations are divided into four categories: *Filter update* refers to the updating of the whole filter. *Covariance rescaling* refers to the fact that the state estimation error covariance matrix $\mathbf{P}_{t|t}$ and the state prediction error covariance matrix $\mathbf{P}_{t+L|t}$ have to be updated to $\mathbf{H}\mathbf{P}_{t|t}\mathbf{H}^*$ and $\mathbf{H}\mathbf{P}_{t+L|t}\mathbf{H}^*$, respectively. *Estimate update* refers to the updating of the state estimates $\hat{\mathbf{x}}_{t|t}$ and the state predictions $\hat{\mathbf{x}}_{t+L|t}$. *Estimate rescaling* refers to the fact that the state estimation and state prediction have to be transformed to channel coefficients $\hat{\mathbf{h}}_{t|t} = \mathbf{H}\hat{\mathbf{x}}_{t|t}$ and $\hat{\mathbf{h}}_{t+L|t} = \mathbf{H}\hat{\mathbf{x}}_{t+L|t}$.

type	number of complex arithmetic operations per second
filter update	$60u \left(n^2 \left(\frac{w}{2} + \frac{3}{2} + \frac{L-1}{2} \right) + n \left(w^2 + \frac{3w}{2} + 3 + (L-1) \right) + \frac{w^3}{6} \right)$
covariance rescaling	$60u(2(nwk + w^2k/2))$
estimate update	$4000 \cdot 108u(u(w+2) + u)$
estimate rescaling	$4000 \cdot 108u(2wk)$

is totally dominated by the updating of estimates and predictions.

8.7 Concluding remarks

This chapter has been devoted to the study of various properties of measured broadband MIMO channels. Channel estimation and prediction performance using a Kalman filter has been evaluated in terms of the normalized mean square error (NMSE). The time dynamics of the measured channels were studied and it was found that they could be adequately represented by an AR model of order 4. For the specific measurement studied here, a higher model order did not contribute to improved estimation and prediction performance. Among the evaluated methods for finding the AR model parameters, it was found that the autocorrelation was the best alternative. A method for compensating for the negative influence of noise when estimating model parameters was suggested. Also, it was shown that using several uncorrelated subchannels in the model estimation process can remedy the problem with noisy regressors somewhat. When the channel fades slowly, we showed that estimation and prediction is improved by using a sub-sampled AR model. Piecewise constant channel models were used. We found that the channel models could be held constant for at least 0.5 seconds.

For the study of multiple subchannels tracked in parallel, we outlined methods for estimating the SNR and frequency correlation properties. We evaluated channel estimation and prediction for four parallel subchannels and three simultaneous inputs coming from antenna elements in the transmitting

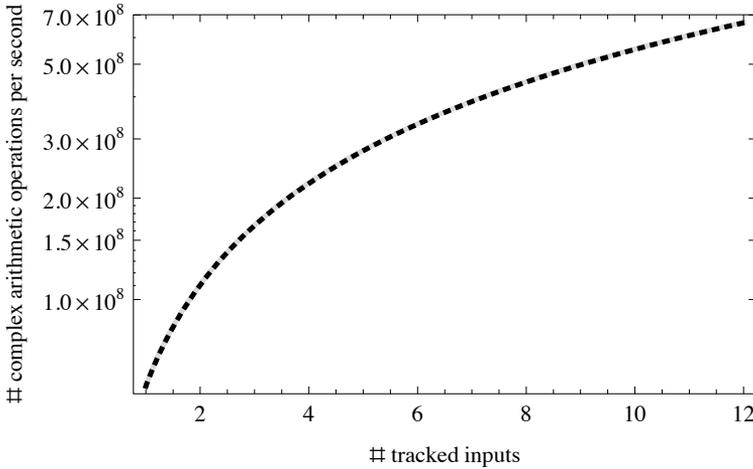


Figure 8.14: The number of complex arithmetic operations that has to be carried out per second, as a function of the number of simultaneously tracked inputs u . Grey curve shows the complexity for the sum of all four types of update in Table 8.1. Black dotted curve shows the complexity for just the estimate update and rescaling. The curves almost completely overlap, meaning that the updating of covariance matrices has a negligible impact on the total numerical complexity.

antenna array. The main conclusion was that when all inputs originate from the same antenna array, then the same channel model can be used for all inputs.

Finally, we examined the convergence rate of the Kalman filter used for channel estimation and prediction. We found that the filter usually converges in a few tens of iterations, even in cases where the Doppler spectra have strong spectral components and when simultaneous inputs are tracked.

8.8 Acknowledgements

The channel measurements used in this Chapter were gratefully provided by Ericsson Research via the Vinnova project Multi-antenna Transmission and Scheduling in IMT-Advanced.

The author would also like to express his gratitude towards Rikke Abildgaard Olesen, who provided the simulation results for Section 8.4.3.

8.A Subsampling

From the sub-sampled AR model (8.3.15), an over-determined set of equations can be set up. For example, with the sub-sampling factor $s = 3$, the model order $k = 4$, and the size of the measurement series $N = 100$, we have

$$\mathbf{A} = \begin{pmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 & h_0 \\ h_4 & h_1 \\ h_5 & h_2 \\ h_6 & h_3 & h_0 \\ h_7 & h_4 & h_1 \\ h_8 & h_5 & h_2 \\ \hline h_9 & h_6 & h_3 & h_0 \\ h_{10} & h_7 & h_4 & h_1 \\ h_{11} & h_8 & h_5 & h_2 \\ h_{12} & h_9 & h_6 & h_3 \\ \vdots \\ \hline h_{96} & h_{93} & h_{90} & h_{87} \\ h_{97} & h_{94} & h_{91} & h_{88} \\ h_{98} & h_{95} & h_{92} & h_{89} \\ h_{99} & h_{96} & h_{93} & h_{90} \\ & h_{97} & h_{94} & h_{91} \\ & h_{98} & h_{95} & h_{92} \\ & h_{99} & h_{96} & h_{93} \\ & & h_{97} & h_{94} \\ & & h_{98} & h_{95} \\ & & h_{99} & h_{96} \\ & & & h_{97} \\ & & & h_{98} \\ & & & h_{99} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \\ h_{10} \\ h_{11} \\ \hline h_{12} \\ h_{13} \\ h_{14} \\ h_{15} \\ \vdots \\ \hline h_{99} \end{pmatrix}, \quad (8.A.1)$$

giving either the autocorrelation method, the covariance method, or the modified covariance, depending on how \mathbf{A} and \mathbf{b} are chosen, as described in Section 8.3.1.

Modelling errors

9.1 Introduction

We saw in Chapters 5 and 6 that channel estimation and prediction performance is dictated by channel fading characteristics and coherence bandwidth. We assumed that the channel properties could be perfectly described by a linear state space model. The inherent difficulties in determining the parameters of such a model give us good reason to assume that the estimation/prediction performance will be influenced not only by the behaviour of the actual channel, but also by the accuracy to which the channel model parameters can be determined.

The impact of modelling errors on optimal linear filters was investigated mainly during the sixties and seventies. Examples include [98], which considers uncertainties in the initial (prior) distribution of the states, and [99], which extends the results in [98] to the case of uncertain noise covariance matrices.

Model parameter uncertainties are relatively rarely considered in digital communications, partly because communications systems commonly include error detecting and correcting coding functionality that can compensate for modelling errors. Among the sparse literature on the subject we may mention [100] and [29], who considered the impact of modelling errors in channel estimation performance in OFDM systems when blockwise MMSE filters are employed.

In this chapter we will derive expressions for the error covariance for the channel coefficient estimates when erroneous models are employed. Such expressions let us study the impact of modelling errors in different scenarios. For illustration, we examine a few case studies.

9.2 Theory

In this section we present two theorems that are useful when considering erroneous channel models. Theorem 9.2.1 assumes *structural consistency*, so that the matrices in the state space models, used to represent the modelled channel and the actual channel, must have the same dimensions. Theorem 9.2.2 is more general in that it does not assume structural consistency.

