Uppsala University
Signals and Systems

PREDICTION OF MOBILE RADIO CHANNELS
Modeling and Design

Torbjörn Ekman

UPPSALA UNIVERSITY 2002

Dissertation for the degree of Doctor of Philosophy
in Signal Processing at Uppsala University, 2002

ABSTRACT
Ekman, T., 2002. Prediction of Mobile Radio Channels: Modeling and Design, 254 pp. Uppsala. ISBN 91-506-1625-0.

Prediction of the rapidly fading envelope of a mobile radio channel enables a number of capacity improving techniques like fast resource allocation and fast link adaptation. This thesis deals with linear prediction of the complex impulse response of a channel and unbiased quadratic prediction of the power. The design and performance of these predictors depend heavily on the correlation properties of the channel. Models for a channel where the multipath is caused by clusters of scatterers are studied. The correlation for the contribution from a cluster can be approximated as a damped complex sinusoid. A suitable model for the dynamics of the channel is an ARMA-process. This motivates the use of linear predictors.
A limiting factor in the prediction is the estimation errors on the observed channels. This estimation error, caused by measurement noise and time variations, is analyzed for a block based least squares algorithm which operates on a Jakes channel model. Efficient noise reduction on the estimated channel impulse responses can be obtained with Wiener-smoothers that are based on simple models for the dynamics of the channel combined with estimates of the variance of the estimation error.
Power prediction that is based on the squared magnitude of the linear prediction of the taps will be biased. Hence, a bias compensated power predictor is proposed and the optimal prediction coefficients are derived for the Rayleigh fading channel. The corresponding probability density functions for the predicted power are also derived. A performance evaluation of the prediction algorithm is carried out on measured broadband mobile radio channels. The performance is highly dependent on the variance of the estimation error and the dynamics of the individual taps.

Keywords: Mobile radio channel, fading, channel model, channel estimation, channel prediction, power prediction

Torbjörn Ekman, Signals and Systems, Uppsala University, P O Box 528, SE-751 20 Uppsala, Sweden. Email: torbjorn.ekman@signal.uu.se.

ISBN 91-506-1625-0
Printed in Sweden by Elanders Gotab AB, Vällingby 2002
Distributed by Signals and Systems, Uppsala University, Uppsala, Sweden

To the teachers

For my loved ones

Acknowledgments

It is all about communication, connecting people. In relation to this thesis I would especially like to express my gratitude for the communication with and connection to the following people.

My patient supervisors, Prof. Anders Ahlén and Prof. Mikael Sternad for thoroughness and enthusiasm, respectively. Prof. Gernot Kubin for extending my views with good humour. All three have been highly involved in the work leading to this thesis.

Dr. Andreas Jakobsson for invaluable support and good fun. Tekn Lic. Nilo Casimiro Ericsson for being more than just a colleague. Dr. Sorour Falahati for applications. Dr. Claes Tidestav for relevance. Dr. Erik Lindskog for intriguing discussions. Dr. David Gesbert for a new field. Prof. J. Bach Andersen for clarification. Prof. Petre Stoica for his fast mind and thorough knowledge. Dr. Sören Andersson, Henrik Asplund and Jan-Erik Berg at Ericsson Radio Systems AB, for providing measurement data. Staffan Ångman and Anneli Fjordevik for unsurpassed hospitality. My parents, Ruth and Elon Ekman, for the opportunity.

All the people at the groups Signals and Systems, System and Control at Uppsala University, where I always have felt at home. The colleagues at the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, where I spent some of my most exciting years, hitherto. The Digital Signal Processing group at the University of Oslo, for their kind hospitality. UniK - University Graduate Center for hiring me.

My much beloved Kathrine Vangen for making it all worth while. Our son Ruben for true perspective and Gulliver, the cat, for devotion and distraction.

Torbjörn Ekman
Oslo, September 2002.

Contents

List of symbols

Symbols	
*	Complex conjugate
^{H}	Complex conjugate transpose of a vector or a matrix
^{T}	Transpose of a vector or a matrix
$\jmath$	$\sqrt{-1}$
$h(t)$	Scalar time varying channel
$\mathbf{h}(t)$	Parameter vector with a time varying impulse response
$h_k(t)$	Tap k of the impulse response $\mathbf{h}(t)$
$\hat{h}_k(t+L\|t)$	Tap prediction at time $t+L$ based on data up to time t
$\tilde{h}_k(t-M\|t)$	Smoothed tap at time $t-M$ based on data up to time t
$p_k(t)$	$\|h_k(t)\|^2$, power of tap k
ω	Frequency measured in rad/s
f	Frequency measured in Hz
f_c	Carrier frequency [Hz]
λ	A wavelength
ω_D	The maximum Doppler frequency in rad/s
f_D	The maximum Doppler frequency in Hz
k	Wave number $k = 2\pi/\lambda$
$\mathbf{k}_n(t)$	Wave vector for wavefront n
$\theta_n(t)$	Angle of incidence for wavefront n
$\tau_n(t)$	Path delay for path n
$\phi_n(t)$	Phase function for path n
$f_x(x)$	Probability density function for the stochastic variable x
$\mathrm{E}\{x\}$	The expected value of x
$\boldsymbol{\varphi}(t)$	Regression vector
$\mathbf{R}_\varphi$	Covariance matrix for regressor, $\mathrm{E}\{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\}$
$r_x(\tau)$	The covariance for $x(t)$, $r_x(\tau) = \mathrm{E}\{(x(t) - \mathrm{E}\{x(t)\})(x(t-\tau) - \mathrm{E}\{x(t-\tau)\})^*\}$
σ_x^2	The variance for $x(t)$, $r_x(0)$
$R_x(\omega)$	The Fourier transform of $r_x(\tau)$
$\boldsymbol{\theta}$	Parameter vector with prediction coefficients
Δt	Delay spacing between the samples used in the regressor
L	Prediction range

Remarks on the notation

Signals and coefficients may be complex valued. Vectors are written with bold face, lower case, e.g. $\mathbf{r}$. Matrices are written with bold face, upper case, e.g. $\mathbf{R}$. Elements of vectors or matrices are addressed with subscripts, e.g. $\mathbf{R}_{kl}$ denoting row k column l, unless there is a reference to the original signal, e.g. x, in the subscript. Then the elements are addressed like $[\mathbf{R}_x]_{k,l}$. The Fourier transform of a scalar signal, e.g. $x(t)$, is denoted by the corresponding symbol in upper case, $X(\omega)$ or $X(e^{\jmath\omega})$.

Chapter 1

Introduction

Radio communication has been mobile since the dawn of this technology. Soon after the early work of Marconi, transmitting Morse signals wirelessly across the Atlantic ocean in 1901, most large ships were equipped with a radio telegraph. Technical advances in this field have promptly been adopted by the public. In the last twenty years mobile radio communication has become a consumer product. In some areas the cellular phones are even more common than phones connected to the fixed net.

The first generation of cellular phones were portable analog radio transceivers, used as ordinary phones. The second generation was digital, e.g. GSM and D-AMPS. Cellular phones are still mainly used for oral communication. However, the extensive use of text messages, such as SMS, indicates that people are interested in using their mobile phones as information terminals. The modern digital radio technology is actually a step back to the digital transmission pioneered by Marconi. In those days, the information was coded as Morse signals, now it is transmitted by digitally modulated symbols. The data rates have however increased dramatically.

The challenge for the third and fourth generation of mobile radio systems is no longer to transmit a conversation between two parties, but rather to transmit large quantities of data. The public has become used to fast Internet access and the aim is to provide that service for mobile users too. The radio frequencies used by mobile radio systems have become a scarce resource. To satisfy a growing number of users with increasing demands on the data rates, we need to come up with mobile communication systems that send more bits faster within a given bandwidth. This means that *the quest is for higher spectral efficiency.*

1.1 The problem

One of the main problems in mobile radio communication is the rapid variation (fading) of the signal power at the receiver. This variation in receiving conditions can however be turned into an advantage. In a multiuser system, different users will experience different time-varying receiving conditions. Which user that has the most favorable receiving condition at a given time will vary, due to the fading. If the receiving conditions for the users are known in advance, then adaptive resource allocation can be used to schedule the use of the radio resource between the users [1], [2], [3]. The *multiuser diversity* is exploited, which leads to better utilization of the available radio resources.

Another tool, that can be combined with the scheduling, is link adaptation. Coding, modulation and/or transmit power are then adjusted to the current receiving conditions to exploit the full capacity of the radio link [4], [5], [6], [7], [8].

Both these schemes require that the receiving conditions, that is the channel states, for the users are known in advance at the transmitter. *Prediction of the channel states thus constitutes a crucial component for the adaptive schemes.* This thesis is devoted to the problem of predicting the mobile radio channel.

Fading

The mobile radio channel is the transfer function between sender and transmitter. The situation in mobile radio transmission is quite similar to what you experience when you hear an echo. Someone (the transmitter) is shouting HELLO and after a while you (the receiver) hear HELLO and then finally HELLO. The sound has traveled in a *direct path*, and then there are a number of reflections (echoes) resulting in delayed and damped versions of the signal. In the same manner the radio waves are reflected, scattered, damped and delayed by objects in the environment, resulting in a multipath channel [9]. When the transceiver moves, the delays and dampings change, as the distance to the reflectors is altered.

The radio transmitter uses a carrier radio frequency and then modulates the signal on this carrier. During the transmission, radio waves with this frequency are transmitted continuously. Radio waves traveling along different paths will interact and form a wave pattern, causing the small scale fading. The situation is not that different from the one experienced in a microwave oven when defrosting a frozen lasagna. The microwaves bounce around in the oven and form a wave pattern. Wherever there is constructive interference, the lasagna will be hot, and where there is destructive interference it will remain frozen. The frozen lasagna corresponds to locations in the environment with low received signal power for the receiver, that is, a fading dip. When the mobile transceiver moves through the wave pattern it will encounter alternating fading dips and nodes. The fading is thus a spatial phenomenon that depends on the position of the transceiver in relation to the contributing reflectors/scatterers. Figure 1.1 illustrates a fading pattern for a single carrier at 2 GHz. There is also a temporal aspect, as the reflectors might move, but this is generally a much slower process than the fading due to the movement of the transceiver.

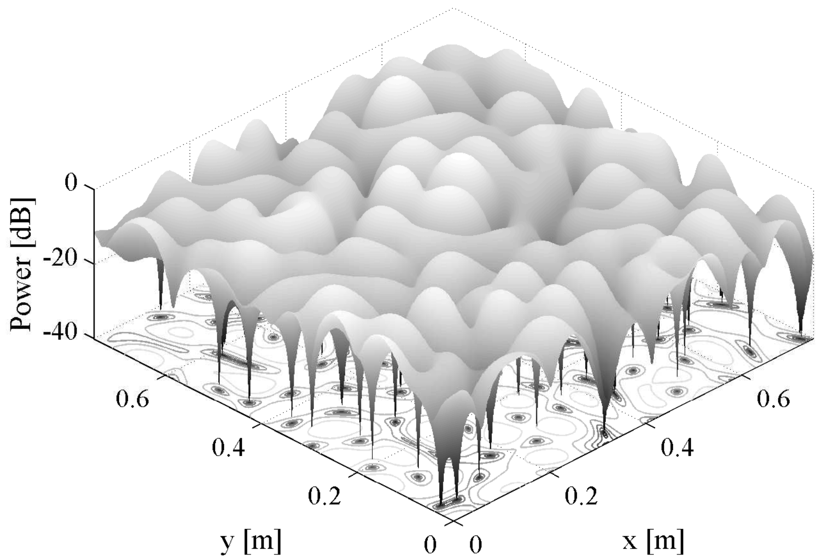

Figure 1.1: Example of a fading pattern for a single carrier with frequency 2 GHz. The multipath propagation environment consists of 50 reflections creating waves arriving from random directions. An area of 0.75×0.75 m is shown. Moving the transceiver just a few centimeters can result in a drop of the power with over 30 dB.

Prediction

The time varying mobile radio channel can be estimated from the received data to obtain snapshots of the impulse response of the channel. Consecutive snapshots give the history of the channel variation. These observations are then used to form a prediction of the future channel impulse response.

There are a number of scales involved in this problem. Even though the channel change on a fraction of a wavelength the environment causing the multipath remains more or less constant for several meters. A model for the radio environment, thus change on a longer scale than the impulse response of the channel. It is possible to exploit this difference in scales to design predictors that function for a large portion of a wavelength.

Fast link adaption requires quite accurate prediction over relatively short prediction ranges, Adaptive resource allocation needs predictions of at least a quarter of a wavelength or more, depending on the velocities of the users. The requirement on the prediction quality is however not that strict in resource allocation problems over these longer ranges.

1.2 Summary of Results and Insights

Channel model

The taps of a broadband mobile radio channel can be modeled as time-varying ARMA or AR-process. The time variability has two components, a slow change due to the changing angles and distances to the reflectors and scatterers and a more abrupt change due to the birth and death process of the reflectors/scatterers. When there are no dominating close scatterers or abrupt changes, the models can be approximated as time invariant for several meters.

Channel estimation

In a broadband communication system the data rate is much faster than the fading rate. The received symbols can be used to obtain frequent snapshots (estimates) of the channel impulse response. These snapshots are obtained at a channel sampling rate that potentially is much faster than the rate of the fading. There are on the order of tens to hundreds of snapshots of the channel for a traveled distance corresponding to a wavelength.

Noise reduction

Simple models for the dynamics of the taps of the impulse response, at the channel sampling rate, together with estimates of the Doppler frequency and estimation error variance can be used to design Wiener-smoothers for the estimated taps. The application of these smoothers to the time series of estimated impulse responses reduces the amount of estimation error on the impulse responses.

Sub-sampled predictors

The tap predictors rely on better models of the dynamics than what is necessary for the noise reduction. The number of predictor coefficients has to be kept low to avoid over fitting of the estimated coefficients, as the amount of data that is stationary will be limited. To be able to obtain reasonable models for the dynamics, utilizing few coefficients, sub-sampled predictors have to be used. Sub-sampling factors on the order of a tenth of a wavelength are found to be sufficient. The sub-sampling can be applied as the noise reduction acts as an anti-aliasing filter.

Direct and indirect predictors

Both direct FIR-predictors and Kalman predictors based on AR and ARMA-models for the complex dynamics of the taps are evaluated. In the Kalman predictor the noise model is designed to give a robust performance of the predictor. The correlation between the taps could be exploited but at the cost of extra complexity and a higher number of estimated coefficients.

Power prediction

To obtain power prediction of the individual taps the squared magnitude of the predictions of the complex valued taps with an added compensation for the bias is used. This power predictor using complex valued regressors, is a quadratic predictor with structural constraints. The coefficients that are MSE optimal for the prediction of a Rayleigh fading tap are shown to be MSE optimal for the proposed unbiased quadratic predictor. The prediction of the total power of the channel is obtained as the sum of the contributions of the taps.

The performance of the predictors on measured impulse responses are seen to be highly dependent on the tap-to-estimation error ratio. The dynamics of the individual taps are also of importance. Taps with lower Doppler spread are generally easier to predict.

Good practice

We have found that the following factors are essential for good prediction:

- The signal power changes abruptly around fading dips, while the channel taps vary smoothly. This is reflected in the different correlation properties of the power and the complex channel. Power prediction based on power is therefore inferior to power prediction based on prediction of the individual complex taps.

- The estimation error on the observed channels limits the performance of the predictors. In highly oversampled channels, it is worth while to filter the data to reduce noise.

- Linear predictors have good generalization properties beyond the data set used for estimating the prediction coefficients.

- Good prediction performance depends on accurate modeling of the fading statistics.

- Direct FIR-predictors or model based Kalman predictors using AR or ARMA models give reasonable performance for prediction of the complex taps.

Scales involved

There are four scales involved in the prediction of the channel:

- The channel sampling, where the snapshots of the channel are estimated from the sampled baseband signal. Data collected from a distance corresponding to a small fraction of a wavelength is used for the channel estimation.

- A separate noise reduction step is performed on the estimated channels. The smoothers use models for the dynamics at the channel sampling rate.

- The prediction is performed at sub-sampled channel sampling rate. The models for the dynamics and the predictors have delay spacings on the order of a tenth of a wavelength.

- The observed noise reduced channel impulse responses are used to estimate the sub-sampled models for the dynamics of the channel, or the predictor coefficients directly. The estimation interval is on the order of meters.

It is the span from the fast baseband sampling rate to the slow change of the models for the dynamics of the channels that is exploited to obtain the channel prediction.

1.3 Outline of the Thesis

1.3.1 Channel models

Chapter 2 gives a review of common channel models and also presents some new results concerning the channel correlation function due to reflections from clusters.

A statistical model

A common approach to describe the channel is to assume that the transmitted radio waves, causing the interference pattern, act as plane wave fronts from different directions, due to the multipath propagation. A statistical description of such mobile radio channels has been developed by Clarke [10] and Jakes [11]. The properties of the channel is deduced from a scattering propagation model which assumes that the field incident to the receiver antenna is composed of an infinite number of randomly phased azimuthal plane waves of arbitrary azimuthal angles. The dynamics of the channel described by the Jakes model is bandlimited. This channel model will be used as a benchmark channel throughout the thesis.

Plane or spherical waves

When the plane wave assumption is valid and the transceiver moves with constant velocity the contribution to the channel from each wave front can be modeled as a complex sinusoid. This is an approximation with limited validity, as we will see in Chapter 2.

The plane wave approximation presuppose that the distance to the source of the waves is long. For scattered waves or waves the are diffracted around corners, spherical or cylindrical waves are more accurate approximations. In Section 2.3 we study how these effects limit the use of the sinusoidal model for the time variation of the channel. This limitation is however only relevant when large contributions to the received power come from scatterers that are close to the transceiver.

Cluster of scatterers

In Section 2.4, we make use of the plane wave approximation but investigate what happens if the assumed reflectors are best described as a cluster of reflectors or scatterers. This is an approach common when channel models for antenna array systems are studied [12], [13]. As buildings generally are not perfectly flat reflecting objects, and the wavelength used for transmission is on the order of the size of a brick, most buildings actually would act more as a cluster of scatterers than a perfect reflector. The geometrical distribution of these clusters are mapped to angular distributions of power around the antenna of the transceiver. For distant clusters, seen under a small angle, an analytical expression for the resulting channel correlation is derived. The contribution to the channel from each cluster can be modeled as a damped (not necessarily exponentially damped) complex sinusoid. These results motivate the use of ARMA-models for the dynamics of the channel. Each tap in the impulse response is then modeled as an ARMA-process that is correlated with the other taps. The poles of the ARMA-model are generally close to the unit circle which implies that the AR-part dominates the behavior. Then an AR-process is an acceptable approximation of the ARMA-process.

The effect on the channel from a building is thus more accurately described by a narrowband filtered Gaussian noise than by a single sinusoid if the assumed reflector actually constitutes a cluster of scatterers. This will limit the performance of any channel predictor, and it tends to favor linear predictors of the complex channel impulse responses.

1.3.2 Estimation errors and noise reduction

Estimation errors in a noisy and time-varying environment

One of the limiting factors in prediction of mobile radio channels is the estimation error on the observed impulse responses. In a broadband communication system, the symbol rate will be very high as compared to the the fading rate. Over a block of symbols the channel can be approximated as time invariant. In Chapter 4, least squares block based estimation of the channel using received symbols is discussed. The effect of the measurement noise is seen to cause an error floor in the estimated impulse responses. The time variation of the channel during the estimation interval is also seen to add to the estimation error. In Section 4.3 this error is analyzed for a Jakes channel estimated by least squares. The estimation error of block based LS estimation of the channel is seen to consist of three independent components [14]: The contribution from the measurement noise, an excess error due to the weighted averaging of the time varying channel and a bias error previously discussed by Lindskog in [15] due to the curvature of the channel. When the estimation interval is only a fraction of a wavelength the excess and bias errors will be small.

Noise reduction of estimated channels

As the symbol rate is very high in a broadband communication system, the channel can be re-estimated quite often in relation to the time constants for the variation of the channel. The channel can thus be oversampled. This can be exploited to reduce the estimation error of the observed impulse responses. It is well known that filtering of the estimated impulse responses, cutting away the frequency content outside the bandlimits given by the maximum Doppler frequency, results in improved channel estimates [16], [17]. In Chapter 5 it is proposed to use IIR and FIR Wiener smoothers based on simple models for the dynamics of the taps together with estimates of the variance of the noise. The Wiener smoother minimizes the smoothing MSE and is thus the preferred solution.

Use of a bank of smoothers to minimize estimation delays

The smoothing lag introduced by the noise reduction operation introduces a corresponding delay when the smoothed taps are used in the regressors for prediction of future values. The predictor has to compensate for the delay by a longer prediction range. We thus want the smoothing-lag to be small. But as a larger smoothing-lag also results in better noise reduction, we also want to use as large smoothing-lags as possible. To circumvent this problem, a bank of noise reduction smoothers with smoothing lags from zero up to as long as needed, can be used. The regressors containing smoothed observations has to be available without delay. Under this condition, each element of the regression vector consists of the smoothed observation with an as large a smoothing-lag as possible.

1.3.3 Prediction of the complex taps

Sinusoidal and statistical models of fading channel coefficients

Given that the taps of the channel act as a sum of sinusoids, then the phase shift and attenuation of each path can be estimated. A prediction of the tap can then be obtained by propagating the estimated sinusoids. This is the most common approach to the prediction of mobile radio channels. Proposed approaches to the estimation of the sinusoids are subspace based methods like root-MUSIC [17] and a modified ESPRIT algorithm [18], [19]. A different approach is to model the sum of sinusoids as an AR-process. The estimated AR-parameters are then used to derive a linear predictor [20], that can be iterated to obtain longer prediction ranges. Overviews of the subspace based methods are found in [21] and the AR-model based prediction is described in [22]. Polynomial filters has also been applied to the prediction of the complex valued channel [23]. The statistical model by Jake describe a bandlimited channel with known correlation function. These properties are exploited in [24] where a continuous time predictor for the Jakes channel is presented. The use of quadratic predictors [25] or MARS-model based predictors [26] for the complex tap results in predictors with poor generalization properties outside the estimation interval [27].

Channel prediction on measured channels

Prediction results for the Jakes channel or channels modeled as a low number of sinusoids should not be taken as an evidence for high predictability of real mobile radio channels. In practice the dynamics of the taps are not perfectly bandlimited, due to birth and death processes of the contributing paths and the changing angles toward the reflectors. The sinusoids will be damped and and their frequencies will change. In this thesis the main focus is on designing predictors that are robust enough to function on real estimated sampled channels. The predictors are therefor *tested on measured broadband mobile radio channels* to validate their performance.

Even though the taps of the sampled channel are mutually correlated in general they are in this thesis predicted separately to reduce the computational complexity. This is suboptimal but reduces the number of coefficients that has to be estimated.

FIR predictors, linear regression

In Chapter 6 an FIR-Wiener-predictor that is designed to minimize the prediction error for a given range is studied. This direct FIR predictor is not depending on a model for the dynamics but depends on estimates of the correlation functions. The structure of the FIR predictor is suitable for taps described by a sum of sinusoids, but it does not depend on that assumption to function. Smoothed regressors with different lags, that give good noise reduction, can easily be used with the direct FIR predictor.

Delay spacing of regressors

The performance of the direct FIR-predictor is found to depend on the delay spacing in the predictor (that is the sub-sampling factor). The optimal delay spacing depends on the tap-to-estimation error ratio, the prediction range and the number of predictor coefficients. The delay spacing for a direct FIR-predictor in Chapter 6 is optimized for taps described by the Jakes model, both for the case with and without noise reduction. The high oversampling of the channel, that is a condition for the noise reduction to perform well, suggests that the predictors should be sub-sampled. A robust choice for the delay spacing of FIR predictors, only depending on the prediction range and the number of predictor coefficients, is proposed. A delay spacing on the order of a tenth of a wavelength is generally a good choice.

The use of sub-sampled predictors can lead to aliasing. The noise reduction pre-processing will however act as an anti-aliasing filter and is thus a crucial component of the predictor.

Model based prediction

The other main prediction approach pursued in this work is the model-based predictor. Sub-sampled AR or ARMA models are used for the dynamics of the taps. These models are then used in a Kalman predictor to obtain predictions at any range that is a multiple of the chosen delay spacing. The AR-model based iterated predictor suggested in [22] is shown to be suboptimal whenever there are estimation errors on the taps. The Kalman predictor with a suitable model for the noise renders better results.

The correlation functions of the taps and the estimation errors change over time and have to be estimated from the observed taps, using a limited amount of data. This limits the number of predictor, or model, coefficients that can be estimated with sufficient accuracy.

1.3.4 Prediction of the power

It usually the power of the channel that is of interest for the system. It can be used to obtain power control or to predict the channel states. Predictors intended for power control usually consider short range prediction, e.g. [28], [29], [30]. The recent interest in long range prediction is due to the developments in link adaptation and scheduling. A common approach to obtain the predicted power of a tap is to use the squared magnitude of the complex prediction, e.g. [22], [21], [18], [17], [23]. When linear predictors are used for the tap prediction the absolute square of the magnitude constitutes a quadratic predictor (with structural constraints) for the power. An alternative approach is to perform prediction directly based on previous estimates of the power. Nonlinear approaches like neural nets [31], [32] and also linear approaches [33] have been proposed.

The quadratic unbiased predictor of the received power

In Chapter 7, the power prediction of the taps based on linear prediction of the complex taps are studied. The taps are assumed to be distributed as complex circular Gaussian stochastic variables. This is a valid assumption if the number of contributing paths is large. (Under the assumption that the reflectors actually are best described as clusters of scatterers, this is a valid assumption even when there is only one contributing cluster.) The taps are thus assumed to be Rayleigh fading, but no assumptions are made about the correlation function. The direct use of the squared magnitude of the linearly predicted complex valued tap, as in [22], is shown to render a biased power predictor. The bias is equal to the variance of the prediction error of the complex tap.

An unbiased quadratic predictor can be obtained by adding a bias compensation term. For a Rayleigh fading tap, it is shown that the same prediction coefficients that are optimal for the linear prediction of the complex valued tap, also are optimal for the quadratic unbiased predictor.

The probability density functions

The probability density functions for the predicted power is of interest in the design of link adaptation systems [7], [34], [35]. These functions are derived for the unbiased quadratic predictor on a Rayleigh fading channel.

The performance on measured channels

The predictors are evaluated on a set of measurements described in Chapter 3. The influence of the estimation error is seen to be one of the limiting factors for prediction. The unbiased quadratic predictor based on the direct FIR predictor and the Kalman predictor using AR-models for the dynamics renders roughly the same performance. The benefit of the unbiased predictor is mainly for prediction ranges above half a wavelength.

1.3.5 Link adaptation

As a final example for what the prediction can be used for, a link adaptation system for a flat fading channel is studied in Chapter 8, which is based on the paper [35]. When adaptive modulation is used to exploit short-term fading in mobile radio channels, signaling delays create problems with outdated channel state information. The use of channel power prediction will improve the performance of the link adaption.

The effect on the link adaptation of uncertainty in the channel states at the transmitter have been discussed in the literature, see e.g. [4], [5],[6], [7]. In Chapter 8, the probability density functions for the predicted power are used to design an optimum adaptive modulation system for a given channel prediction error variance, that maximizes the spectral efficiency while satisfying a certain BER requirement. The proposed system utilizes the unbiased quadratic predictor derived in Chapter 7 to predict the channel quality at the receiver. The predicted SNR for which the system changes modulation, the rate regions, are highly dependent on the variance of the prediction error. Especially is transmission at low predicted SNR avoided by the scheme when the prediction error variance is high. This is due to the high relative prediction error for low SNR, described in Chapter 7.

Chapter 2

Channel Models

This chapter begins with an overview of the commonly used continuous-time and discrete-time models used to describe terrestrial radio communication channels for mobile communication. Such models are based on assumptions of planar wave-fronts, constant vehicle velocity and propagation via reflectors and scattering objects. The channel can then be described as a weighted sum of complex sinusoids with fixed frequencies. The assumptions and approximations in the derivation of this kind of model will be discussed below. Furthermore, a generalization of this model is studied, where spherical instead of plane waves are assumed. Finally the effect of clusters of scatterers and reflectors are discussed. Interested readers are also referred to textbooks on the subject, see for example [36] and [37].

From linear systems theory it is known that sinusoidal based channel models can be perfectly extrapolated in space and time, i.e. predicted, if the number of sinusoids are known and if the predictor is properly tuned. Experience with prediction on measured channels show that this is not the case in reality.

In Section 2.3, the limitations of the simple sinusoidal based model is discussed and a class of models with more general validity is proposed. By taking spherical wavefronts and non-constant vehicle velocities into account, a model is obtained in which the sinusoids may have time-varying frequencies, such as chirps. While it would be unrealistic to expect unbounded predictability in such scenarios, it is still reasonable to believe that prediction over a large fraction of a wavelength of the carrier frequency may be attainable. The rest of this thesis constitutes an exploration of methods with which such long-range predictions might be realized.