9.2.1 Model

Let a process of state vectors $\{\mathbf{x}_t\}$ and measurement vectors $\{\mathbf{y}_t\}$ be described by a linear model

$$\mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{u}_t, \quad (9.2.1)$$

$$\mathbf{y}_t = \mathbf{J}_t \mathbf{x}_t + \mathbf{v}_t, \quad (9.2.2)$$

in which $t \geq 0$ and the zero-mean Gaussian processes \mathbf{u}_t and \mathbf{v}_t are white with known covariance matrix

$$\left\| \left[\begin{array}{c} \mathbf{u}_t \\ \mathbf{v}_t \end{array} \right] \right\|^2 = \left[\begin{array}{cc} \mathbf{Q}_t & 0 \\ 0 & \mathbf{R}_t \end{array} \right]. \quad (9.2.3)$$

The initial state \mathbf{x}_0 is zero-mean Gaussian with covariance matrix Π_0 and uncorrelated with \mathbf{u}_t and \mathbf{v}_t . It is well-known that the optimal filtered state estimation vector $\hat{\mathbf{x}}_{t|t}$ and the optimal one-step state prediction vector $\hat{\mathbf{x}}_t$ are given by the Kalman filter recursions as we have seen in Chapter 3, although we here consider general time-varying matrices:

$$\mathbf{R}_{e,t} = \mathbf{J}_t \mathbf{P}_t \mathbf{J}_t^* + \mathbf{R}_t, \quad (9.2.4)$$

$$\mathbf{K}_{f,t} = \mathbf{P}_t \mathbf{J}_t^* \mathbf{R}_{e,t}^{-1}, \quad (9.2.5)$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_t + \mathbf{K}_{f,t} (\mathbf{y}_t - \mathbf{J}_t \hat{\mathbf{x}}_t), \quad (9.2.6)$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_{f,t} \mathbf{J}_t) \mathbf{P}_t, \quad (9.2.7)$$

$$\hat{\mathbf{x}}_{t+1} = \mathbf{F}_t \hat{\mathbf{x}}_{t|t}, \quad (9.2.8)$$

$$\mathbf{P}_{t+1} = \mathbf{F}_t \mathbf{P}_{t|t} \mathbf{F}_t^* + \mathbf{G}_t \mathbf{Q}_t \mathbf{G}_t^*. \quad (9.2.9)$$

The error covariance for the filtered estimates and the one-step predictions are given by $\mathbf{P}_{t|t}$ and \mathbf{P}_t , respectively. As in previous chapters, we use the short-hand notation $\hat{\mathbf{x}}_{t|t+1} \triangleq \hat{\mathbf{x}}_t$ and $\mathbf{P}_{t|t-1} \triangleq \mathbf{P}_t$. The measurement update (9.2.6) and time update (9.2.8) for the state estimation vector can be combined into the predictor update

$$\hat{\mathbf{x}}_{t+1} = \mathbf{F}_t \hat{\mathbf{x}}_t + \mathbf{K}_{p,t} \mathbf{e}_t, \quad (9.2.10)$$

where the Kalman predictor gain $\mathbf{K}_{p,t} = \mathbf{F}_t \mathbf{K}_{f,t}$ and the innovations $\mathbf{e}_t = \mathbf{y}_t - \mathbf{J}_t \hat{\mathbf{x}}_t$. The predictor update can also be written

$$\hat{\mathbf{x}}_{t+1} = \mathbf{F}_{p,t} \hat{\mathbf{x}}_t + \mathbf{K}_{p,t} \mathbf{y}_t, \quad (9.2.11)$$

where

$$\mathbf{F}_{p,t} = \mathbf{F}_t - \mathbf{K}_{p,t} \mathbf{J}_t. \quad (9.2.12)$$

9.2.2 Modelling errors with structural consistency

We now turn to studying what happens when the fading channel is governed by one model (the signal model), but the Kalman filter is based on another model (the design model), so that we have a modelling error. We first consider modelling errors when the signal model and the design model have the same structure.

Theorem 9.2.1 (The impact of modelling errors). *Consider a standard state space model*

$$\mathbf{x}_{t+1}^a = \mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a, \quad \mathbf{y}_t = \mathbf{J}_t^a \mathbf{x}_t^a + \mathbf{v}_t^a, \quad (9.2.13)$$

where the noises $\{\mathbf{u}_t^a\}$ and $\{\mathbf{v}_t^a\}$ and the initial state \mathbf{x}_0^a are zero-mean white Gaussian and uncorrelated with covariance matrices \mathbf{Q}_t^a , \mathbf{R}_t^a , and Π_0^a , respectively. Suppose that there are modelling errors present, so that one-step predictions of $\{\mathbf{x}_t\}$ are based on the Kalman filter recursions, but with some matrices $\{\mathbf{F}_t^b, \mathbf{G}_t^b, \mathbf{J}_t^b, \Pi_0^b, \mathbf{Q}_t^b, \mathbf{R}_t^b\}$ in place of the real model $\{\mathbf{F}_t^a, \mathbf{G}_t^a, \mathbf{J}_t^a, \Pi_0^a, \mathbf{Q}_t^a, \mathbf{R}_t^a\}$, giving a suboptimal Kalman gain

$$\mathbf{K}_{p,t}^b = \mathbf{F}_t^b \mathbf{P}_t^b \mathbf{J}_t^{b*} (\mathbf{J}_t^b \mathbf{P}_t^b \mathbf{J}_t^{b*} + \mathbf{R}_t^b)^{-1}, \quad (9.2.14)$$

where

$$\mathbf{P}_{t+1}^b = \mathbf{F}_t^b \mathbf{P}_t^b \mathbf{F}_t^{b*} + \mathbf{G}_t^b \mathbf{Q}_t^b \mathbf{G}_t^{b*} - \mathbf{F}_t^b \mathbf{P}_t^b \mathbf{J}_t^{b*} (\mathbf{J}_t^b \mathbf{P}_t^b \mathbf{J}_t^{b*} + \mathbf{R}_t^b)^{-1} \mathbf{J}_t^b \mathbf{P}_t^b \mathbf{F}_t^{b*}, \quad \mathbf{P}_0^b = \Pi_0^b, \quad (9.2.15)$$

that is, the usual discrete-time Riccati difference equation (4.3.18), but based on the design model. Assume structural consistency so that $\{\mathbf{x}_t^a, \hat{\mathbf{x}}_t^b\}$, $\{\mathbf{F}_t^a, \mathbf{F}_t^b\}$, $\{\mathbf{G}_t^a \mathbf{Q}_t^a \mathbf{G}_t^{a*}, \mathbf{G}_t^b \mathbf{Q}_t^b \mathbf{G}_t^{b*}\}$, $\{\mathbf{J}_t^a, \mathbf{J}_t^b\}$, and $\{\mathbf{R}_t^a, \mathbf{R}_t^b\}$ have pairwise the same dimensions. Then, the state-error covariance $\mathbf{P}_t = \|\mathbf{x}_t^a - \hat{\mathbf{x}}_t^b\|^2$ will obey the following recursion:

$$\begin{aligned} \mathbf{P}_{t+1} &= \mathbf{F}_{p,t}^b \mathbf{P}_t \mathbf{F}_{p,t}^{b*} + \mathbf{G}_t^a \mathbf{Q}_t^a \mathbf{G}_t^{a*} + \mathbf{K}_{p,t}^b \mathbf{R}_t^a \mathbf{K}_{p,t}^{b*} \\ &\quad + (\mathbf{F}_{p,t}^b + \Delta \mathbf{F}_{p,t}) \Pi_t^a (\mathbf{F}_{p,t}^b + \Delta \mathbf{F}_{p,t})^* - \mathbf{F}_{p,t}^b \Pi_t^a \mathbf{F}_{p,t}^{b*} \\ &\quad - \mathbf{F}_{p,t}^b \mathbf{T}_t \Delta \mathbf{F}_{p,t}^* - \Delta \mathbf{F}_{p,t} \mathbf{T}_t^* \mathbf{F}_{p,t}^{b*}, \quad \mathbf{P}_0^a = \Pi_0^a, \end{aligned} \quad (9.2.16)$$

where $\mathbf{F}_{p,t}^b = \mathbf{F}_t^b - \mathbf{K}_{p,t}^b \mathbf{J}_t^b$ (cf. (9.2.12)) and

$$\Delta \mathbf{F}_{p,t} = (\mathbf{F}_t^a - \mathbf{F}_t^b) - \mathbf{K}_{p,t}^b (\mathbf{J}_t^a - \mathbf{J}_t^b) = \mathbf{F}_t^a - \mathbf{F}_{p,t}^b - \mathbf{K}_{p,t}^b \mathbf{J}_t^a, \quad (9.2.17)$$

and where $\mathbf{T}_t \triangleq \langle \hat{\mathbf{x}}_t^b, \mathbf{x}_t^a \rangle$ and $\Pi_t^a \triangleq \|\mathbf{x}_t^a\|^2$ obey

$$\Pi_{t+1}^a = \mathbf{F}_t^a \Pi_t^a \mathbf{F}_t^{a*} + \mathbf{G}_t^a \mathbf{Q}_t^a \mathbf{G}_t^{a*}, \quad (9.2.18)$$

$$\mathbf{T}_{t+1} = \mathbf{F}_{p,t}^b \mathbf{T}_t \mathbf{F}_t^{a*} + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \Pi_t^a \mathbf{F}_t^{a*}, \quad \mathbf{T}_0 = \mathbf{0}. \quad (9.2.19)$$

Proof. Beginning with (9.2.18) and (9.2.19), the recursion (9.2.18) follows directly from Theorem 3.4.1. Recursion (9.2.19) follows from

$$\begin{aligned} \mathbf{T}_{t+1} &= \langle \hat{\mathbf{x}}_{t+1}^b, \mathbf{x}_{t+1}^a \rangle = \langle \mathbf{F}_{p,t}^b \hat{\mathbf{x}}_t^b + \mathbf{K}_{p,t}^b \mathbf{y}_t, \mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a \rangle \\ &= \langle \mathbf{F}_{p,t}^b \hat{\mathbf{x}}_t^b + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \mathbf{x}_t^a + \mathbf{K}_{p,t}^b \mathbf{v}_t^a, \mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a \rangle \\ &= \langle \mathbf{F}_{p,t}^b \hat{\mathbf{x}}_t^b + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \mathbf{x}_t^a, \mathbf{F}_t^a \mathbf{x}_t^a \rangle \\ &= \mathbf{F}_{p,t}^b \mathbf{T}_t \mathbf{F}_t^{a*} + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \Pi_t^a \mathbf{F}_t^{a*}, \end{aligned} \quad (9.2.20)$$