The situation with a low number of discrete reflectors and scatterers is an oversimplification of the radio environment. In Section 2.4 we consider clusters of scatterers and reflectors. The geometrical distribution of these clusters are mapped to angular distributions of power around the antenna of the transceiver. For distant clusters, seen under a small angle, an analytical expression for the resulting channel correlation is derived. The contribution to the channel from each cluster can be modeled as a damped (not necessary exponentially damped) complex sinusoid. These results motivate the use of ARMA-models for the dynamics of the channel.

We will in the following always refer to the moving transmitter/receiver as the mobile station. Whether it acts as a receiver or transmitter does not matter, as the channel is the same in both directions due to reciprocity. Thus both the up- and down-links are treated.

2.1 Multipath Propagation via Reflectors and Scatterers

In the classical papers by Clarke [10] and Jakes [11], a statistical passband description of the mobile radio channel is developed. The model treats a scenario where N plane waves are arriving at the receiver from random directions. The different propagation paths cause the waves to have different attenuation and phase shifts. In Figure 2.1, a scenario with a mobile receiving waves from two different directions is depicted. The base-station antenna acts as a point source emitting a spherical electro-magnetic wave. On large distances from the antenna the spherical waves are locally perceived as plane waves. In Figure 2.1 we have no line of sight propagation (direct path) due to the shadowing of a building. The buildings further away will also contribute, although the longer propagation paths result in a higher path loss. In Figure 2.1 two buildings act as *remote reflectors* and the corresponding wave-fronts are considered to be almost plane at the mobile station. A lamppost nearby the mobile would act as a *close scatterer*. Close scatterers cause spherical wave-fronts at the mobile station, whereas rays that are scattered around corners of houses close by cause cylindrical wave-fronts. Further away from the scatterers and the corners the corresponding spherical and cylindrical waves are locally perceived as plane waves. The characteristic size and distance of objects that make them behave as reflectors rather than scatterers is discussed in Appendix 2.A.

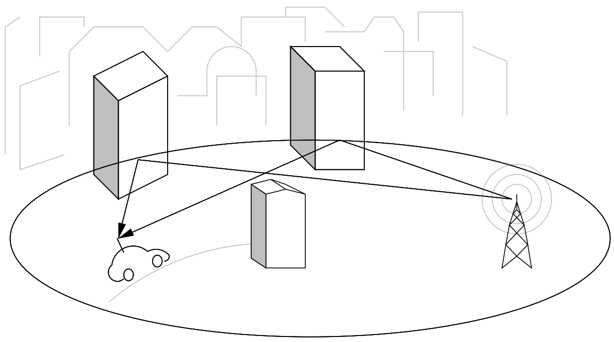

Figure 2.1: A mobile antenna receiving two reflected rays that have the same path distance.

2.2 The Continuous Time Channel for Plane Wave Fronts

The voltage, $V(\mathbf{r}, t)$, which can be measured at an antenna is a function of the spatial electric field and the antenna properties. The incident field at the receiver, $\mathbf{E}(\omega, \mathbf{r}, t)$, is a function of the angular frequency ω, the position $\mathbf{r}$ and the time t.[1] We represent a transmitted single-frequency radio signal as a complex exponential,[2] $e^{\jmath\omega t}$. The complex open circuit voltage at the receiving antenna can then be approximated as the product between $e^{\jmath\omega t}$ and a space-dependent transfer function, $H(\omega, \mathbf{r})$, such that

$$V(\mathbf{r}, t) = H(\omega, \mathbf{r})e^{\jmath\omega t}. \tag{2.1}$$

The in-phase component is obtained as $\mathrm{Re}[V(\mathbf{r}, t)]$ and the quadrature component is $\mathrm{Im}[V(\mathbf{r}, t)]$. The wavefronts contributing to the electric field can be thought of as rays with different delays coming from different directions. The transfer function $H(\omega, \mathbf{r})$ is therefore approximated as the summation of the contributions from reflectors and scatterers according to

$$H(\omega, \mathbf{r}) = \sum_{n=1}^{N} a_n(\mathbf{r})e^{\jmath(\psi_n(\mathbf{r}) - kr_n(\mathbf{r}))}. \tag{2.2}$$

Here $kr_n(\mathbf{r})$ is the (scalar) electrical distance (with $k = 2\pi/\lambda = \omega/c$ being the wave number, where c is the velocity of light) and $r_n(\mathbf{r})$ is the physical

[1]The origin of the coordinate system for the position $\mathbf{r}$, is arbitrary.

[2]The electric and magnetic fields are the real and imaginary parts respectively

distance of path n. The number of contributors N may be arbitrary large. The phase factor $\psi_n(\mathbf{r})$ depends on the properties of the nth reflector. The effective complex gain of contributor n is $a_n(\mathbf{r})e^{\jmath\psi_n(\mathbf{r})}$. It is influenced by, among other things, antenna pattern weighting effects and path attenuation. The attenuation $a_n(\mathbf{r})$ and the phase shift $\psi_n(\mathbf{r})$ depend on physical parameters such as the path distance and the texture of the reflectors. These parameters are fairly constant over short distances. Therefore we shall regard $a_n(\mathbf{r})$ and $\psi_n(\mathbf{r})$ as being space independent, at least over distances of a few wavelengths. Thus we set

$$a_n(\mathbf{r}) = a_n, \tag{2.3}$$
$$\psi_n(\mathbf{r}) = \psi_n. \tag{2.4}$$

In addition we also assume a_n to be independent of ω, at least over the bandwidth around the carrier frequency, used by a mobile radio communication system.

An alternative description of the exponential term of (2.2) uses the scalar product between the wave-vector of the nth path $\mathbf{k}_n(\mathbf{r})$, and the position $\mathbf{r}$,

$$H(\omega, \mathbf{r}) = \sum_{n=1}^{N} a_n(\mathbf{r})e^{\jmath(\varphi_n(\mathbf{r}) - \mathbf{k}_n(\mathbf{r})\cdot\mathbf{r})}. \tag{2.5}$$

A wave-vector is a vector in space that is perpendicular to the wavefront, pointing in the direction of propagation. The magnitude of the wave-vector is the wave number, $|\mathbf{k}| = k = 2\pi/\lambda$. The phase $\varphi_n(\mathbf{r})$ in (2.5) includes the phase shift due to reflection and/or scattering just as $\psi_n(\mathbf{r})$ in (2.2), but $\varphi_n(\mathbf{r})$ also includes a spatially invariant contribution that depends on the choice of the origin of the coordinate system. The two expressions (2.2) and (2.5) are fully equivalent but can be used to illustrate different properties of the channel.

A more detailed description of the exponential term of (2.2) will be derived next.

The electrical distance and the phase of a received sinusoid

The model (2.1), (2.2) is valid in a local region if the phase, $\psi_n(\mathbf{r})$, and the amplitude $a_n(\mathbf{r})$, for each scatterer remain constant in that region. Define a right hand coordinate system with the x-axis along the direction of movement and the y-axis pointing to the left in the plane on which the mobile station is moving (the z-axis will then point up from the plane). Thus the position along the direction of movement is x ($\mathbf{r} = [x, 0, 0]$).

The electrical distance between the base station and the mobile along path n will be a function of x. It can be expressed as

$$kr_n(x) = \omega \frac{r_n(x)}{c} = \omega\tau_n(x), \tag{2.6}$$

where $r_n(x)$ as before is the location-dependent physical distance between the base station and the mobile station along path n and $\tau_n(x)$ is the corresponding path delay.

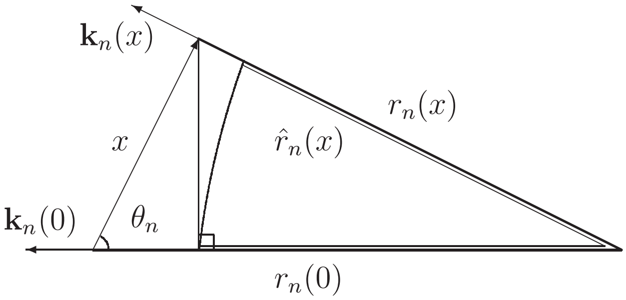

Figure 2.2: The change of path distance when moving the mobile. Here $\hat{r}_n(x)$ is the approximate and $r_n(x)$ is the true distance after moving the distance x.

The difference in electrical distance between two points corresponds to the phase difference for a sinusoid carrier signal. We will below derive an expression for the difference in electrical distance when the mobile moves a distance x at an angle θ_n to the direction of the scatterer/reflector which is causing the incoming wave front. The path distance from the closest point of reflection will then change from $r_n(0)$ to $r_n(x)$ according to

$$r_n^2(x) = r_n^2(0) + x^2 - 2xr_n(0)\cos(\theta_n). \tag{2.7}$$

A second-order Taylor expansion of the square root of this expression around $x = 0$ gives

$$r_n(x) \approx r_n(0) - x\cos(\theta_n) + \frac{x^2\sin^2(\theta_n)}{2r_n(0)}. \tag{2.8}$$

The corresponding change of electrical distance (the change of phase) from position 0 to x is thus approximated by

$$kr_n(x) - kr_n(0) \approx \frac{\omega}{c}\left(-x\cos(\theta_n) + \frac{x^2\sin^2(\theta_n)}{2r_n(0)}\right) \tag{2.9}$$

If the last term in (2.8) is disregarded, then the common linear approximation, cf. Figure 2.2, is obtained as

$$\hat{r}_n(x) = r_n(0) - x\cos(\theta_n). \tag{2.10}$$

This is the plane wave approximation which is valid if the mobile is moving over distances which are small in comparison to the distance to the reflector. For a plane wave the sides in the triangle in Figure 2.2 would be parallel (and thus not forming a triangle) as in Figure 2.3. The corresponding change of

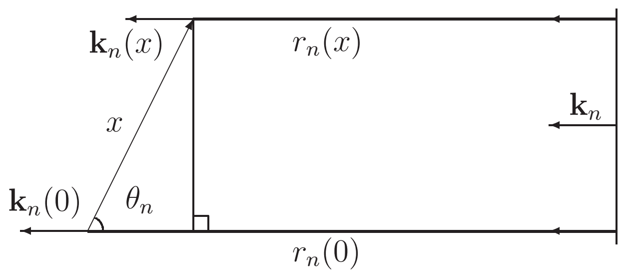

Figure 2.3: The change of path distance when moving the mobile a distance x in a plane wave scenario.

electrical distance is then

$$kr_n(x) - kr_n(0) = \omega\tau_n(x) - \omega\tau_n(0) \approx -\frac{\omega x \cos(\theta_n)}{c}, \tag{2.11}$$

where $\tau_n(\cdot)$ denotes the path delay associated with $r_n(\cdot)$. The change of electrical distance (2.11) can also be easily derived using (2.5) assuming plane waves, that is constant wave-vectors $\mathbf{k}_n$, as

$$\begin{aligned} \mathbf{k}_n \cdot \mathbf{r}(x) - \mathbf{k}_n \cdot \mathbf{r}(0) &= \frac{\omega}{c}[-\cos\theta_n, \sin\theta_n, 0] \cdot [x, 0, 0] - 0 \\ &= -\frac{\omega x \cos(\theta_n)}{c}, \end{aligned} \tag{2.12}$$

where the position vectors are $\mathbf{r}(x) = [x, 0, 0]$, $\mathbf{r}(0) = [0, 0, 0]$ and the wave-vector is $\mathbf{k}_n = \frac{\omega}{c}[-\cos\theta_n, \sin\theta_n, 0]$. The scalar product representation of the phase is especially handy in the plane wave case.

To summarize, movement of the transceiver will change the phase according to (2.9). For plane waves the first order approximation (2.11) will be valid.

A continuous-time baseband description

A frequency component ω of a radio signal with nonzero bandwidth can be expressed as the baseband frequency, ω_b, shifted by the carrier frequency, ω_c

$$\omega = \omega_b + \omega_c. \tag{2.13}$$

The complex open circuit antenna voltage (2.1) can then be expressed as

$$V(\mathbf{r},t) = H_p(\omega_b + \omega_c, \mathbf{r})e^{\jmath(\omega_b+\omega_c)t}, \tag{2.14}$$

where the subscript p on the transfer function denotes that it is the passband channel. The radio transmission is performed at the high frequencies of the passband but it is more convenient to do the signal processing on the symbols in the baseband. The received signal $V(\mathbf{r},t)$ is therefore transfered from the passband to the baseband by multiplication with $e^{-\jmath\omega_c t}$. The corresponding baseband signal $y(\mathbf{r},t)$ can be expressed as

$$\begin{aligned} y(\mathbf{r},t) = V(\mathbf{r},t)e^{-\jmath\omega_c t} &= H_p(\omega_b + \omega_c, \mathbf{r})e^{\jmath\omega_b t} \\ &= H_b(\omega_b, \mathbf{r})e^{\jmath\omega_b t}. \end{aligned} \tag{2.15}$$

The passband channel $H_p(\omega_b + \omega_c, \mathbf{r})$ is thus equivalent to the baseband channel $H_b(\omega_b, \mathbf{r})$ within the band-limits of the system. The model (2.2) for the passband channel is also a model for the baseband channel,

$$H_b(\omega_b, \mathbf{r}) = \sum_{n=1}^{N} a_n(\mathbf{r})e^{\jmath(\psi_n(\mathbf{r}) - kr_n(\mathbf{r}))}. \tag{2.16}$$

For the wave number we obtain

$$k = k_b + k_c = \frac{\omega_b}{c} + \frac{\omega_c}{c}. \tag{2.17}$$

By using the linearization (2.11) of equation (2.9), the electrical distance can be rewritten as

$$kr_n(x) \approx \omega_c\tau_n(0) + \omega_b\tau_n(0) - k_c\cos(\theta_n)x - k_b\cos(\theta_n)x, \tag{2.18}$$

where the dependence of the electrical distance on the local spatial movement is collected in the last two terms. The first term is constant and may be included in a combined complex attenuation and phase factor α_n that collects all (nearly) space-independent factors. The second term is similar to the first term, but since it depends on the baseband frequency we need it for the transfer function description. The fourth term will be small relative to the third term if the relative bandwidth is small, i.e., if ω_b/ω_c is small. In an application with a carrier frequency at 1800 MHz and a bandwidth of 5 MHz the relative bandwidth will be less than 0.3% and the fourth term can thus be neglected. The nearly time-invariant attenuation and phase factors are collected in a complex valued factor α_n as

$$\alpha_n = a_n e^{\jmath(\psi_n - \omega_c\tau_n(0))}. \tag{2.19}$$

Thus, the baseband transfer function (2.16) can for planar waves and straight line motion with constant velocity of the mobile station, be approximated using (2.18) and (2.19) as

$$H(\omega, x) = \sum_{n=1}^{N} \alpha_n e^{\jmath(-\omega\tau_n(0) + \frac{\omega_c}{c} x \cos(\theta_n))}, \tag{2.20}$$

where the subscript b has been dropped, so that ω from now on will denote *baseband frequency*. The phase shifts described by the exponential terms $\frac{\omega_c}{c} x \cos(\theta_n)$ will cause the rapid variation of the channel when moving short distances. Since $\omega_c/c = 2\pi/\lambda_c$, these terms are significant already when the traveled distances is a small fraction of the carrier wavelength.