In the above, we have used the fact that $\langle \hat{\mathbf{x}}_t^b, \mathbf{u}_t^a \rangle = \mathbf{0}$, $\langle \mathbf{x}_t^a, \mathbf{u}_t^a \rangle = \mathbf{0}$, and $\langle \mathbf{v}_t^a, \mathbf{x}_t^a \rangle = \mathbf{0}$. To derive the recursion (9.2.16), first note that $\mathbf{x}_{t+1}^a - \hat{\mathbf{x}}_{t+1}^b$ can be written

$$\begin{aligned} \mathbf{x}_{t+1}^a - \hat{\mathbf{x}}_{t+1}^b &= \mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a - \mathbf{F}_{p,t}^b \hat{\mathbf{x}}_t^b - \mathbf{K}_{p,t}^b \mathbf{J}_t^a \mathbf{x}_t^a - \mathbf{K}_{p,t}^b \mathbf{v}_t^a \\ &= \underbrace{(\mathbf{F}_t^a - \mathbf{K}_{p,t}^b \mathbf{J}_t^a)}_{=\mathbf{F}_{p,t}^b + \Delta \mathbf{F}_{p,t}} \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a - \mathbf{F}_{p,t}^b \hat{\mathbf{x}}_t^b - \mathbf{K}_{p,t}^b \mathbf{v}_t^a \\ &= \mathbf{F}_{p,t}^b (\mathbf{x}_t^a - \hat{\mathbf{x}}_t^b) + \Delta \mathbf{F}_{p,t} \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a - \mathbf{K}_{p,t}^b \mathbf{v}_t^a. \end{aligned} \quad (9.2.21)$$

The one-step prediction error covariance matrix \mathbf{P}_{t+1} can now be expressed as

$$\begin{aligned} \mathbf{P}_{t+1} &= \|\mathbf{x}_{t+1}^a - \hat{\mathbf{x}}_{t+1}^b\|^2 \\ &= \|\mathbf{F}_{p,t}^b (\mathbf{x}_t^a - \hat{\mathbf{x}}_t^b) + \Delta \mathbf{F}_{p,t} \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a - \mathbf{K}_{p,t}^b \mathbf{v}_t^a\|^2 \\ &= \mathbf{F}_{p,t}^b \mathbf{P}_t \mathbf{F}_{p,t}^{b*} + \mathbf{G}_t^a \mathbf{Q}_t^a \mathbf{G}_t^{a*} + \mathbf{K}_{p,t}^b \mathbf{R}_t^a \mathbf{K}_{p,t}^{b*} \\ &\quad + \mathbf{F}_{p,t}^b \Pi_t^a \Delta \mathbf{F}_{p,t}^* + \Delta \mathbf{F}_{p,t} \Pi_t^a \mathbf{F}_{p,t}^{b*} + \Delta \mathbf{F}_{p,t} \Pi_t^a \Delta \mathbf{F}_{p,t}^* \\ &\quad - \mathbf{F}_{p,t}^b \mathbf{T}_t \Delta \mathbf{F}_{p,t}^* - \Delta \mathbf{F}_{p,t} \mathbf{T}_t^* \mathbf{F}_{p,t}^{b*}. \end{aligned} \quad (9.2.22)$$

To arrive at (9.2.16), it remains to rewrite the Π_t^a -terms in (9.2.22). Making use of the fact that $\mathbf{F}_{p,t}^b + \Delta \mathbf{F}_{p,t} = \mathbf{F}_t^a - \mathbf{K}_{p,t}^b \mathbf{J}_t^a$ by (9.2.17), it follows that

$$\begin{aligned} &\mathbf{F}_{p,t}^b \Pi_t^a \Delta \mathbf{F}_{p,t}^* + \Delta \mathbf{F}_{p,t} \Pi_t^a \mathbf{F}_{p,t}^{b*} + \Delta \mathbf{F}_{p,t} \Pi_t^a \Delta \mathbf{F}_{p,t}^* \\ &= (\Delta \mathbf{F}_{p,t} + \mathbf{F}_{p,t}^b) \Pi_t^a \Delta \mathbf{F}_{p,t}^* + \Delta \mathbf{F}_{p,t} \Pi_t^a \mathbf{F}_{p,t}^{b*} \\ &= (\Delta \mathbf{F}_{p,t} + \mathbf{F}_{p,t}^b) \Pi_t^a ((\mathbf{F}_t^a - \mathbf{K}_{p,t}^b \mathbf{J}_t^a)^* - \mathbf{F}_{p,t}^{b*}) + \Delta \mathbf{F}_{p,t} \Pi_t^a \mathbf{F}_{p,t}^{b*} \\ &= \Delta \mathbf{F}_{p,t} \Pi_t^a (\mathbf{F}_t^a - \mathbf{K}_{p,t}^b \mathbf{J}_t^a)^* + \mathbf{F}_{p,t}^b \Pi_t^a (\mathbf{F}_t^a - \mathbf{K}_{p,t}^b \mathbf{J}_t^a)^* - \mathbf{F}_{p,t}^b \Pi_t^a \mathbf{F}_{p,t}^{b*} \\ &= (\mathbf{F}_{p,t}^b + \Delta \mathbf{F}_{p,t}) \Pi_t^a (\mathbf{F}_{p,t}^b + \Delta \mathbf{F}_{p,t})^* - \mathbf{F}_{p,t}^b \Pi_t^a \mathbf{F}_{p,t}^{b*}, \end{aligned} \quad (9.2.23)$$

which proves the theorem. \square

Note that, if the true model is used so that $\mathbf{F}^a = \mathbf{F}^b = \mathbf{F}$ and so on, so that the superscript a or b are omitted, the recursion (9.2.16) reduces to the ordinary discrete-time Riccati difference equation, which is trivially seen by setting $\Delta\mathbf{F}_{p,t} = 0$. Also, with a correct design model we should have that $\mathbf{T}_t = \langle \hat{\mathbf{x}}_t, \mathbf{x}_t \rangle = \langle \hat{\mathbf{x}}_t, \hat{\mathbf{x}}_t + (\mathbf{x}_t - \hat{\mathbf{x}}_t) \rangle = \|\hat{\mathbf{x}}_t\|^2 = \Sigma_t$ so that \mathbf{T}_t should obey the same recursion as that for Σ_t (see (3.4.30)):

$$\Sigma_{t+1} = \mathbf{F}_t \Sigma_t \mathbf{F}_t^* + \mathbf{K}_{p,t} \mathbf{R}_{e,t} \mathbf{K}_{p,t}^*. \quad (9.2.24)$$

Using the fact that $\mathbf{P}_t = \Pi_t - \Sigma_t$ (see (3.4.31)), we can write

$$\begin{aligned} \Sigma_{t+1} &= \mathbf{F}_t \Sigma_t \mathbf{F}_t^* + \mathbf{K}_{p,t} \mathbf{R}_{e,t} \mathbf{K}_{p,t}^* \\ &= \mathbf{F}_t \Sigma_t \mathbf{F}_t^* + \mathbf{K}_{p,t} \mathbf{J}_t \mathbf{P}_t \mathbf{F}_t^* \\ &= \mathbf{F}_t \Sigma_t \mathbf{F}_t^* + \mathbf{K}_{p,t} \mathbf{J}_t (\Pi_t - \Sigma_t) \mathbf{F}_t^* \\ &= (\mathbf{F}_t + \mathbf{K}_{p,t} \mathbf{J}_t) \Sigma_t \mathbf{F}_t^* + \mathbf{K}_{p,t} \mathbf{J}_t \Pi_t \mathbf{F}_t^* \\ &= \mathbf{F}_{p,t} \Sigma_t \mathbf{F}_t^* + \mathbf{K}_{p,t} \mathbf{J}_t \Pi_t \mathbf{F}_t^*, \end{aligned} \quad (9.2.25)$$

which is the same form as (9.2.19).