The parameters a_n, ψ_n and θ_n are also space dependent but on a different scale, governed by the overall geometry, the distance to and the structure of the reflectors. It is reasonable to assume that these parameters will remain fairly constant over at least a small number of wavelengths, whereas the resulting channel varies significantly.

As there exists a parameterized description of the process, where x is the only variable, it is conceivable that there exists a mapping from observations of the channel coefficients to the corresponding coefficients at a location within in a distance of a few wavelengths. Thus, it is likely that the channel parameters can be predicted reasonably well on this small geometrical scale of a few wavelengths. The utilized predictor may, but does not have to, rely on estimates of the parameters a_n, ψ_n and θ_n.

Time dependent transfer function

We have described how the channel depends on the position of the mobile station. By introducing the velocity of the mobile station we can also describe the channel as a function of time. If the vehicle drives straight ahead along the x-axis, at constant velocity, then the traveled distance depends on time as $x = vt$, where v is the speed of the vehicle. This introduces a rotation of each term in the channel (2.20) with an angular frequency

$$\omega_{D_n} = \omega_c \frac{v \cos\theta_n}{c} = kv\cos\theta_n, \tag{2.21}$$

that corresponds to the Doppler-shift of the carrier frequency due to the motion.[3] The maximum Doppler shift ω_D, encountered when driving straight

[3] The Doppler shift, ω_{D_n}, is caused by the movement through the wave pattern. The wavelength is perceived as shorter (or longer) when traveling towards (or away from) a wave front.

towards the transceiver, is

$$\omega_D = \omega_c \frac{v}{c} = \frac{2\pi v}{\lambda_c} = kv. \tag{2.22}$$

If we introduce any form of acceleration or change of direction, i.e., driving in a curve, the Doppler shift will become time dependent.

When the Doppler shift is introduced in equation (2.20) the following time-frequency domain representation is obtained

$$H(\omega, t) = \sum_{n=1}^{N} \alpha_n e^{\jmath(-\omega\tau_n(0)+\omega_{D_n} t)}. \tag{2.23}$$

The time domain equivalent of this baseband description of the continuous-time channel is [9]

$$h(\tau, t) = \sum_{n=1}^{N} \alpha_n e^{\jmath\omega_{D_n} t} \delta(\tau - \tau_n(0)), \tag{2.24}$$

where $\delta(\tau)$ is Dirac's delta function and α_n is given by (2.19). The time-varying channel impulse response is thus described as the sum of N complex sinusoids, with fixed frequencies between plus/minus the maximum Doppler frequency ω_D.

It is useful to study the Doppler domain representation of equation (2.24), defined as

$$\begin{aligned} H(\tau, \Omega) &= \int_{-\infty}^{+\infty} \sum_n \alpha_n e^{\jmath\omega_{D_n} t} \delta(\tau - \tau_n(0)) e^{-\jmath\Omega t} dt \\ &= 2\pi \sum_n \alpha_n \delta(\tau - \tau_n(0)) \delta(\Omega - \omega_{D_n}), \end{aligned} \tag{2.25}$$

where Ω is the Doppler frequency (in rad/s). In the Doppler domain not only the different delays $\tau_n(0)$ for the incoming waves, as in the impulse response, but also their Doppler frequency ω_{D_n} can be studied. Using the Doppler domain representation (2.25), two rays that arrive with the same delay τ_n from different directions, as in Figure 2.1, can be separated by their different Doppler frequencies.

Let us recapitulate under what conditions the expressions (2.23)-(2.25) are valid.

- The channel models obtained in (2.23)-(2.25) are valid only locally, that is for a few wavelengths and corresponding short times (t) when moving the transceiver.

- Narrow relative bandwidth is presupposed. This holds for most mobile radio systems.
- Under the assumption that the scatterers are fixed and not too close to the receiver, the linearization (2.8) without the quadratic term, resulting in equation (2.23) is valid. This is the plane wave approximation.
- The velocity of the mobile is assumed constant, which excludes curves and accelerations.
- Furthermore, the models do not take polarization effects into consideration.

The use of a spherical wave propagation model is discussed next.

2.3 Sampled Channel with Time Varying Frequencies

The plane-wave approximation can be justified only at a large distance from the wave source. The assumption of spherical wave propagation may be more realistic in the following cases:

- A relatively close primary source visible via the direct path,
- A relatively close primary source visible via one or multiple reflections, modeled by the mirror-image of the primary source.
- A secondary source induced by a distant-source wave front impinging on a close point scatterer (such as a lamp-post).

Using a more elaborate model based on ray-optics including the last term in (2.8), we will see that in the presence of close scatterers the description of the channel as a sum of time invariant weighted complex sinusoids is an oversimplification. This may motivate the use of adaptive or nonlinear predictors.

The sampled channel: Several reflectors may contribute to each tap
We shall repeat and expand the discussion performed in the last section but now in the time domain and for sampled channels. Recall equations (2.2)

and (2.17) with $\omega_b = \omega$ in a base-band formulation. Let the position vector $\mathbf{r}$ depend on time. Thus we obtain

$$H(\omega, \mathbf{r}(t)) = \sum_{n=1}^{N} a_n(\mathbf{r}(t)) e^{\jmath(\psi_n(\mathbf{r}(t)) - (\omega_c + \omega) r_n(\mathbf{r}(t))/c)}. \tag{2.26}$$

The nth path delay is $\tau_n(t) = r_n(t)/c$. We can view the expression (2.26) as a function of time directly by dropping the position $\mathbf{r}(t)$ and just using t as variable. The time-varying impulse response $h(\tau, t)$ for a baseband channel at time t in a multi-path environment can thus be described by

$$h(\tau, t) = \sum_{n=1}^{N} a_n(t) e^{\jmath(\psi_n(t) - \omega_c \tau_n(t))} \delta(\tau - \tau_n(t)), \tag{2.27}$$

where ω_c is the carrier frequency (in rad/s) and $a_n(t)$ is a time-varying attenuation factor covering antenna effects, path loss and attenuation due to reflection and scattering for the nth path [9]. The phase shift caused by reflectors and scatterers is described by $\psi_n(t)$ whereas $\tau_n(t)$ denotes the propagation delay for the nth path. The expression (2.27) corresponds to (2.24) without the approximations due to linearization of the change of path distance and small relative bandwidth that render the Doppler frequency description of the phase.

Let $g(\cdot)$ be a time invariant impulse response due to pulse shaping and receiver filtering and let the symbol interval be T. The discrete-time channel impulse response can then be described by an FIR-filter with the mth tap given by [38]

$$\begin{aligned} h_m(t) &= \int_0^{MT} g(Tm - \tau) h(\tau, t) d\tau \\ &= \sum_{n=1}^{N} g(Tm - \tau_n(t)) a_n(t) e^{\jmath(\psi_n(t) - \omega_c \tau_n(t))}, \end{aligned} \tag{2.28}$$

where MT covers the length of the continuous-time impulse response. Note that N may be arbitrary large. In an ideal noiseless and lossless environment the radio waves could be reflected between objects forever, resulting in paths of unbounded length, requiring an IIR description of the channel. In practice we can however assume that the impulse response $h(\tau, t)$ will be of finite length, as the long paths are sufficiently attenuated, through propagation losses and losses at the reflecting/scattering surfaces, to fall below the background noise level. With an effective support of $g(\cdot)$ on the closed interval $[-KT, KT]$, the number of reflectors and scatterers contributing to the mth tap will be limited to paths with delays in the interval $[T(m-K), T(m+K)]$. A limited number of contributors is an advantage when the tap is to be predicted.

Effective source
If we base our modeling on ray optics and omit the effects of diffraction and Fresnel optics, a *scatterer* can be modeled as a *secondary effective source* induced by a wave front whereas a *reflector* generates a *secondary effective source* as the *mirror image* of the emitting source. Thus both scatterers and the mirror images can be viewed as secondary effective sources emitting *spherical wavefronts*. Even with this simplification we will encounter a model where the phases of the rotating channel coefficients are nonlinear functions of time. Such coefficients can not be predicted accurately by a linear predictor.

Path delays and phase
To simplify the expression of the phase in (2.28) we separate time-invariant and time depending factors. The path delay $\tau_n(t)$ can be decomposed into the sum of a time-varying delay from the effective source to the mobile ($\tau_n^{MS}(t)$) and a time-invariant path delay from the base station to produce the secondary effective source, τ_n^{BS} as

$$\tau_n(t) = \tau_n^{MS}(t) + \tau_n^{BS}. \tag{2.29}$$

For a scatterer τ_n^{BS} is the path delay from the base station to the scatterer and $\tau_n^{MS}(t)$ is the path delay from the scatterer to the mobile station. For a reflector the secondary source is the mirrored image of the primary source. The path delay from the secondary source to the mobile station is thus equal to the path delay from the primary source to the mobile station, that is $\tau_n^{MS}(t) = \tau_n(t)$. Accordingly is $\tau_n^{BS} = 0$ for reflections.

The geometry for the scatterers and reflectors are assumed to be time-invariant. The phase shift due to the time invariant delay, τ_n^{BS}, can be included in a complex attenuation factor which is now defined as

$$\alpha_{n,m}(t) = g(Tm - \tau_n(t))a_n(t)e^{\jmath(\psi_n(t) - \omega_c \tau_n^{BS})}, \tag{2.30}$$

while the time-varying term

$$\phi_n(t) \stackrel{\triangle}{=} -\omega_c \tau_n^{MS}(t) = -k r_n^{MS} = -k||\mathbf{r}_n^{MS}(t)||, \tag{2.31}$$

remains in the exponential factor of (2.28). Note that $a_n(t)$ and $\psi_n(t)$ and thus $\alpha_{n,m}(t)$ are assumed to be time-varying on a much slower time scale than $\phi_n(t)$. The discrete-time channel (2.28) can thus be expressed as

$$h_m(t) = \sum_n \alpha_{n,m}(t) e^{\jmath \phi_n(t)}. \tag{2.32}$$

In (2.31), $\mathbf{r}_n^{MS}(t)$ is a vector in space pointing from the nth effective source to the mobile station (see Figure 2.4). Consequently the norm of $\mathbf{r}_n^{MS}$ is nothing but the physical distance r_n^{MS}. For spherical waves the position vector $\mathbf{r}_n^{MS}$ is parallel to the wave vector, $\mathbf{k}_n(t)$. The phase $\phi_n(t)$ is thus solely a function of the electrical distance to the effective source. When the distance changes by as little as one wavelength the phase $\phi_n(t)$ changes by 2π, causing the effect of small-scale (fast) fading. In the model based on the plane wave-approximation (2.24) the phase $\phi_n(t)$ has only a linear dependence on time. In the following we will discuss how much the linearized model can deviate from a phase modeled using spherical waves.

Straight-line motion

Consider, as before, the simplest mobile dynamics, a straight-line motion at constant velocity $\mathbf{v}$. For notational convenience we denote the initial position by $\mathbf{r}_n^{MS}(0) = \mathbf{r}_n^{MS}$ without explicit time index. We then have the phase function

$$\phi_n(t) = -k||\mathbf{r}_n^{MS}(t)|| = -k||\mathbf{r}_n^{MS} + \mathbf{v}t||. \tag{2.33}$$

Let $\Omega_n(t)$ denote the angle between the wave vector $\mathbf{k}_n(t)$ (or equivalently the position vector $\mathbf{r}_n^{MS}(t)$) and the velocity vector $\mathbf{v}$, see Figure 2.4. The angle towards the effective source, also called the angle of incidence, is

$$\theta_n(t) = \pi - \Omega_n(t).$$

Furthermore let

$$\theta_n = \theta_n(0),\ r_n^{MS}(0) = ||\mathbf{r}_n^{MS}(0)||,\ v = ||\mathbf{v}||,\ r_n^{MS}(t) = ||\mathbf{r}_n^{MS}(t)||.$$

Then, by using the cosine theorem as in (2.7) the square of the path distance can be expressed as

$$r_n^{MS2}(t) = ||\mathbf{r}_n^{MS}(t)||^2 = r_n^{MS2} - 2r_n^{MS} vt\cos\theta_n + v^2t^2. \tag{2.34}$$

The phase function at the position $\mathbf{r}_n^{MS}(t)$ at time t can be rewritten as

$$\phi_n(t) = -kr_n^{MS}(t) = -kr_n^{MS}\sqrt{1 - 2\frac{vt}{r_n^{MS}}\cos\theta_n + \left(\frac{vt}{r_n^{MS}}\right)^2}. \tag{2.35}$$

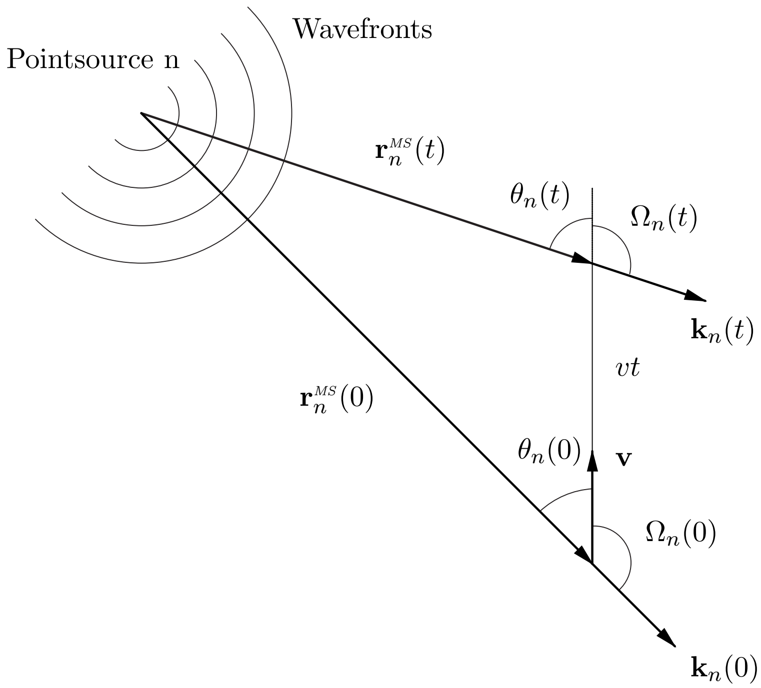

Figure 2.4: The change of path distance, $\mathbf{r}_n^{MS}(t)$, when moving with constant velocity, $\mathbf{v}$. Here $\Omega_n(t)$ is the angle between the velocity vector $\mathbf{v}$ and the wave vector $\mathbf{k}_n(t)$, whereas the angle $\theta_n(t) = \pi - \Omega_n(t)$ is the angle between $\mathbf{v}$ and the direction to the effective source at time t (that is $-\mathbf{r}_n^{MS}(t)$).