Another fact that can be seen from (9.2.16), and was first pointed out by [98], is this: if the only errors present in the model lie in the covariance matrices Π_0 , \mathbf{Q}_t , and \mathbf{R}_t (so that $\Delta\mathbf{F}_{p,t} = 0$), such that

$$\Pi_0^b \geq \Pi_0^a, \quad \mathbf{Q}_t^b \geq \mathbf{Q}_t^a, \quad \mathbf{R}_t^b \geq \mathbf{R}_t^a, \quad (9.2.26)$$

then it holds that

$$\mathbf{P}_t^b \geq \mathbf{P}_t^a. \quad (9.2.27)$$

This can be seen by writing (9.2.15) on the form

$$\mathbf{P}_{t+1}^b = \mathbf{F}_{p,t}^b \mathbf{P}_t^b \mathbf{F}_{p,t}^{b*} + \mathbf{G}_t^b \mathbf{Q}_t^b \mathbf{G}_t^{b*} + \mathbf{K}_{p,t}^b \mathbf{R}_t^b \mathbf{K}_{p,t}^{b*}, \quad (9.2.28)$$

and comparing this with (9.2.16) when $\Delta\mathbf{F}_{p,t} = 0$,

$$\mathbf{P}_{t+1}^a = \mathbf{F}_{p,t}^a \mathbf{P}_t^a \mathbf{F}_{p,t}^{a*} + \mathbf{G}_t^a \mathbf{Q}_t^a \mathbf{G}_t^{a*} + \mathbf{K}_{p,t}^a \mathbf{R}_t^a \mathbf{K}_{p,t}^{a*}. \quad (9.2.29)$$

This means that if upper bounds for Π_0 , \mathbf{Q}_t , and \mathbf{R}_t are used, then the actual state estimation error covariance \mathbf{P}_t is upper bounded by the calculated quantity \mathbf{P}_t^b .

9.2.3 Modelling errors with structural inconsistency

In Section 9.2.2 we demanded that the true model and the design model have the same dimensions of the state spaces. We now relax this constraint.

Theorem 9.2.2 (Impact of modelling errors). *Consider the modelling error case described in Theorem 9.2.1, but without the assumption of structural consistency. Consider linear combinations $\mathbf{z}_t^a = \mathbf{A}_t^a \mathbf{x}_t^a$ and $\mathbf{z}_t^b = \mathbf{A}_t^b \mathbf{x}_t^b$ that have the same dimensions. It then holds that*

$$\|\mathbf{z}_t^a - \hat{\mathbf{z}}_t^b\|^2 = \mathbf{A}_t^a \Pi_t^a \mathbf{A}_t^{a*} - \mathbf{A}_t^a \mathbf{T}_t^* \mathbf{A}_t^{b*} - \mathbf{A}_t^b \mathbf{T}_t \mathbf{A}_t^{a*} + \mathbf{A}_t^b \Sigma_t^b \mathbf{A}_t^{b*}, \quad (9.2.30)$$

where Π_t^a and \mathbf{T}_t obey the same recursions as in Theorem 9.2.1, and where

$$\begin{aligned} \Sigma_{t+1}^b &= \mathbf{F}_{p,t}^b \Sigma_t^b \mathbf{F}_{p,t}^{b*} + \mathbf{F}_{p,t}^b \mathbf{T}_t \mathbf{J}_t^{a*} \mathbf{K}_{p,t}^{b*} + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \mathbf{T}_t^* \mathbf{F}_{p,t}^{b*} \\ &\quad + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \Pi_t^a \mathbf{J}_t^{a*} \mathbf{K}_{p,t}^{b*} + \mathbf{K}_{p,t}^b \mathbf{R}_t^a \mathbf{K}_{p,t}^{b*}. \end{aligned} \quad (9.2.31)$$

Proof. The equality (9.2.30) follows trivially. The recursion (9.2.31) follows from (9.2.11) and (9.2.1) as

$$\begin{aligned} \Sigma_{t+1}^b &= \|\hat{\mathbf{x}}_{t+1}^b\|^2 = \|\mathbf{F}_{p,t}^b \hat{\mathbf{x}}_t^b + \mathbf{K}_{p,t}^b \mathbf{y}_t\|^2 \\ &= \|\mathbf{F}_{p,t}^b \hat{\mathbf{x}}_t^b + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \mathbf{x}_t^a + \mathbf{K}_{p,t}^b \mathbf{v}_t^a\|^2 \\ &= \mathbf{F}_{p,t}^b \Sigma_t^b \mathbf{F}_{p,t}^{b*} + \mathbf{F}_{p,t}^b \mathbf{T}_t \mathbf{J}_t^{a*} \mathbf{K}_{p,t}^{b*} + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \mathbf{T}_t^* \mathbf{F}_{p,t}^{b*} \\ &\quad + \mathbf{K}_{p,t}^b \mathbf{J}_t^a \Pi_t^a \mathbf{J}_t^{a*} \mathbf{K}_{p,t}^{b*} + \mathbf{K}_{p,t}^b \mathbf{R}_t^a \mathbf{K}_{p,t}^{b*}, \end{aligned} \quad (9.2.32)$$

hence proving the theorem. \square

Note that we have restricted the discussion to concern errors for one-step predictions. Similar expression as in Theorems 9.2.1 and 9.2.2 can be derived also for many-steps predictions, smoothing estimates, and filtered estimates (see Appendix 9.A). However, such expressions tend to be more complicated than the expressions used here. To keep the presentation as simple as possible, we confine the discussion of the case studies to one-step predictions.

9.3 A few case studies

To demonstrate the use of Theorems 9.2.1 and 9.2.2, we here study a few model error cases. Throughout this section, when not explicitly stated otherwise, we assume a Doppler spectrum as illustrated in Figure 9.1, expressed by a 4th order AR model. The time t_p between pilot bearing OFDM symbols is 200 μs , the carrier frequency is 3 GHz, and the velocity is set to a pedestrian speed of $v = 7.2$ km/h. The SNR was set to 10 dB. On pilot bearing OFDM symbols, pilots are located on every fourth subchannel. The WINNER II C2 NLOS channel model (see Section 2.3.2) was used. The prediction horizon here is $t_p f_c v / c_0 = 0.004$ wavelengths, i.e. the results practically correspond to filtered estimates.

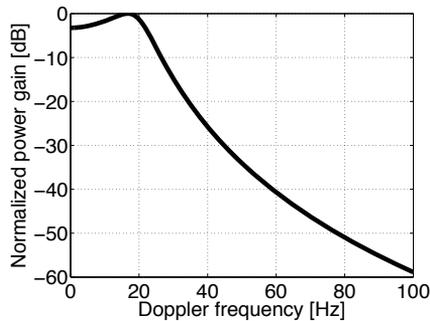


Figure 9.1: Doppler spectrum for a 4th order AR model at pedestrian speed 7.2 km/h.

Because the channel model gives rise to a frequency selective channel, NMSE performance will vary between outer and inner subchannels over a filter band of width w . Assuming that w is even, $w/2$ different NMSE curves are produced. In the figures below, the uppermost curves correspond to the outermost subchannels in a filter band, whereas the lowermost curves correspond to innermost subchannels.

9.3.1 Post-smoothing of channel estimates

As we have seen in Chapter 6, the filter width w is a parameter with which the system designer can set a trade-off between complexity and filtering performance. Filters running in parallel, each taking w adjacent pilot subchannels into account, produce channel estimates (and predictions) over the whole desired bandwidth. The hence produced grid of channel estimates will exhibit “discontinuities” on the borderlines between the filter bands, since no consideration is given to the correlation between adjacent subchannels lying in different filter bands.

Instead of increasing the filter width to reduce the number of discontinuities, which would lead to a considerable increase in computational complexity (see Section 4.4), one could attempt to perform a post-smoothing of the estimates. The smoothing should be carried out over a bandwidth considerably larger than the filter width. We now analyze the utility of such post-smoothing.

Assume that each filter has width w subchannels and that post-smoothing over a band of cw subchannels is to be undertaken. Each filter i produces a vector $\bar{\mathbf{h}}_i$ of channel estimates. Our “measurement” signal is then the c w -vectors $\{\bar{\mathbf{h}}_i\}_{i=1}^c$. Defining $\bar{\mathbf{h}} = [\bar{\mathbf{h}}_1^T, \dots, \bar{\mathbf{h}}_c^T]^T$, we wish to produce a final estimate $\hat{\mathbf{h}}$ based only on $\bar{\mathbf{h}}$. Forming the MMSE estimate (see Section 3.1),

we find that

$$p(\mathbf{h}|\bar{\mathbf{h}}, I) = \mathcal{CN}(\mathbf{h}; \langle \mathbf{h}, \bar{\mathbf{h}} \rangle \|\bar{\mathbf{h}}\|^{-2} \bar{\mathbf{h}}, \|\mathbf{h}\|^2 - \langle \mathbf{h}, \bar{\mathbf{h}} \rangle \|\bar{\mathbf{h}}\|^{-2} \langle \bar{\mathbf{h}}, \mathbf{h} \rangle) \quad (9.3.1)$$

and we are interested in calculating the mean square error (MSE), i.e. the diagonal elements of $\|\mathbf{h}\|^2 - \langle \mathbf{h}, \bar{\mathbf{h}} \rangle \|\bar{\mathbf{h}}\|^{-2} \langle \bar{\mathbf{h}}, \mathbf{h} \rangle$, and compare it with the MSE obtained if one single filter covering the entire band of cw subchannels had been used. We therefore need to calculate the cross-correlation $\langle \mathbf{h}, \bar{\mathbf{h}} \rangle$ and the autocorrelation $\|\bar{\mathbf{h}}\|^2$.

We can do this by considering an equivalent modelling error situation, in which we run a filter of width cw , but erroneously model the subchannels as independent between bands of width w . This is accomplished by letting the covariance matrix for the subchannels in the design model be block-diagonal with blocks of size $w \times w$. The suboptimal estimate $\bar{\mathbf{h}}$ that we then obtain for the cw -vector \mathbf{h} is exactly the same as had we run c parallel filters.

We model the true channel with a state-space model $\{\mathbf{F}, \mathbf{G}, \mathbf{J}, \Pi_0^a, \mathbf{Q}^a, \mathbf{R}\}$ and use the design parameters $\{\mathbf{F}, \mathbf{G}, \mathbf{J}, \Pi_0^b, \mathbf{Q}^b, \mathbf{R}\}$ to represent the erroneous model, as described in Chapter 4. Note that the two models only differ in the process correlation matrices. In this investigation, we consider a one-user system. We may therefore, without losing generality, set the pilot matrix $\Phi_t = \mathbf{I}$, so that the true channel is given by $\mathbf{h} = \mathbf{J}\mathbf{x}^a$ and the sub-optimal one-step predictions are given by $\bar{\mathbf{h}} = \mathbf{J}\hat{\mathbf{x}}^b$. Using Theorem 9.2.1, we find that $\langle \mathbf{h}, \bar{\mathbf{h}} \rangle = \mathbf{J}\langle \mathbf{x}^a, \hat{\mathbf{x}}^b \rangle \mathbf{J}^* = \mathbf{J}\mathbf{T}^* \mathbf{J}^*$ and that $\|\bar{\mathbf{h}}\|^2 = \mathbf{J}\|\hat{\mathbf{x}}^b\|^2 \mathbf{J}^* = \mathbf{J}\Sigma^b \mathbf{J}^*$. In the above, time indices are omitted to indicate that we consider the converged filter.

The matrix $\|\bar{\mathbf{h}}\|^2$ may be badly conditioned so that direct inversion is not advisable. Instead we solve for \mathbf{A} in $(\mathbf{J}\Sigma^b \mathbf{J}^*)\mathbf{A} = \mathbf{J}\mathbf{T}\mathbf{J}^*$. The MSE values of the post-smoothed filter are then found in the diagonal elements of $\|\mathbf{h}\|^2 - \mathbf{J}\mathbf{T}^* \mathbf{J}^* \mathbf{A}$. The prior covariance matrix $\|\mathbf{h}\|^2$ has the SNR along its diagonal, and we here set the SNR to 10 dB.

We evaluate the performance of post-smoothing when the original narrow-band filter spans w subchannels and post-smoothing is carried out over c filter bands. We investigate five settings for the parameters w and c in increasing order of performance: $\{w, c\} = \{1, 16\}$, $\{w, c\} = \{2, 8\}$, $\{w, c\} = \{4, 4\}$, $\{w, c\} = \{8, 2\}$, and $\{w, c\} = \{16, 1\}$, so that the total filter width is always 16 subchannels. In Figure 9.2, the NMSE performance of the one-step prediction after post-smoothing (solid lines) is compared with the NMSE without post-smoothing (dashed lines). We also compare with the optimal performance of a high-complexity filter spanning all 16 subchannels (dotted lines). The NMSE improvement provided by the post-smoothing is quite dramatic, especially for low values of w and high values of c .

In the context of post-smoothing it is worth noting that the narrowband filters, although they perform considerably worse than the post smoothed and

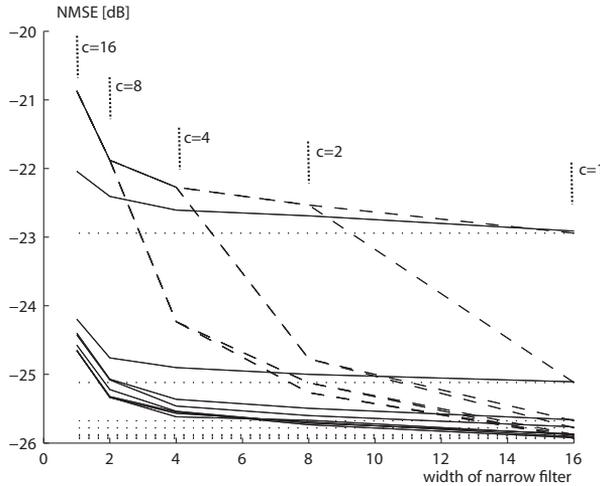


Figure 9.2: Post-smoothing one-step prediction NMSE performance (solid) versus filter width w of the narrowband filter. The post-smoothing is carried out over c parallel filter bands, where c is chosen so that $cw = 16$. Dashed lined mark the performance of the narrowband filters prior to post-smoothing. Dotted lines indicate the performance of an optimal filter operating over 16 subchannels.

the optimal filters, are still optimal, *given measurements over w subchannels only*. The post-smoothed filter, however, is *sub-optimal* given measurements over all 16 subchannels, but optimal given the vector of estimates $\hat{\mathbf{h}}$.

9.3.2 Model order sensitivity

We here study the impact of choosing an inappropriate model order and/or modelling the Doppler spectrum incorrectly. In Figure 9.3, the signal model uses a 4th order AR model for the time dynamics of the channel, giving the Doppler spectrum illustrated in Figure 9.1. The SNR is set to 10 dB. We investigate how the NMSE of the one-step predictions is affected by using a different Doppler spectrum in the design model (although the model order is correct). The model matrices $\{\mathbf{F}^a, \mathbf{G}^a, \mathbf{J}^a, \Pi_0^a, \mathbf{Q}^a, \mathbf{R}^a\}$ for the signal model and the corresponding matrices $\{\mathbf{F}^b, \mathbf{G}^b, \mathbf{J}^b, \Pi_0^b, \mathbf{Q}^b, \mathbf{R}^b\}$ for the design model are constructed according to the framework in Chapter 4. Theorem 9.2.1 (or Theorem 9.2.2) is then used to produce $\|\mathbf{h}^a - \hat{\mathbf{h}}^b\|^2 = \|\mathbf{J}^a \mathbf{x}^a - \mathbf{J}^b \hat{\mathbf{x}}^b\|^2$, from whose diagonal we find the MSE. The solid lines in the figure displays NMSE performance when a range of different Doppler spectra are used in the design model, from a Jakes-like Doppler spectrum with poles close to the unit circle (low flatness), to a flat Doppler spectrum (high flatness). The

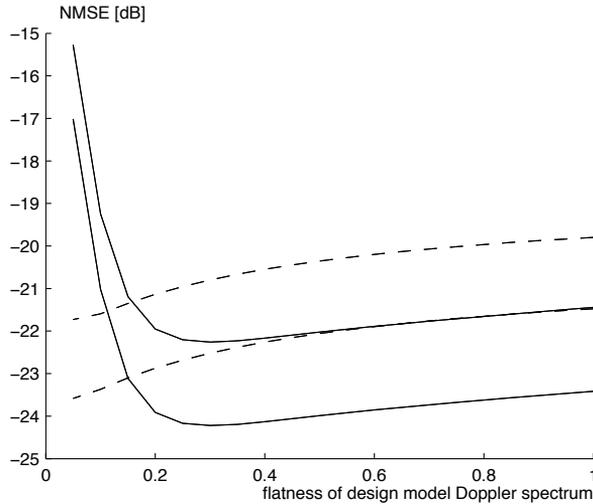


Figure 9.3: One-step prediction NMSE versus flatness of the design Doppler spectrum when the true model has flatness parameter 0.3. Solid lines indicate results for a correct AR model order in the design model (order 4). Dashed lines indicate results for an incorrect model order in the design model (order 2).

true model has flatness 0.3. The degree of flatness is defined so that flatness 1 means that four poles are distributed evenly over a half-circle in the left complex plane, corresponding to a continuous-time Butterworth filter. This filter is then transformed to a discrete-time filter according to Section 4.2.1. For lower values of the flatness, the angles of the respective poles are shifted towards the imaginary line, so that flatness 0 means that all four poles are located on the imaginary line (the unit circle in discrete time). Evidently, the best performance is achieved when the correct model is used. Also, the NMSE performance decays considerably faster if the poles are placed too close to the unit circle, than if they are located too far from the unit circle. It is therefore advisable to use caution when designing the fading model, and place poles somewhat farther from the unit circle than what measurements might suggest.

We also investigate what happens when the incorrect model order is used. Dashed lines in Figure 9.3 indicate NMSE performance when the design model has order 2 instead of the correct model order 4. Note that Theorem 9.2.2 has to be employed here, due to the inconsistency in model structures. The NMSE increases with about 2 dB because of the erroneous model order, but the sensitivity due to placing poles too close to the unit circle is not seen here. This is due to the fact that the lower model order will cause the Doppler spectrum to be smoother than if a high model order is used.