We now use a second-order approximation

$$\sqrt{1+y} \approx 1 + y/2 - y^2/8,$$

with y representing the sum of the two last terms under the square root sign of (2.35). Furthermore, we neglect terms in vt/r_n^{MS} of higher order than two since $vt/r_n^{MS} \ll 1$ and thus obtain

$$\begin{aligned} \phi_n(t) &\approx -kr_n^{MS}\left[1 - \frac{vt}{r_n^{MS}}\cos\theta_n + \frac{1}{2}\left(\frac{vt}{r_n^{MS}}\right)^2 \sin^2\theta_n\right] \\ &= -kr_n^{MS} + \omega_{D_n} t - \frac{kr_n^{MS}}{2}\left(\frac{vt}{r_n^{MS}}\right)^2 \sin^2\theta_n, \end{aligned} \tag{2.36}$$

with the Doppler frequency $\omega_{D_n} = kv\cos(\theta_n)$ as in (2.21). No terms of higher order than $(vt/r_n^{MS})^2$ are kept in the series expansion. The first term, the initial phase, is a time invariant phase shift and can be included in the complex attenuation factor $\alpha_{n,m}(t)$. The second term represents the phase rotation at the Doppler frequency, encountered in Section 2.2. In addition to this linear increase of the phase, we obtain a third, quadratic term originating from the sphericity of the wave fronts. The time-varying part of the phase in the channel model (2.32), can thus be approximated as

$$\phi_n(t) \approx \omega_{D_n} t - \frac{k r_n^{MS}}{2} \left(\frac{vt}{r_n^{MS}} \right)^2 \sin^2 \theta_n. \tag{2.37}$$

The quadratic time dependence of the phase can be interpreted as a *chirp*, whereas a strict linear time dependence would represent a complex sinusoid with time-invariant frequency. The first term in (2.37) is thus sufficient in the case of plane waves.

2.3.1 The instantaneous frequency

The instantaneous frequency, that is the time derivative of the phase function, can offer further insights into the relation between phase and Doppler frequency. The instantaneous frequency can be derived directly from (2.35) and will then be expressed in the initial angle θ_n, speed, and time. It can also be derived by operating on (2.33) to obtain an instantaneous frequency expressed in the instantaneous angle $\theta_n(t)$. In the latter approach the time-derivative of $r_n^{MS}(t)$ is needed. An infinitesimal step in time, Δt, result in a change of path distance as

$$r_n^{MS}(t + \Delta t) = ||\mathbf{r}_n^{MS}(t) + \mathbf{v}\Delta t||. \tag{2.38}$$

As Δt is infinitesimally small we can use the approximation (2.10) in (2.38),

$$r_n^{MS}(t + \Delta t) = r_n^{MS}(t) - v \cos \theta_n(t) \Delta t. \tag{2.39}$$

The time-derivative of the path-length is thus

$$\frac{dr_n^{MS}(t)}{dt} = \lim_{\Delta t \to 0} \frac{r_n^{MS}(t + \Delta t) - r_n^{MS}(t - \Delta t)}{2\Delta t} = -v \cos \theta_n(t). \tag{2.40}$$

The instantaneous frequency is thus given by

$$\begin{aligned} \dot{\phi}_n(t) &= -k \frac{d}{dt} ||\mathbf{r}_n^{MS}(t)|| = kv \cos \theta_n(t) \\ &= \omega_D \cos \theta_n(t) = \omega_{D_n}(t), \end{aligned} \tag{2.41}$$

where ω_D is the maximum Doppler frequency (2.22) and $\omega_{D_n}(t)$ is the instantaneous Doppler frequency for path n. If instead the approximate expression (2.36) is used in the derivation of the instantaneous frequency, then we obtain

$$\dot{\phi}_n(t) \approx k\left[v\cos\theta_n - \frac{t}{r_n^{MS}}\left(v\sin\theta_n\right)^2\right] = \omega_D\cos\theta_n - \omega_D\frac{tv}{r_n^{MS}}\sin^2\theta_n. \quad (2.42)$$

The instantaneous frequency can thus either be described as an instantaneous Doppler shift depending on the momentary angle of incidence $\theta_n(t)$ as in (2.41) or approximately by the difference between the Doppler shift at time $t = 0$ and a time dependent correction, as in (2.42).

Phase as a function of instantaneous angle of incidence

When comparing the expression (2.41) to (2.42), one can be led to assume that if the instantaneous Doppler frequency was used instead of the Doppler frequency corresponding to the angle at $t = 0$, in (2.36), then there would be no need for a correction term in (2.36). This is however not the case as we will see below. To express the distance $\mathbf{r}_n^{MS}(t)$ as a function of the instantaneous angle of incidence $\theta_n(t)$, we can by using the cosine theorem and the fact that $\cos\Omega_n(t) = \cos(\pi - \theta_n(t)) = -\cos\theta_n(t)$, obtain

$${r_n^{MS}}^2 = (vt)^2 + r_n^{MS}(t)^2 + 2vtr_n^{MS}(t)\cos\theta_n(t). \quad (2.43)$$

Solving for $r_n^{MS}(t)$ we obtain

$$r_n^{MS}(t) = -tv\cos\theta_n(t) + r_n^{MS}\sqrt{1 - \left(\frac{vt}{r_n^{MS}}\right)^2\sin^2\theta_n(t)}. \quad (2.44)$$

The phase function at the position $\mathbf{r}_n^{MS}(t)$ at time t is $-k\mathbf{r}_n^{MS}(t)$ and can thus be expressed as

$$\phi_n(t) \quad = \quad \omega_{D_n}(t)t - kr_n^{MS}\sqrt{1 - \left(\frac{vt}{r_n^{MS}}\right)^2\sin^2\theta_n(t)}, \quad (2.45)$$

or using the same approximation as in (2.36) as

$$\phi_n(t) \approx -kr_n^{MS} + \omega_{D_n}(t)t + \frac{kr_n^{MS}}{2}\left(\frac{vt}{r_n^{MS}}\right)^2\sin^2\theta_n(t). \quad (2.46)$$

The initial phase is obviously the same as in (2.36). In the second term the difference lies in that the instantaneous Doppler frequency $\omega_{D_n}(t)$ is used instead of the initial Doppler frequency ω_{D_n}. Observe that even though we use the instantaneous Doppler frequency there is a quadratic correction term in the phase, in this case involving $\sin^2\theta_n(t)$ instead of $\sin^2\theta_n$.

The sine of the instantaneous angle of incidence can however by application of the sine and cosine theorem, be expressed in terms of the time and the initial values of distance and angle, as

$$\sin^2\theta_n(t) \quad = \quad \frac{r_n^{MS\,2}\sin^2\theta_n}{r_n^{MS\,2}(t)} = \frac{r_n^{MS\,2}\sin^2\theta_n}{r_n^{MS\,2} + (vt)^2 - 2r_n^{MS}vt\cos\theta_n}. \tag{2.47}$$

Using the relation (2.47) we can show that the term $(\frac{vt}{r_n^{MS}})^2\sin^2\theta_n(t)$ occurring in the correction terms of (2.45) and (2.46) is bounded:

$$0 \leq \left(\frac{vt}{r_n^{MS}}\right)^2 \sin^2\theta_n(t) \leq 1. \tag{2.48}$$

Thus the correction terms in (2.45) and (2.46) are limited whereas the correction term in (2.36) grows continuously in magnitude. It is thus more convenient for modeling, to use a time varying Doppler frequency than to assume a fixed one, as in (2.36).

2.3.2 Linearized model

Assuming plane incoming waves, that is waves with a constant angle of incidence $\theta_n(t) = \theta_n$, and a time invariant complex attenuation factor $\alpha_{n,m}$ in (2.32), we obtain, by approximating $\phi_n(t)$ in (2.36) by $-kr_n^{MS} + \omega_{D_n}t$ and including the phase factor $-kr_n^{MS}$ in $\alpha_{n,m}$, the following commonly used approximation [11]

$$h_m(t) = \sum_{n=1}^{N} \alpha_{n,m} e^{\jmath\omega_D \cos\theta_n t} = \sum_{n=1}^{N} \alpha_{n,m} e^{\jmath\omega_{D_n} t}. \tag{2.49}$$

This is the sampled version of the continuous impulse response (2.24). This model has been used as a basis expansion model for blind equalization [39] and also for long-range prediction of mobile radio channels by Duel-Hallen and co-workers [20].

As we have seen in (2.46) the Doppler frequencies actually vary, and therefore a more suitable model is

$$h_m(t) = \sum_{n=1}^{N} \alpha_{n,m}(t) e^{\jmath\omega_{D_n}(t)t}, \tag{2.50}$$

where $\omega_{D_n}(t)$ is the instantaneous Doppler frequency for the nth path and $\alpha_{n,m}(t)$ is the instantaneous complex attenuation. Both these parameters can be assumed to vary slowly.

Phase error in sinusoid models
The linear phase model, that is the deterministic sinusoid model (2.49), will accumulate a phase error when used in a situation with nearby scatterers where the plane wave approximation does not hold. The size of this phase error defines the maximum time interval over which the linear deterministic model can be used as an approximation in the spherical-wave case. The difference in phase can be approximated by the second order term in (2.36)

$$\Delta\phi_n(t) = -\frac{kv^2}{2r_n^{MS}} t^2 \sin^2\theta_n = -\frac{\omega_D v}{2r_n^{MS}} t^2 \sin^2\theta_n. \tag{2.51}$$

The largest phase deviation in the model (2.49) relative to (2.35) occurs for a *transversal vehicle movement* at $\theta_n = \pi/2$, when the velocity vector is orthogonal to the direction of the incident wave at time $t = 0$ and the nominal Doppler shift vanishes.

Example 2.1 When is the linear model inadequate?

Consider a situation where there is only one path contributing to the channel and the velocity vector is orthogonal to the direction of the incident waves at time $t = 0$. The contribution by the path to the channel tap is a chirp $h(t) = \alpha e^{\jmath\zeta t^2}$, where the phase is obtained from (2.36). The phase factor $-kr_{MS}$ is part of α, $\omega_{D_n} t = 0$ since ω_{D_n} is zero as the waves arrive perpendicular to the direction of motion, and the third of term (2.36) yields ζt^2 with $\zeta = -kv^2/(2r_{MS})$ since $\theta = \pi/2$. The true parameters, α and ζ, at time $t = 0$ are now used either in the linear sinusoid model (2.49) or in the model (2.32), with the quadratic phase expression (2.35), and the time is propagated. Two different predictions of $h(t)$ into the future are then obtained. The phase of $h(t)$ at time $t = T$ is ζT^2, and this will also be the result in the model containing the quadratic term. The linear phase model, with the accumulated phase error of $\Delta\phi(T)$, will predict the contribution to be

$$\hat{h}(t) = \alpha e^0 = \alpha e^{\jmath(\zeta T^2 + \Delta\phi(T))}. \tag{2.52}$$

The channel tap prediction error, for the predictor using the linear model, at time $t = T$ is then

$$\begin{aligned} \varepsilon(T) &= \alpha(e^{\jmath\zeta T^2} - e^{\jmath(\zeta T^2+\Delta\phi(T))}) = \alpha e^{\jmath(\zeta T^2+\Delta\phi(T)/2)}(-2\jmath)\sin\left(\frac{\Delta\phi(T)}{2}\right) \\ &= \alpha e^{\jmath(\zeta T^2-\Delta\phi(T)/2)}(-2\jmath)\sin\left(\frac{\Delta\phi(T)}{2}\right) \end{aligned} \tag{2.53}$$

and the relative power of the error is

$$\frac{|\varepsilon(T)|^2}{|\alpha|^2} = 4\sin^2\frac{\Delta\phi(T)}{2}. \tag{2.54}$$

Compare this to the result if we would predict the tap by its mean, $\hat{h}(T) = 0$, which would give a relative error of one. The time limit for how long the linear model can be used as a predictor, with better performance than the mean, is obtained by

$$4\sin^2\frac{\Delta\phi(T)}{2} \leq 1, \tag{2.55}$$

that is

$$|\Delta\phi(T)| \leq \pi/3. \tag{2.56}$$

Thus, predictions based on the extrapolation of a sinusoidal model will perform better than the zero predictor only up to prediction horizon T where the accumulated phase error has grown to $\pi/3$.

Based on the above example for a situation with just one path with a tap acting as a chirp, the time interval $T_{\pi/3}$ denotes the time over which a linear predictor based on the model (2.49) actually renders predictions better than using the mean as the prediction. Defining this time interval, $T_{\pi/3}$, for the worst case $\theta_n = \pi/2$, as the interval after which the phase approximation error has grown to $60°$, $|\ \Delta\phi_n(t)|_{T_{\pi/3}}\ | = \pi/3$, we obtain from (2.51)

$$T_{\pi/3} = \sqrt{\frac{\lambda r_n^{MS}}{3v^2}}. \tag{2.57}$$

For this time span, the above second-order approximation (2.36) can be well justified as we have

$$\left.\frac{vt}{r_n^{MS}}\right|_{t=T_{\pi/3}} = \sqrt{\frac{\lambda}{3r_n^{MS}}} \ll 1 \tag{2.58}$$

where the last inequality holds because even a close point scatterer will be many wavelengths away from the mobile transceiver in outdoor mobile radio scenarios. There is thus no need for more terms in the series expansion of the phase in (2.36).

The time interval $T_{\pi/3}$ can be used to determine if the linear model (2.49) is adequate over a certain time interval. In environments where $T_{\pi/3}$ is large, there is no need for the quadratic term in the phase. Whenever $T_{\pi/3}$ is small, either a time-varying model or a phase with a quadratic term is called for.

Example 2.2

A simple scenario including a vehicle driving at 90 km/h past a close point scatterer, 10 m beside the road, demonstrates the limitations of the linear model. With a carrier frequency of 1800 MHz the resulting $T_{\pi/3}$ is as short as 30 ms (corresponding to a traveled distance of 4.5 wavelengths). A linear deterministic model will thus not be valid for more than a few wavelengths in this scenario. An attempt to estimate such a model for a longer data interval (estimation window) would result in an average linear model for the estimation interval, with poor properties for extrapolation outside the interval.

However, this problem does not exist for more distant (primary and reflected) sources, say, at $r_n^{MS} = 250$ m and beyond where we obtain $T_{\pi/3} =$ 150 ms, which is more than a distance of 22 wavelengths. The linear model will then be a good approximation of the true dynamics of the channel for a longer distance, than in the previous case.

2.3.3 Path loss

The path loss for a scattered or reflected path is proportional to the path distance as [13]

$$a_n(t) \propto (r_n^{MS}(t) \otimes r_n^{BS}(t))^{-\gamma/2} \tag{2.59}$$

where $r_n^{MS}(t)$, $r_n^{BS}(t)$ denotes the distance from the nth scatterer/reflector to the mobile station and to the base station respectively and γ is the power attenuation exponent. In (2.59) $r_n^{MS}(t) \otimes r_n^{BS}(t)$ denotes $r_n^{MS}(t) \cdot r_n^{BS}(t)$ for scattering and $r_n^{MS}(t) + r_n^{BS}(t)$ for specular reflection. The power attenuation exponent corresponding to the path loss in free space is $\gamma = 2$. Okamura [40] suggest that the power attenuation exponent should be $\gamma = 2 + 2m$ where

m depends on the distance between the base station and the mobile station. Up to 10 km $m = 0.5$ gives a good approximation for the power attenuation observed in measurements. For close scatterers the power attenuation exponent corresponding to the path loss in free space, $\gamma = 2$, can be used. Even fairly close scatterers will give contributions with amplitudes which are an order of magnitude weaker than the specular reflections. Thus, it is appropriate to use the linear model (2.49) in situations where there are direct line of sight (LOS) or large contributions from specular reflections, that is strong reflections from large buildings. In all other cases the effect of spherical waves from nearby scatterers introduces a significant deviation from the linear model (2.49).

2.3.4 Curves and plane waves

Phase functions containing quadratic terms occur when the vehicle accelerates or makes a turn, even if the plane-wave approximation holds. Assume that the mobile moves at constant tangential speed v around a circle of radius R with its center at the origin of the coordinate system such that

$$||\mathbf{r}(t)|| = R, \tag{2.60}$$

where $\mathbf{r}(t)$ denotes the position of mobile station. As the mobile station now changes direction we use a fixed coordinate system as shown in Figure 2.5. Let β_n be the angle between the constant wave vector $\mathbf{k}_n$ (plane wave) and the initial position vector $\mathbf{r}(0)$ as in Figure 2.5. Recall that the phase could be calculated as the scalar product between the wave vector and the position vector added to a time invariant phase factor as for the continuous channel (2.5). We take $\phi_n(t)$ to be the time varying part of the phase in (2.31). If we initiate $\phi(t)$ to be zero when the position vector and the wave vector are orthogonal, then the phase function at time t is given by

$$\begin{aligned} \phi_n(t) &= -\mathbf{k}_n \cdot \mathbf{r}(t) \\ &= -k[\cos\beta_n, \sin\beta_n, 0] \cdot R[\cos(vt/R), sin(vt/R), 0] \\ &= -kR\cos(vt/R - \beta_n). \end{aligned} \tag{2.61}$$

The phase does not increase as a linear function of time here. Instead the Doppler shift oscillates (slowly and) symmetrically around 0. The maximum deviation from a linear phase function occurs at the time instances when the cosine function has its maximum curvature, which coincide with the cosine maxima. These are reached when the vehicle moves *transversal to the incident wave*, that is when $vt/R = \beta_n$, where the instantaneous Doppler

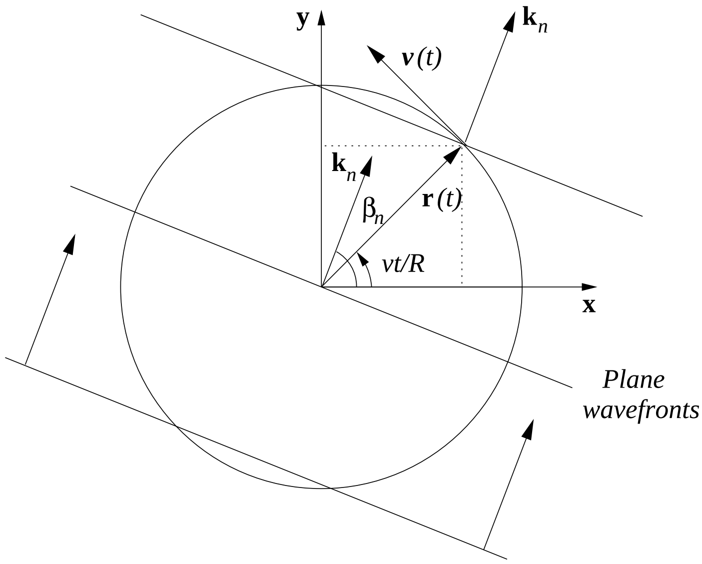

Figure 2.5: Plane waves and a vehicle driving with velocity v in a curve with radius $||\mathbf{r}(t)|| = R$.

shift vanishes but the corresponding chirp rate (according to (2.36)) has its maximum. Consider the time interval $T_{\pi/3}$ over which a phase error of $\pi/3$ is accumulated. To this end, a second-order approximation of the cosine function is used around its maximum, i.e.

$$\phi_n(t) \approx -kR\left[1 - \frac{1}{2}(vt/R)^2\right], \tag{2.62}$$

where t is relative to the time where the cosine maximum is reached during the circular motion. Again we have obtained a second-order polynomial phase behavior. The constant term corresponds to the linear deterministic model which, for transversal motion, would not show any Doppler shift at all. The quadratic term represents the deviation from the linear behavior. Using $T_{\pi/3}$ as the measure for how fast these deviations occur we obtain

$$T_{\pi/3} = \sqrt{\frac{\lambda R}{3v^2}}. \tag{2.63}$$

For this time span, the above second-order approximation can be well justified as we have

$$\left.\frac{v\tau}{R}\right|_{\tau=T_{\pi/3}} = \sqrt{\frac{\lambda}{3R}} \ll 1 \tag{2.64}$$

where the last inequality holds because the minimal turning circle diameter of any car is orders of magnitudes larger than the radio wavelengths used in digital mobile communications systems.

Example 2.3 Is the linear model adequate for driving in curves?

We evaluate equation (2.63) for some typical speeds and sizes of curves. With a carrier frequency of 1800MHz, a speed of v =90 km/h and a curve radius of R =100 m we have $T_{\pi/3}$ =94 ms (2.4 m or 14 wavelength of traveled distance) and, for v =20 km/h and R =10 m, $T_{\pi/3}$=130 ms (0.75 m or 4.5 wavelengths of traveled distance). We conclude that normal deviations from the exact straight-line motion may result in significant phase errors of the linear deterministic fading model over estimation time windows of the order of 0.1 s.

2.4 Statistical Channel Model

The smallest objects that radio-waves interact with are on the order of a wavelength long. Facades of old houses in any European town consist not so much of flat surfaces as sculptured windows and decorations, thus a lot of object with shapes on the scale of a wavelength, about 15 cm, for a 2GHz radio interface. Such a facade will not act as a reflector but as a cluster of scatterers. In the previous sections we have treated the effect of discrete point source reflectors and scatters. Here we will instead study clusters of objects and their effect on the channel.

In antenna array measurements of the angular distribution of power [41] it is observed that most reflecting objects indeed have an angular spread. The object thus does not act as a perfect reflector. Rather it act more like a cluster of scatterers.

A distant group of buildings which act as reflectors can also constitute a cluster. Each building then contributes a plane wave to the energy received by the transmitter, but from a slightly different direction and with a different phase as compared to the contribution from adjacent other houses. In

Section 2.2 we saw that, due to the movement of the transceiver, a plane wave results in a channel that can be approximated by a complex sinusoid. The channel discussed here will then consist of a large number of complex sinusoids with different phases but with almost the same frequency. When the number of sinusoids is high enough, the law of large numbers causes the sum of sinusoids to result in a complex Gaussian channel and the contribution to the channel from the cluster of buildings can be approximated as a narrow-band filtered Gaussian noise.

The optimal predictor for a signal being a linearly filtered Gaussian noise, which is disturbed by an additive Gaussian noise, is linear. To study the linear prediction properties for scenarios with angular spread, we need to describe the correlation and Doppler spectrum of the channel.

In the following sections we will recapitulate some results from angular power distributions, link these results to geometrical properties of clusters of scatterers and finally motivate the choice of simple linear AR and ARMA models for the dynamics of the channel.

2.4.1 Correlation

The point source reflector and scatterer model lead to the linearized model (2.49) where each tap of the channel is modeled as the sum of complex sinusoids caused by plane waves,

$$h(t) = \sum_n \alpha_n \mathrm{e}^{\jmath \omega_D \cos\theta_n t}. \tag{2.65}$$

The linear prediction properties depend on the the auto-covariance function for the tap, which is

$$r_h(\tau) = \mathrm{E}\{h(t)h^*(t-\tau)\} = \mathrm{E}\left\{ \sum_k \sum_l \alpha_k \alpha_l^* \mathrm{e}^{\jmath \omega_D (\cos\theta_k - \cos\theta_l)t} \mathrm{e}^{\jmath \omega_D \cos\theta_l \tau} \right\}, \tag{2.66}$$

where we take the average over time of (2.66) to obtain

$$r_h(\tau) = \sum_n |\alpha_n|^2 \mathrm{e}^{\jmath \omega_D \cos\theta_n \tau}. \tag{2.67}$$

It is the magnitude and direction of arrival of the waves that matters for the correlation.

Instead of point source scatterers, resulting in discrete angles of arrival, we shall here, according to the discussion above, use a continuous distribution over the angle of arrival. In other words, we assume a diffuse scattering processes. The sum in (2.67) is then exchanged for an integration over the angle of arrival [9]. The normalized auto-covariance function for the channel can then be expressed as

$$r_h(\tau) = \int_\gamma f_\gamma(\gamma) e^{\jmath\omega_D \cos\gamma\tau} d\gamma, \tag{2.68}$$

where γ is the angle between the direction of movement and the direction to the contributing scatterer, as defined in Figure 2.6, and $f_\gamma(\gamma)$ is the angular probability density function. If the antenna is not omni-directional then the characteristics of the antenna gain has to be included into $f_\gamma(\gamma)$. It is

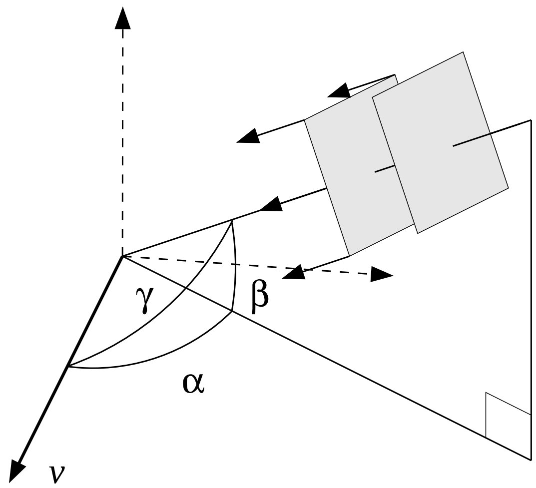

Figure 2.6: The angle γ is the angle between the direction of motion and the direction to the source of the incoming waves. (Note that it can be a secondary source.) The angle in the azimuthal plane is α and the elevation angle is β

the factor $\cos\gamma$ that is the important property and as it is a function of a stochastic variable with known distribution we can obtain the PDF for $\cos\gamma$ by transformation of $f_\gamma(\gamma)$ [42]. We begin the transformation by finding the solution to the inverse function of $\cos\gamma$,

$$c = g(\gamma) = \cos\gamma \tag{2.69}$$

$$\gamma = \pm|\arccos c|. \tag{2.70}$$

There are thus two solutions to the inverse of $g(\gamma)$. To obtain the PDF for

c we also need the derivative of $g(\gamma)$,

$$g'(\gamma) = \frac{dc}{d\gamma} = -\sin\gamma = -\sqrt{1-c^2}. \tag{2.71}$$

Now the corresponding PDF for c, with $\gamma = |\arccos c|$, can be formed as

$$f_c(c) = \frac{f_\gamma(\gamma)}{|g'(\gamma)|} + \frac{f_\gamma(-\gamma)}{|g'(-\gamma)|} = \begin{cases} \frac{f_\gamma(|\arccos c|)+f_\gamma(-|\arccos c|)}{\sqrt{1-c^2}} & -1 < c \leq 1 \\ 0 & \textit{otherwise.} \end{cases} \tag{2.72}$$

The correlation (2.68) for the channel is then obtained as

$$r_h(\tau) = \int_{-1}^{1} f_c(c)\mathrm{e}^{\jmath\omega_D c\tau} dc. \tag{2.73}$$

An important property of the channel is the Doppler spectrum, which can be obtained as the Fourier transform of the auto-correlation of the channel,

$$R_h(\omega) = \int_{-\infty}^{\infty} r_h(\tau)\mathrm{e}^{-\jmath\omega\tau} d\tau = \int_{-1}^{1} f_c(c) \underbrace{\int_{-\infty}^{\infty} \mathrm{e}^{\jmath(\omega_D c-\omega)\tau} d\tau}_{2\pi\delta(\omega-\omega_D c)} dc \tag{2.74}$$

$$= \frac{2\pi}{\omega_D}\int_{-\omega_D}^{\omega_D} f_c\left(\frac{u}{\omega_D}\right)\delta(\omega-u)du = \frac{2\pi}{\omega_D} f_c\left(\frac{\omega}{\omega_D}\right), \tag{2.75}$$

where we in the second last equality have made the change of variables $u = \omega_D c$ and in the last equality we have exploited the properties of the Dirac delta function. The Doppler spectrum thus directly depend on the distribution for c and can, utilizing (2.72), be expressed as

$$R_h(\omega) = \begin{cases} 2\pi\frac{f_\gamma(\arccos\frac{\omega}{\omega_D})+f_\gamma(-\arccos\frac{\omega}{\omega_D})}{\sqrt{\omega_D^2-\omega^2}} & |\omega| \leq \omega_D \\ 0 & \textit{otherwise,} \end{cases} \tag{2.76}$$

which is known as Gans mapping [43]. *The Doppler spectrum is thus band-limited.* This is a result which is valid locally but *as soon as the time variation of the angular power distribution is taken into account, the spectrum is no longer band-limited.*

There is thus a direct relation between the angular distribution of power, which is determined by the geometry of the scattering environment, and the correlation of the channel. A remaining problem is to obtain reasonable distributions for γ. To get some ideas of that we will consider some special cases in the following.