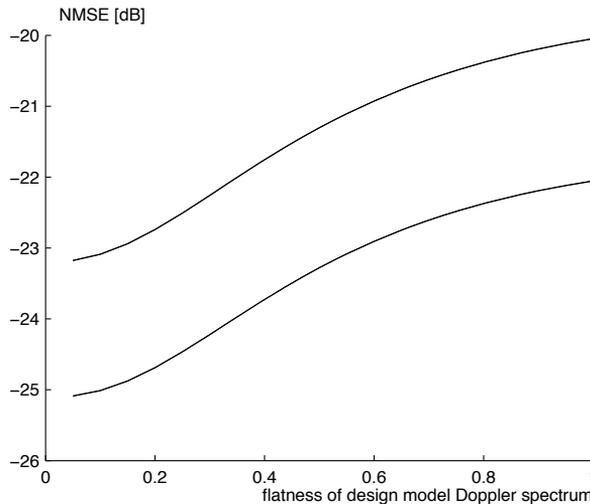


Figure 9.4: One-step prediction NMSE versus flatness of the true Doppler spectrum when the design model is held fixed with a flatness parameter of 0.3. The fading channel is generated by an AR process of order 4, with varying flatness parameter.

Finally in this experiment, we investigate what happens is the design model is held fixed with a Doppler spectrum with flatness 0.3, while the Doppler spectrum in the signal model is allowed to vary. See Figure 9.4. Interestingly, the NMSE performance is better the less flat the true Doppler spectrum is. There is no global minimum when the true model is used. This illustrates the importance of the fading behaviour of the actual channel.

9.3.3 Tap drift

The topic here is to examine the impact of erroneous estimation of delays of taps in a channel impulse response. We consider a flat fading channel, so that the impulse response has a single tap. The true delay of this tap, as represented by the signal model, is 0 s. In the design model, we let the tap delay vary and we study the impact on the NMSE of the one-step predictions that this has. See Figure 9.5. The largest delay estimation error investigated in the figure, 500 ns, corresponds to misjudging the path length by 150 m, which is quite extreme. For moderate errors, say up to 50 ns, the performance decrease is minute for the middle subchannels (lower curve). For the outer subchannels (upper curve), however, the NMSE increase is several dB. The reason is that misjudging delays lead to erroneous expectations on correlation between subchannels. This error is largest for the outer subchannels. It

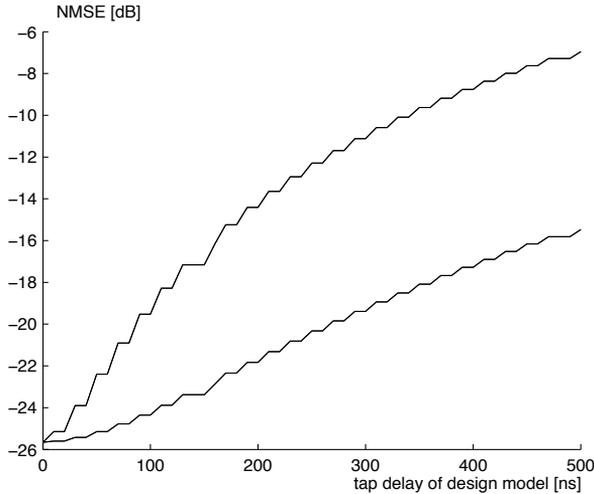


Figure 9.5: If the delay of the single tap in a flat fading channel is erroneously estimated, the estimation and prediction performance will deteriorate. The true tap delay is 0 s. One-step prediction NMSE versus the estimated tap delay is illustrated. The filter width was $w = 4$. Results for outer subchannels (upper curve) and inner subchannels (lower curve) are shown. The SNR is 10 dB.

should be noted here, that we consider misjudging the delay of the entire energy in the impulse response. In a frequency-selective channel where only some of the tap delays are poorly estimated, the performance drop is expected to be less.

9.3.4 Impulse response length

The WINNER II C2 NLOS channel model has 24 taps (see Section 2.3.2). We examine what happens if only a subset of those taps are included in the design model. Figure 9.6 displays NMSE performance of the one-step predictions as a function of the number of taps in the design impulse response. At all times, the total power is normalized so that the design model represents a channel with correct SNR, which we here set to 10 dB, but with the wrong frequency correlation properties. Evidently, capturing as large a part of the impulse response as possible in the design model is crucial for good one-step prediction performance.

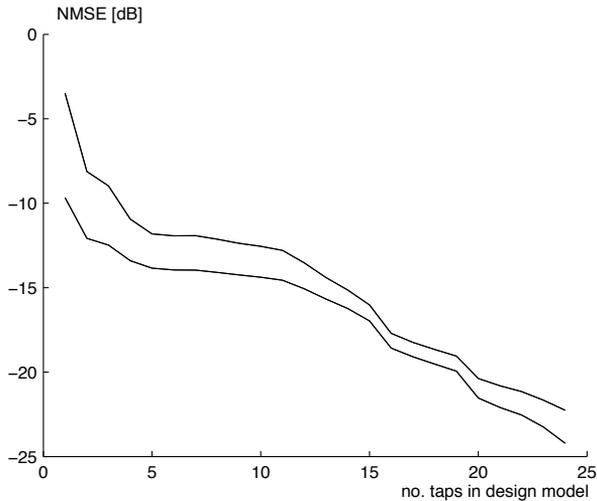


Figure 9.6: One-step prediction NMSE performance for one-step prediction when only the first c out of a total of 24 significant taps in the channel impulse response is included in the design model. The number of taps c is featured on the x-axis. The filter width was $w = 4$. Results for outer subchannels (upper curve) and inner subchannels (lower curve) are shown. The SNR is 10 dB.

9.4 Summary

In this chapter we studied the impact of modelling errors. Two theorems were presented, with which one-step state prediction error covariance matrices can be calculated in cases where the design model used to derive the channel tracking Kalman filter differs from the channel model that exactly describes the statistical properties of the actual channel. The state prediction error covariance matrices may be used e.g. to derive normalized mean square error values for channel predictions. We illustrated the utility of the theorems with a number of examples.

9.A Filtered estimates

In Theorem 9.2.1 and 9.2.2 we derived expressions for the error of the *one-step predictions*, $\|\mathbf{x}_t^a - \hat{\mathbf{x}}_t^b\|^2$, when modelling errors are present. We here investigate how these expressions change when we instead consider the error of the *filtered estimates*, $\|\mathbf{x}_t^a - \hat{\mathbf{x}}_{t|t}^b\|^2$. First, note that

$$\begin{aligned}\hat{\mathbf{x}}_{t+1|t+1}^b &= \hat{\mathbf{x}}_{t+1}^b + \mathbf{K}_{f,t+1}^b(\mathbf{y}_{t+1} - \mathbf{J}_{t+1}^b \hat{\mathbf{x}}_{t+1}^b) \\ &= (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^b) \hat{\mathbf{x}}_{t+1}^b + \mathbf{K}_{f,t+1}^b \mathbf{y}_{t+1} \\ &= (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^b) \mathbf{F}_t^b \hat{\mathbf{x}}_{t|t}^b + \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a \mathbf{x}_{t+1}^a + \mathbf{K}_{f,t+1}^b \mathbf{v}_{t+1}^a.\end{aligned}\tag{9.A.1}$$

Then it follows that

$$\begin{aligned}\|\mathbf{x}_{t+1}^a - \hat{\mathbf{x}}_{t+1|t+1}^b\|^2 &= \\ &= \|\mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a - (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^b) \mathbf{F}_t^b \hat{\mathbf{x}}_{t|t}^b \\ &\quad - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a (\mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a) - \mathbf{K}_{f,t+1}^b \mathbf{v}_{t+1}^a\|^2 \\ &= \|(\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a) \mathbf{F}_t^a \mathbf{x}_t^a + (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a) \mathbf{G}_t^a \mathbf{u}_t^a \\ &\quad - (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^b) \mathbf{F}_t^b \hat{\mathbf{x}}_{t|t}^b - \mathbf{K}_{f,t+1}^b \mathbf{v}_{t+1}^a\|^2 \\ &= \|(\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^b) \mathbf{F}_t^b (\mathbf{x}_t^a - \hat{\mathbf{x}}_{t|t}^b) + ((\mathbf{F}_t^a - \mathbf{F}_t^b) - \mathbf{K}_{f,t+1}^b (\mathbf{J}_{t+1}^a - \mathbf{J}_{t+1}^b)) \mathbf{x}_t^a \\ &\quad + (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a) \mathbf{G}_t^a \mathbf{u}_t^a - \mathbf{K}_{f,t+1}^b \mathbf{v}_{t+1}^a\|^2\end{aligned}\tag{9.A.2}$$

This is to be compared with the second line in (9.2.22). We also need a recursion for $\langle \hat{\mathbf{x}}_{t|t}^b, \mathbf{x}_t^a \rangle$, which corresponds to \mathbf{T}_t in Theorem 9.2.1:

$$\begin{aligned}\langle \hat{\mathbf{x}}_{t+1|t+1}^b, \mathbf{x}_{t+1}^a \rangle &= \\ &= \langle (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^b) \mathbf{F}_t^b \hat{\mathbf{x}}_{t|t}^b + \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a (\mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a) \\ &\quad + \mathbf{K}_{f,t+1}^b \mathbf{v}_{t+1}^a, \mathbf{F}_t^a \mathbf{x}_t^a + \mathbf{G}_t^a \mathbf{u}_t^a \rangle \\ &= (\mathbf{I} - \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^b) \mathbf{F}_t^b \langle \hat{\mathbf{x}}_{t|t}^b, \mathbf{x}_t^a \rangle \mathbf{F}_t^{a*} + \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a \mathbf{\Pi}_t^a \mathbf{F}_t^{a*} \\ &\quad + \mathbf{K}_{f,t+1}^b \mathbf{J}_{t+1}^a \mathbf{G}_t^a \mathbf{Q}_t^a \mathbf{G}_t^{a*}.\end{aligned}\tag{9.A.3}$$

Similar recursions can be derived for many-steps predictions or for smoothing, but apparently, the above expressions look more intimidating than the one-step expressions, which is why we have restricted the exposition to one-step predictions in this chapter.