2.4.2 The flat world

A common simplification of the problem with the angular distribution of power is to assume that all scattering processes are in the azimuthal plane. Then $\beta = 0$ and $\alpha = \gamma$ in Figure 2.6 and the covariance (2.68) becomes an integral from $-\pi$ to π over the angles in the plane,

$$r_h(\tau) = \int_\alpha f_\alpha(\alpha) e^{\jmath\omega_D \cos\alpha\tau} d\alpha. \tag{2.77}$$

The angular distribution $f_\alpha(\alpha)$ simply states from which direction a certain amount of power comes, in the horizontal plane. Just as for the spatial angle γ we can form the PDF for $a = \cos\alpha$,

$$f_a(a) = \frac{f_\alpha(\alpha) + f_\alpha(-\alpha)}{\sqrt{1-a^2}} = \begin{cases} \frac{f_\alpha(|\arccos a|)+f_\alpha(-|\arccos a|)}{\sqrt{1-a^2}} & -1 < a \leq 1 \\ 0 & \textit{otherwise}, \end{cases} \tag{2.78}$$

so the Doppler spectrum can then obtained as in (2.75)

$$R_h(\omega) = \frac{2\pi}{\omega_D} f_a\left(\frac{\omega}{\omega_D}\right). \tag{2.79}$$

The mapping of equation (2.79) can be understood in geometrical terms, just as with the point-source scattering model discussed earlier. An incident wave with the angle of arrival α to the direction of motion of the moving transceiver is perceived as compressed by a factor proportional to the speed and $\cos\alpha$ (the projection of a unit distance on the direction given by the angle α). The distribution $f_a(a)$ can be interpreted as the normalized Doppler spectrum measured in Hz instead of rad/s, with the maximum Doppler frequency set to 1Hz.

Omni-directional scattering

In a rich scattering environment waves from all directions are equally likely and the angular PDF is

$$f_\alpha(\alpha) = \frac{1}{2\pi}. \tag{2.80}$$

The corresponding Doppler spectrum, obtained by inserting (2.80) into (2.78) and (2.79), is

$$R_h(\omega) = \begin{cases} \frac{2}{\sqrt{\omega_D^2-\omega^2}} & |\omega| \leq \omega_D \\ 0 & \textit{otherwise}, \end{cases} \tag{2.81}$$

which is the Jakes spectrum [11]. The corresponding correlation function is

$$r_h(\tau) = \mathrm{J}_0(\omega\tau) \tag{2.82}$$

where $\mathrm{J}_0(\cdot)$ is the zero order Bessel-function of the first kind.

Furthermore, all angles of incidence are assumed to be time-invariant and equally probable. That is, the angles θ_n are mutually uncorrelated with uniform distribution in the interval $[0, 2\pi[$. The Doppler frequency is given as $\omega_D \cos\theta_n$. As the probability distribution for $\cos\theta_n$ has high peaks at ± 1, the Doppler frequencies ω_{D_n} are likely to be close to the limits $\pm\omega_D$. This causes the well known bathtub shape of the Doppler spectrum for a

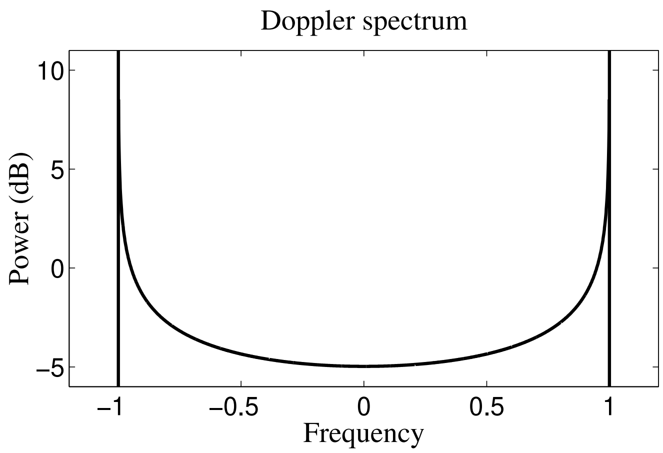

Figure 2.7: Doppler spectrum for a Rayleigh fading tap. The frequency is normalized by the Doppler frequency ω_D and the power is normalized by the power of the tap. There is no power outside the the bounds set by the Doppler frequency (± 1).

Rayleigh fading tap shown in Figure 2.7. This theoretical Doppler spectrum for a Rayleigh fading tap is a good approximation for taps in a narrow-band channel, when the assumptions of a high number of reflectors are met. For broad-band channels there will be more taps in the the discrete-time baseband channel resulting in higher spatial resolution. As a result, fewer reflectors will contribute to each tap. This cause the Doppler spectra for the taps in broad-band channels to have much more fine grained structure, as illustrated in Appendix A. Still, the Jakes model will be used as a benchmark channel through out the thesis.

2.4.3 Elevation

Radio waves that are diffracted over roofs down into a street will not reach the transceiver in the horizontal plane, neither will reflected waves from scatterers on high buildings in the vicinity. The received power might thus have a distribution in elevation too.

The PDF for the elevation angle power distribution is $f_\beta(\beta)$ where the elevation angle β is between $-\pi/2$ and $\pi/2$. The PDF for $b = \cos\beta$ can just as for γ be obtained as

$$f_b(b) = \frac{f_\beta(\beta) + f_\beta(-\beta)}{\sqrt{1-b^2}} = \begin{cases} \frac{f_\beta(|\arccos b|)+f_\beta(-|\arccos b|)}{\sqrt{1-b^2}} & 0 < b \leq 1 \\ 0 & otherwise, \end{cases} \tag{2.83}$$

Note that the borders are different to (2.72) due to the different region of valid angles.

A simple trigonometric calculation renders that $\cos\gamma = \cos\alpha\cos\beta$. Thus, if the azimuthal and elevation angles are independent, then the correlation can be obtained as

$$\begin{aligned} r_h(\tau) = \int_c f_c(c) e^{\jmath\omega_D c\tau} dc &= \int_\gamma f_\gamma(\gamma) e^{\jmath\omega_D \cos\gamma\tau} d\gamma \\ &= \int_\alpha\int_\beta f_\alpha(\alpha) f_\beta(\beta) e^{\jmath\omega_D \cos\alpha\cos\beta\tau} d\alpha d\beta. \end{aligned} \tag{2.84}$$

As $\cos\gamma = \cos\alpha\cos\beta$, the stochastic variable c is the product between $a = \cos\alpha$ and $b = \cos\beta$. The PDF for the product of two stochastic variables can be obtained as [42](pp. 147)

$$f_c(c) = \int_{-1}^{1} \frac{f_a(a)}{|a|} f_b\left(\frac{c}{a}\right) da = \int_0^1 \frac{f_b(b)}{b} f_a\left(\frac{c}{b}\right) db \tag{2.85}$$

From (2.75) we have a direct relationship between $f_c(c)$ and the Doppler spectrum. Thus, if we have both an azimuthal and an elevation angular distribution, then the Doppler spectrum can easily be obtained with the aid of the integral (2.85). There have been a few suggestions for the distribution of elevation angle, e.g. [44] and [45].

Example 2.4

Aulin [44] proposed the elevation angular distribution

$$f_\beta(\beta) = \begin{cases} \frac{\cos\beta}{2\sin\beta_0} & |\beta| \leq \beta_0 < \frac{\pi}{2} \\ 0 & otherwise. \end{cases} \tag{2.86}$$

The parameter β_0 is the maximum elevation angle, which corresponds to the angle under which the highest building is observed. By inserting (2.86) into (2.83) we obtain

$$f_b(b) = \frac{1}{\sin\beta_0}\frac{b}{\sqrt{1-b^2}} = \frac{1}{\sqrt{1-b_0^2}}\frac{b}{\sqrt{1-b^2}} \tag{2.87}$$

where $b_0 = \cos\beta_0$, which imposes $0 < b_0 \leq 1$. The PDF for c is then

$$f_c(c) = \frac{1}{\sqrt{1-b_0^2}}\int_{b_0}^{1}\frac{1}{\sqrt{1-b^2}} f_a\left(\frac{c}{b}\right) db \tag{2.88}$$

The main advantage of this distribution is that it offers an analytical solution for the Doppler spectrum in the case of omni-directional scattering in the azimuthal plane [44], that is $f_\alpha(\alpha) = 1/2\pi$,

$$R_h(\omega) = \begin{cases} \frac{\pi}{2\omega_D \sin\beta_0} & \omega_D\cos\beta_0 \leq |\omega| \leq \omega_D \\ \frac{2\pi}{\omega_D}\left(\frac{\pi}{2} - \arcsin\frac{2\cos^2\beta_0 - 1 - (\omega/\omega_D)^2}{1-(\omega/\omega_D)^2}\right) & |\omega| < \omega_D\cos\beta_0 \\ 0 & |\omega| > \omega_D \end{cases} \tag{2.89}$$

An important property of the Doppler spectrum (2.89) in Example 2.4 is that it remains finite everywhere, as opposed to the Jakes spectrum. *The spread in elevation angle of the incident waves smears out the distinct peaks of the Jakes Doppler spectrum, due to the integral over the elevation distribution.* Peaks will still be present, but not as high and narrow as in the Jakes spectrum. Most reasonable elevation distributions result in this effect that change the shape of all Doppler spectra in the same manner as the Jakes spectrum.

2.4.4 Spatial and angular distribution

It is the scattering environment that results in the angular distribution of the power. There exists a mapping from the spatial distribution of the scatterers to the angular distribution. The simplest of these mappings has already been treated, namely the flat distribution of scatterers in all directions. In this section we assume the single scattering model (SSM). Then, only single scattering processes contribute to the power as multiple scattering events are assumed to be too weak to contribute. The power will thus propagate from a scatterer directly to the receiver without interaction with the other scatterers in the cluster. A cluster is just a geometrically distributed group of point scatterers. The scenario in Figure 2.8 is a typical mobile radio channel model. There are distant spatially distributed clusters with different path delays, that each contribute to the channel. Usually a ring with local scatterers is also included in the model of the channel [13]. We will in the following only consider distant clusters.

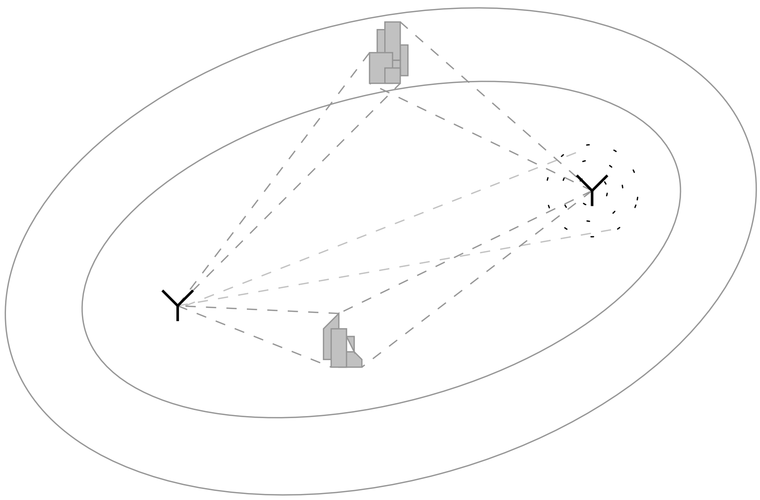

Figure 2.8: The two distant clusters of scatterers (group of buildings), seen in different directions and under different angles, will contribute to different taps of the channel. There is a ring of local scatterers around the transceiver to the right. The scattering in the local environment will give a contribution to all taps. In the SCM model it is assumed that the power after two scatterng events can be neglected, then the local scattering will contribute only to the first tap.

In the following two examples, we introduce a one-dimensional and a two-dimensional distribution of scatters, respectively, and their angular distribution. The results are obtained for a flat world, but can be generalized to include elevation using (2.85). For distant clusters it can be assumed that the azimuthal extension will be much larger than the extension in elevation (a group of buildings seen from a distance is generally quite wide but not so high). In that case the flat world approximation is quite usefull.

In Section 2.2 we defined the angle of incidence θ as the angle between the direction of movement and the direction of the scatterer. This can be generalized so that the angle of incidence θ denotes the angle to the center of the flat cluster. The correlation and the Doppler spectrum then depends on $a = \cos(\theta+\alpha)$, since the cluster has θ as angle of incidence. This imposes a slight change of the mapping (2.78) to

$$f_a(a) = \frac{f_\alpha(|\arccos a| - \theta) + f_\alpha(-|\arccos a| - \theta)}{\sqrt{1-a^2}}[\mathrm{U}(a+1) - \mathrm{U}(a-1)], \tag{2.90}$$

where we assume that $f_\alpha(\alpha)$, as defined in the interval $[-\pi\ \pi]$, is repeated with a period of 2π. This assumption is used throughout this Chapter and makes the handling of the limits much easier.

Example 2.5 Scatterers spread on a line

Here the scatterers are assumed to be spread on a line along a wall, following a Gaussian distribution,

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \mathrm{e}^{-\frac{x^2}{2\sigma_x^2}}, \tag{2.91}$$

where x denotes the coordinate along the wall. The wall is on the distance L from the transceiver and oriented as in Figure 2.9, with the wall perpendicular to the direction of the antenna from the center of the cluster. Let the attenuation depend only on the distance from the antenna to the

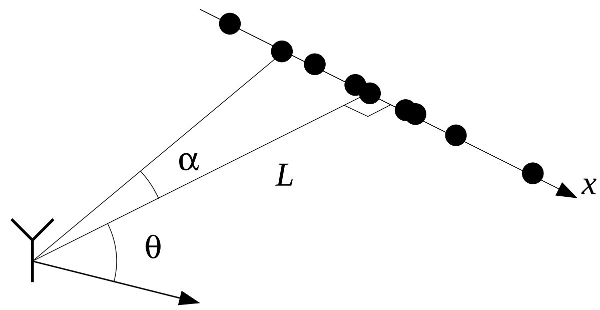

Figure 2.9: A scatter distribution along a wall.

center of the cluster L and not on the distance to the individual scatterers. This is a valid approximation if $\sigma_x \ll L$. Furthermore assume that all the scatterers reflect the same power and that the phases are random. The angle α depends on the scatterer position through $\tan\alpha = x/L$ and is thus a function of a random variable. The PDF for the angular distribution can then be obtained by a transformation of the Gaussian PDF as

$$f_\alpha(\alpha) = \begin{cases} \frac{L}{\sqrt{2\pi\sigma_x^2}\cos^2\alpha} e^{-\frac{L^2}{2\sigma_x^2}\tan^2\alpha} & -\pi/2 < \alpha \leq \pi/2 \\ 0 & otherwise, \end{cases} \tag{2.92}$$

For $\sigma_x^2/L \ll 1$ only small angles will contribute to the power and we can approximate $\cos^2\alpha \approx 1$ and $\tan^2\alpha \approx \alpha^2$. The angular distribution for a small or distant wall of Gaussian distributed scatterers can thus be approximated as

$$f_\alpha(\alpha) \approx \frac{L}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{L^2}{2\sigma_x^2}\alpha^2}, \tag{2.93}$$

which is a Gaussian distribution with variance σ_x^2/L^2. When L increases or σ^2 decreases, the cluster is seen under a decreasing angle. In the limit $\sigma_x^2/L^2 \to 0$, the cluster thus resembles a point source.