Bibliography

- [1] T. Eyceoz, A. Duel-Hallen, and H. Hallen. Deterministic channel modeling and long range prediction of fast fading mobile radio channels. *Communications Letters, IEEE*, 2(9):254–256, 1998.
- [2] S. Hu, H. Hallen, and A. Duel-Hallen. Physical channel modeling, adaptive prediction and transmitter diversity for flat fading mobile channel. In *Signal Processing Advances in Wireless Communications, 1999. SPAWC'99. 1999 2nd IEEE Workshop on*, pages 387–390. IEEE, 1999.
- [3] T. Eyceoz, S. Hu, and A. Duel-Hallen. Performance analysis of long range prediction for fast fading channels. In *Proc. of 33rd Annual Conf. on Inform. Sciences and Systems CISS'99*, volume 2, pages 656–661, 1999.
- [4] T. Ekman. *Prediction of Mobile Radio Channels : Modeling and Design*. PhD thesis, Uppsala University, Signals and Systems Group, 2002.
- [5] S. Semmelrodt and R. Kattenbach. Investigation of different fading forecast schemes for flat fading radio channels. In *Vehicular Technology Conference, 2003. VTC 2003-Fall*, volume 1, pages 149–153, 2003.
- [6] A. Duel-Hallen. Fading channel prediction for mobile radio adaptive transmission systems. *Proceedings of the IEEE*, 95(12):2299–2313, 2007.
- [7] L. Lindbom. *A Wiener Filtering Approach to the Design of Tracking Algorithms, with Applications in Mobile Radio Communications*. PhD thesis, Uppsala University, Signals and Systems Group, 1995.
- [8] T.S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, 2 edition, 2002.

-
- [9] I.C. Wong and B.L. Evans. Joint channel estimation and prediction for OFDM systems. In *Proc. IEEE Globecom 2005*, volume 4, pages 2255–2259, 2005.
- [10] I.C. Wong and B.L. Evans. Low-complexity adaptive high-resolution channel prediction for OFDM systems. In *Proc. IEEE Globecom 2006*, 2006.
- [11] P.D. Teal and R.G. Vaughan. Simulation and performance bounds for real-time prediction of the mobile multipath channel. In *IEEE Signal Processing Workshop on Statistical Signal Processing*, pages 548–551, 2001.
- [12] R.G. Vaughan, P.D. Teal, and R. Raich. Short-term mobile channel prediction using discrete scatterer propagation model and subspace signal processing algorithms. In *Vehicular Technology Conference, 2000. IEEE VTS-Fall 2000*, volume 2, pages 751–758, 2000.
- [13] M. Chen, T. Ekman, and M. Viberg. New approaches for channel prediction based on sinusoidal modeling. *EURASIP Journal on Applied Signal Processing*, 2007(1):197, 2007.
- [14] J. B. Andersen, J. Jensen, S. H. Jensen, and F. Frederiksen. Prediction of future fading based on past measurements. In *IEEE Vehicular Technology Conference, 1999*, volume 1, pages 151–155, Amsterdam, September 1999.
- [15] L. Dong, G. Xu, and H. Ling. Prediction of fast fading mobile radio channels in wideband communication systems. In *Global Telecommunications Conference, 2001*, volume 6, pages 3287–3291, San Antonio, TX, November 2001.
- [16] I. C. Wong and B. L. Evans. Sinusoidal modeling and adaptive channel prediction in mobile OFDM systems. *IEEE Transactions on Signal Processing*, 56(4):1601–1615, April 2008.
- [17] J. Hwang and J. H. Winters. Sinusoidal modeling and prediction of fast fading processes. In *Global Telecommunications Conference, 1998*, volume 2, pages 892–897, Sydney, NSW, November 1998.
- [18] K.E. Baddour and N.C. Beaulieu. Improved pilot-assisted prediction of unknown time-selective Rayleigh channels. In *Proc. on IEEE International Conference on Communications, 2006. ICC'06.*, volume 11, pages 5192–5199. IEEE, 2006.

- [19] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.
- [20] Z. Yuanjin. A novel channel estimation and tracking method for wireless OFDM systems based on pilots and Kalman filtering. *IEEE Transactions on Consumer Electronics*, 49(2):275–283, May 2003.
- [21] T. Y. Al-Naffouri and A. Paulraj. A forward-backward Kalman filter for the estimation of time-variant channels in OFDM. In *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications, 2005*, pages 670–674, June 2005.
- [22] T. Y. Al-Naffouri. An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM. *IEEE Transactions on Signal Processing*, 55:3924–3930, July 2007.
- [23] D. Schafhuber, G. Matz, and F. Hlawatsch. Kalman tracking of time-varying channels in wireless MIMO-OFDM systems. In *Asilomar Conference on Signals, Systems and Computers, 2003*, volume 2, pages 1261–1265, November 2003.
- [24] K. J. Kim, J. Yue, R. A. Iltis, and J. D. Gibson. A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems. *IEEE Transactions on Wireless Communications*, 4(2):710–721, March 2005.
- [25] Z. Cheng and D. Dahlhaus. Time versus frequency domain channel tracking using Kalman filters for OFDM systems with antenna arrays. In *IEEE Vehicular Technology Conference, 2003 spring*, volume 1, pages 651–655, April 2003.
- [26] W. Chen and R. Zhang. Kalman-filter channel estimator for OFDM systems in time and frequency-selective fading environment. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 377–380, May 2004.
- [27] S. B. Bulumulla, S. A. Kassam, and S. S. Venkatesh. A systematic approach to detecting OFDM signals in a fading channel. *IEEE Transactions on Communications*, 48(5):725–728, May 2000.
- [28] T. Roman, M. Enescu, and V. Koivunen. Time-domain method for tracking dispersive channels in MIMO OFDM systems. In *International Conference on Multimedia and Expo (ICME)*, volume 2. IEEE, 2003.

-
- [29] O. Edfors, M. Sandell, J. J. van de Beek, S. K. Wilson, and P. O. Börjesson. OFDM channel estimation by singular value decomposition. *IEEE Transactions on Communications*, 46(7):931–939, July 1998.
- [30] 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 (ICASSP)*, volume 3, pages 1845–1848, April 1997.
- [31] IST WINNER and WINNER II projects. Online: <https://ist-winner.org>. Partly funded by the European Commission.
- [32] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.
- [33] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming. *3G Evolution: HSPA and LTE for Mobile Broadband*. Academic Press, 2007.
- [34] Overview of 3GPP release 8. Online: www.3gpp.org.
- [35] M. Wennström. *On MIMO Systems and Adaptive Arrays for Wireless Communication*. PhD thesis, Uppsala University, Signals and Systems Group, 2002.
- [36] A.M. Tulino and S. Verdú. *Random Matrix Theory and Wireless Communications*. Now Publishers Inc., 2004.
- [37] W.C. Jakes. *Microwave Mobile Communications*. Wiley, New York, 1974.
- [38] IST-4-027756 WINNER II, D1.1.2v1.2: WINNER II channel models. Online: <https://ist-winner.org>, November 2007.
- [39] IST-4-027756 WINNER II, D7.1.5 v1.0: Final report. Online: <https://ist-winner.org>, February 2008.
- [40] IST-2003-507581 WINNER, D5.4 v1.4: Final report on link level and system level channel models. Online: <https://ist-winner.org>, November 2005.
- [41] S. Stiglmayr, M. Bossert, and E. Costa. Adaptive coding and modulation in OFDM systems using BICM and rate-compatible punctured codes. In *European Wireless*, Barcelona, 2007.
- [42] IEEE standard for local and metropolitan area networks, part 16: Air interface for fixed broadband wireless access systems, IEEE P802.16a/D12, February 2005.

-
- [43] D. Aronsson. Channel estimation and prediction from a Bayesian perspective. Technical report, Uppsala University, 2007.
- [44] E. T. Jaynes. *Probability Theory: the Logic of Science*. Cambridge University Press, 2003.
- [45] R. T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 1946.
- [46] D. S. Sivia. *Data Analysis: a Bayesian Tutorial*. Oxford University Press, USA, 1996.
- [47] A. Duel-Hallen, S. Hu, and H. Hallen. Long-range prediction of fading signals. *Signal Processing Magazine, IEEE*, 17(3):62–75, 2000.
- [48] T. Söderström. *Discrete-time Stochastic Systems*. Springer, 2 edition, 2002.
- [49] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, 2000.
- [50] A. Ben-Artzi and I. Gohberg. Orthogonal polynomials over Hilbert modules. *Operator theory*, 73:96–126, 1994.
- [51] A.E. Bryson and M. Frazier. Smoothing for linear and nonlinear dynamic systems. *Aero. Syst. Div.*, pages 353–364, 1963.
- [52] R.H. Clarke. A statistical theory of mobile-radio reception. *Bell System Technical Journal*, 47(6):957–1000, 1968.
- [53] M.H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996.
- [54] R. Vaughan and J.B. Andersen. *Channels, Propagation and Antennas for Mobile Communications*. Inst. of Electrical Engineers, 2003.
- [55] B. Sklar. *Digital Communications: Fundamentals and Applications*. Prentice-Hall, 2001.
- [56] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.
- [57] T. Ekman, M. Sternad, and A. Ahlén. Unbiased power prediction of Rayleigh fading channels. In *IEEE Vehicular Technology Conference, 2002 fall*, volume 1, pages 280–284, 2002.

- [58] M. Sternad, T. Ekman, and A. Ahlén. Power prediction on broadband channels. In *IEEE Vehicular Technology Conference, 2001 Spring*, volume 4, pages 2328–2332, 2001.
- [59] M. Sternad, L. Lindbom, and A. Ahlén. Wiener design of adaptation algorithms with time-invariant gains. *IEEE Transactions on Signal Processing*, 50(8):1895–1907, 2002.
- [60] M. Sternad and D. Aronsson. Channel estimation and prediction for adaptive OFDM downlinks. In *IEEE Vehicular Technology Conference, 2003 fall*, pages 1283–1287, Orlando, Fla., October 2003.
- [61] M. Verhaegen and P. van Dooren. Numerical aspects of different kalman filter implementations. *IEEE Transactions on Automatic Control*, 31(10):907–917, 1986.
- [62] M.T. Heath. *Scientific Computing: An Introductory Survey*. McGraw-Hill, 2002.
- [63] S. Falahati, A. Svensson, M. Sternad, and H. Mei. Adaptive trellis-coded modulation over predicted flat fading channels. In *IEEE Vehicular Technology Conference, 2003 fall*, volume 3, pages 1532–1536, October 2003.
- [64] S. Falahati, A. Svensson, T. Ekman, and M. Sternad. Adaptive modulation systems for predicted wireless channels. *IEEE Transactions on Communications*, 52(2):307–316, February 2004.
- [65] M. Sternad, T. Svensson, T. Ottosson, A. Ahlén, A. Svensson, and A. Brunström. Towards systems beyond 3G based on adaptive OFDMA transmission. *Proceedings of the IEEE*, 95(12):2432–2455, 2007.
- [66] IST-2003-507581 WINNER, D2.4: Assessment of Adaptive Transmission Technologies. Online: <https://ist-winner.org>, February 2005.
- [67] C. Tidestav and E. Lindskog. Bootstrap equalization. In *IEEE International Conference on Universal Personal Communications, 1998*, volume 2, pages 1221–1225, Florence, Italy, October 1998. IEEE.
- [68] R. Koetter, A. C. Singer, and M. Tüchler. Turbo equalization. *IEEE Signal Processing Magazine*, January 2004.
- [69] Y. Liu, L. Brunel, and J.J. Boutros. EM channel estimation for coded OFDM transmissions over frequency-selective channel. In *IEEE International Symposium on Spread Spectrum Techniques and Applications, 2008*, pages 544–549, 2008.

- [70] ARTIST4G, D2.2: Advanced receiver signal processing techniques: evaluation and characterization. Online: <https://ict-artist4g.eu/>, December 2010.
- [71] J.G. Proakis. *Digital Communications*. McGraw-Hill New York, 4 edition, 2001.
- [72] IST-4-027756 WINNER II, D2.2.2v1.0: WINNER II link adaptation procedures. Online: <https://ist-winner.org>, October 2006.
- [73] T. Svensson, T. Frank, T. Eriksson, D. Aronsson, M. Sternad, and A. Klein. Block interleaved frequency division multiple access for power efficiency, robustness, flexibility, and scalability. *EURASIP Journal on Wireless Communications and Networking*, 2009.
- [74] IST-4-027756 WINNER II, D6.13.14 v1.1: WINNER II System Concept Description. Online: <http://www.signal.uu.se/Publications/wip.html>, January 2008.
- [75] IST-4-027756 WINNER II, D2.3.3: Link level procedures for the WINNER System. Online: <http://www.signal.uu.se/Publications/wip.html>, November 2007.
- [76] C. Lam, D.Falconer, and F. Danilo-Lemoine. Channel estimation for sub-chunk-based DFT-precoded OFDM systems. In *Wireless World Research Forum*, June 2007.
- [77] J. Bonnet and G. Auer. Chunk-based channel estimation for uplink OFDM. In *IEEE Vehicular Technology Conference, 2006-Spring*, May 2006.
- [78] IST-4-027756 WINNER II, D6.13.7: Test Scenarios and Calibration Cases; Issue 2. Online: <http://www.signal.uu.se/Publications/wip.html>, December 2006.
- [79] IST-4-027756 WINNER II, D1.1.2 v1.2 Part 1: Channel Models. Online: <http://www.signal.uu.se/Publications/wip.html>, September 2007.
- [80] Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description: Stage 2 (release 8), 3GPP TS 36. 300, April 2007.
- [81] M. Sternad, T. Ottosson, A. Ahlén, and A. Svensson. Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks. In *IEEE Vehicular Technology Conference, 2003-Fall*, 2003.

- [82] M. Sternad, S. Falahati, T. Svensson, and D. Aronsson. Adaptive TDMA/OFDMA for wide-area coverage and vehicular velocities. In *IST Mobile and Vehicular Summit*, June 2005.
- [83] G. Auer. Analysis of pilot-symbol aided channel estimation for OFDM systems with multiple transmit antennas. In *IEEE International Conference on Communications*, June 2004.
- [84] M. Sternad and D. Aronsson. Channel estimation and prediction for adaptive OFDMA/TDMA uplinks based on overlapping pilots. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2005.
- [85] IST-4-027756 WINNER II, D6.13.7: Test Scenarios and Calibration Cases; Issue 2. Online: <https://ist-winner.org>, December 2006.
- [86] R. Abildgaard Olesen. Channel prediction for coordinated multipoint transmission. Master's thesis, Uppsala University, 2011.
- [87] G. Caire, G. Taricco, and E. Biglieri. Bit-interleaved coded modulation. *IEEE Transactions on Information Theory*, 44(3):927–946, 1998.
- [88] M. Sternad and S. Falahati. Maximizing throughput with adaptive M-QAM based on imperfect channel predictions. In *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004*, volume 3, 2004.
- [89] C. Jones, T. Tian, A. Matache, R. Wesel, and J. Villasenor. Robustness of LDPC codes on periodic fading channels. In *IEEE Global Telecommunications Conference*, volume 2, 2002.
- [90] Y. Li, W.E. Ryan, A.T.I.R. Inc, and PA Yardley. Mutual-information-based adaptive bit-loading algorithms for LDPC-coded OFDM. *IEEE Transactions on Wireless Communications*, 6(5):1670–1680, 2007.
- [91] S.T. Chung and AJ Goldsmith. Degrees of freedom in adaptive modulation: a unified view. *IEEE Transactions on Communications*, 49(9):1561–1571, 2001.
- [92] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada. Distribution of L-values in gray-mapped M²-QAM signals: Exact expressions and simple approximations. In *IEEE Global Telecommunications Conference*, 2007.

-
- [93] L. Szczecinski, A. Alvarado, and R. Feick. Distribution of max-log metrics for QAM-based BICM in fading channels. *IEEE Transactions on Communications*, 57(9):2558–2563, 2009.
- [94] IST-4-027756 Winner II D2.2.3, Modulation and Coding schemes for the Winner II System. Online: www.ist-winner.org, November 2007.
- [95] M. Binelo, A.L.F. de Almeida, J. Medbo, H. Asplund, and F.R.P. Cavalcanti. MIMO Channel Characterization and Capacity Evaluation in an Outdoor Environment. In *Vehicular Technology Conference Fall (VTC 2010-Fall)*, pages 1–5. IEEE, 2010.
- [96] B.K. Lau, J. Medbo, and J. Furuskog. Downlink cooperative MIMO in urban macrocell environments. In *Antennas and Propagation Society International Symposium (APSURSI)*, pages 1–4. IEEE, 2010.
- [97] N.L. Johnson and S. Kotz. *Continuous Univariate Distributions: Distributions in Statistics*. John Wiley & Sons, 1970.
- [98] T. Nishimura. On the a priori information in sequential estimation problems. *IEEE Transactions on Automatic Control*, 11(2):197–204, 1966.
- [99] H. Heffes. The effect of erroneous models on the Kalman filter response. *IEEE Transactions on Automatic Control*, 11(3):541–543, 1966.
- [100] Y.G. Li, LJ Cimini Jr, and NR Sollenberger. Robust channel estimation for OFDM systems with rapid dispersive fading channels. In *IEEE International Conference on Communications*, 1998.