Example 2.6 A Gaussian two dimensional cluster

Here we consider a two dimensional cluster of scatterers, as in Figure 2.10, with normal distribution of the x and y coordinates,

$$f_{xy}(x, y) = \frac{1}{2\pi\sigma_r^2} e^{-\frac{x^2+y^2}{2\sigma_r^2}}. \tag{2.94}$$

The standard deviation σ_r of the spatial distribution of scatterers can be seen as an effective radius of the cluster. Let the attenuation depend only on the distance from the antenna to the center of the cluster L and not on the distance to the individual scatterers. This assumption is well fulfilled when $\sigma_r^2 \ll L$. The angular distribution is then obtained as [46]

$$f_\alpha(\alpha) = \frac{1}{2\pi} e^{-\frac{L^2}{2\sigma_r^2}} \left(1 + \mathrm{erf}\left(\frac{L\cos\alpha}{\sqrt{2\sigma_r^2}}\right)\right)\left(1 + \sqrt{\frac{\pi L^2}{2\sigma_r^2}} \cos\alpha e^{\frac{L^2}{2\sigma_r^2}\cos^2\alpha}\right), \tag{2.95}$$

which is equivalent to the mathematical expression Rice [47] obtained for finding the phase distribution of a sine wave in additive white Gaussian noise. the function $\mathrm{erf}(\cdot)$ is the error function for a Gaussian distribution. Note that if we set $L = 0$ in (2.95), that is we stand in the middle of the

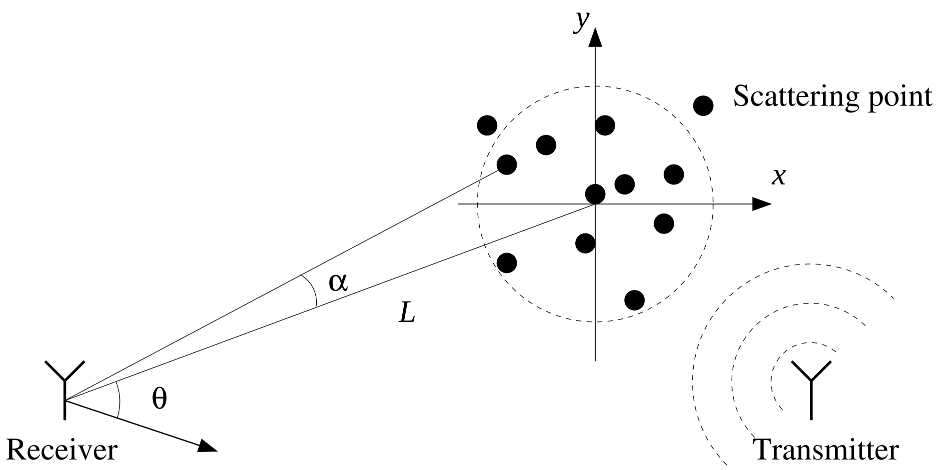

Figure 2.10: A two dimensional Gaussian cluster of scatterers.

cluster, we obtain $f_\alpha(\alpha) = 1/2\pi$, as then all directions are equally likely. If $\sigma_r^2 \lll L$ then (2.95) can be approximated as

$$f_\alpha(\alpha) \approx \frac{L}{\sqrt{2\pi\sigma_r^2}} \mathrm{e}^{-\frac{L^2}{2\sigma_r^2}\alpha^2}. \tag{2.96}$$

Thus, for small or distant clusters of scatterers with spatially Gaussian distribution, the azimuthal angular distribution is approximately Gaussian with variance σ_r^2/L^2. The angular distribution does thus mainly depend on the one dimensional projection of the cluster, as in Example 2.5.

Both the one dimensional and the two dimensional clusters result in the same azimuthal angular distribution when they are seen under a sufficiently small angle. In both cases the mapping from spatial to angular distribution depends on $\tan\alpha \approx \alpha$ for small angles. *The angular distribution will thus be similar to the one dimensional projection of the spatial distribution for distant or small clusters.*

A mapping from the distribution $f_x(x)$ of the one dimensional projection of a cluster, as in Figure 2.9, to the corresponding angular power distribution can be derived. As the mapping from position to angle is $\alpha = \arctan(x/L)$, the angular distribution is in Appendix 2.B shown to be

$$f_\alpha(\alpha) = \begin{cases} \frac{L f_x(L\tan\alpha)}{\cos^2\alpha} & -\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2} \\ 0 & \textit{otherwise.} \end{cases} \tag{2.97}$$

The corresponding distribution for $a = \cos\alpha$ can be obtained through (2.78). If there is no elevation spread, then the Doppler spectrum can be directly obtained using (2.97) in (2.78) and (2.79), otherwise the integral in (2.85) has to be calculated.

2.4.5 Channel correlation due to a cluster

The correlation is given by the integral (2.77) which now also includes the angle to the center of the cluster

$$r_h(\tau) = \int_{-\pi}^{\pi} f_\alpha(\alpha) e^{\jmath\omega_D \cos(\alpha+\theta)\tau} d\alpha. \tag{2.98}$$

There is a connection between how fast the correlation decays and the

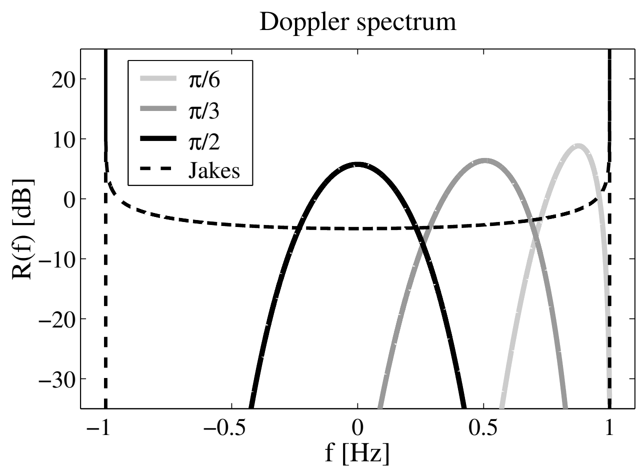

Figure 2.11: The Doppler spectrum for a Gaussian cluster along a line for three different angles of incidence θ, 30°, 60° and 90° (that is $\pi/6$, $\pi/3$ and $\pi/2$ radians). The Jakes spectrum is plotted for comparison. The Doppler shift is normalized to $f_D = \omega_D/2\pi = 1$ Hz. The main bulk (90%) of the power received from the cluster is here seen under an angle of 20° from the transceiver. This effect is obtained with a Gaussian cluster, with standard deviation $\sigma = 10$m, along a wall, 100m away from the transceiver.

predictability of the channel using a linear predictor, as will be evident in Chapter 6. Using an optimal linear predictor with just one coefficient the NMSE of a prediction τ time steps ahead of a noise free channel is $1 - |r_h(\tau)|^2/r_h^2(0)$, where $|r_h(\tau)| \leq r_h(0)$. This follows from equation (6.13)

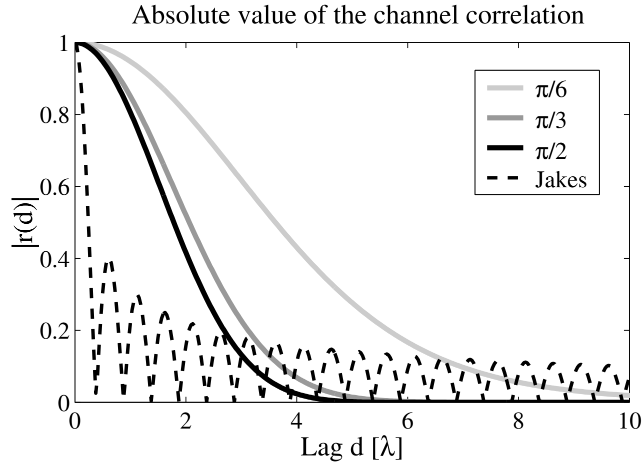

Figure 2.12: The absolute value of the correlation for the channel with the Doppler spectrum given as in Figure 2.11.

in Section 6.2. When there is an angular distribution due to e.g. diffuse scattering, the correlation decays with increasing correlation lag, as opposed to a single reflector and the plane wave assumption, for which the magnitude of the correlation is constant. In the case of omni-directional scattering, Jakes model, the magnitude of the correlation of the channel reduces to half the initial value when moving just a quarter of a wavelength (0.242 wavelengths). The Gaussian clusters causing the Doppler spectra of Figure 2.11 are more localized in space and thus result in a slower decay of the correlations, as seen in Figure 2.12. Thus, a low number of contributing clusters, each highly localized in space, result in slower decay of the channel correlation and a higher predictability of the channel, than an omni-directional scattering environment. Broad-band channels where each tap is influenced by only a small number of clusters is thus beneficial for the channel prediction.

The decay of the correlations, seen in Figure 2.12, for the channels corresponding to a Gaussian cluster at different angles, have the familiar Gaussian shape. In the following we will see how this can be the case.

Highly centered angular distributions

If we assume that the distribution is highly centered around $\alpha = 0$ (cf. Figure 2.9) and that it decays rapidly, then we can make some approximations that lead to an analytical expression for the correlation of the channel due to distant clusters of scatters. The integrand (2.98) depends partly on $\cos(\alpha+\theta)$ which for small angles α can be approximated as

$$\cos(\alpha+\theta) = \cos\alpha\cos\theta - \sin\alpha\sin\theta \approx \cos\theta - \alpha\sin\theta. \tag{2.99}$$

Thus if $f_\alpha(\alpha)$ has power close to $\alpha = 0$ only and is virtually zero elsewhere we can use (2.99) in (2.98) to approximate the correlation as

$$r_h(\tau) \approx \int_{-\pi}^{\pi} f_\alpha(\alpha) \mathrm{e}^{-\jmath\alpha\tau\omega_D \sin\theta} d\alpha\, \mathrm{e}^{\jmath\tau\omega_D\cos\theta} = \Phi_\alpha(\tau\omega_D\sin\theta)\mathrm{e}^{\jmath\tau\omega_D\cos\theta}, \tag{2.100}$$

which is the characteristic function[4] $\Phi_\alpha(\omega)$ of $f_\alpha(\alpha)$ at $\omega = \tau\omega_D\sin\theta$, times an oscillating component with the frequency given by the Doppler frequency $\omega_D\cos\theta$.

From (2.6) we have that

$$\tau\omega_D = \tau v k = dk, \tag{2.101}$$

where d is the traveled distance, in a time interval τ and k is the wavenumber. In the following we will express the correlation in terms of d instead of τ to obtain a representation that is independent of the speed of the vehicle. The correlation of the channel over distance is thus approximated as

$$r_h(d) \approx \Phi_\alpha(dk\sin\theta)\mathrm{e}^{\jmath dk\cos\theta}. \tag{2.102}$$

Correlation due to a Gaussian cluster

The Gaussian PDF for the angle in (2.92) fulfills the assumption of rapid decay if σ_x^2/L^2 is sufficiently small. Then we can approximate the angular distribution as a Gaussian as in (2.93). The Fourier transform of a Gaussian function is also Gaussian, so the characteristic function is

$$\begin{aligned}\Phi_\alpha(\omega) &= \int_{-\pi}^{\pi} f_\alpha(\alpha)\mathrm{e}^{-\jmath\omega\alpha}d\alpha \\ &\approx \int_{-\infty}^{\infty} \frac{L}{\sqrt{2\pi\sigma_x^2}} \mathrm{e}^{-\frac{L^2}{2\sigma_x^2}\alpha^2} \mathrm{e}^{\jmath\omega\alpha} d\alpha = \mathrm{e}^{-\frac{\sigma_x^2}{2L^2}\omega^2}.\end{aligned} \tag{2.103}$$

Insert (2.103) into (2.102) to obtain the approximate correlation function

$$r_h(d) \approx \mathrm{e}^{-d^2k^2\sin^2\theta\sigma_x^2/2L^2}\mathrm{e}^{\jmath dk\cos\theta} = \rho^{(dk/2\pi)^2}\mathrm{e}^{\jmath dk\cos\theta}, \tag{2.104}$$

[4] The characteristic function of a random variable is the Fourier transform of the distribution, which here is $\Phi_\alpha(\omega) = \int_{-\pi}^{\pi} f_\alpha(\alpha)\mathrm{e}^{-\jmath\omega\alpha}d\alpha$

where the damping coefficient ρ is given by

$$\rho = \mathrm{e}^{-2\pi^2 \sin^2 \theta \sigma_x^2 / L^2}. \tag{2.105}$$

The damping coefficient ρ is normalized so that if the distance d is measured in units of wavelengths, that is $\lambda = 1$ and $k = 2\pi$, then the damping is ρ^{d^2}.

A distant Gaussian cluster of scatterers thus result in a correlation of the channel that is a damped complex sinusoid with the frequency determined by the cosine of the angle between the direction of motion and the center of the cluster. The damping is a function of the squared distance, as measured in wavelengths. The channel is thus highly correlated for a short distance, whereas its correlation decays rapidly as the distance increases.

To evaluate the accuracy of the approximation for the correlation it is compared to the true correlation for the channel, obtained by numerical calculations using the angular distribution (2.92) for a number of angles θ. This result is compared to the approximation given by (2.104). The angular distribution has 90% of the power within an angle of 5°. This corresponds to a Gaussian cluster with a standard deviation of 20m at a distance of 750m. As seen in Figure 2.13, the approximation in (2.104) is better for clusters that appear on the side than for clusters that are close to right in front of or behind the vehicle.

The approximation error increases not only with decreasing angle of incidence but also with the angular width of the cluster. This is natural as the approximation is derived under the assumption of a narrow angular distribution. Let the width be defined as the angle under which 90% of the power is seen. For a width of 10°, the approximation error for the correlation will have a magnitude that is around or below 1/100, for angles of incidence that are larger than 30°. The approximation (2.104) can of course be used on clusters that are seen under a larger angle (width) than 10°, but with lower accuracy. If a correlation error on the order of 1/10 for $\theta > 30°$ is acceptable, then the width may be up to 60°.

The Cauchy cluster

Measurements of the local scattering environment, e.g. [48], suggest that the angular distribution has a sharper peak than found in the Gaussian distribution. In [48] a Laplacian distribution for the scatterers is proposed. Another possible candidate is obtained if, instead of the Gaussian cluster along the wall, we assume a more Cauchy like distribution. The channel correlation will then take the familiar form of the correlation for an AR1 process, as seen in the following.

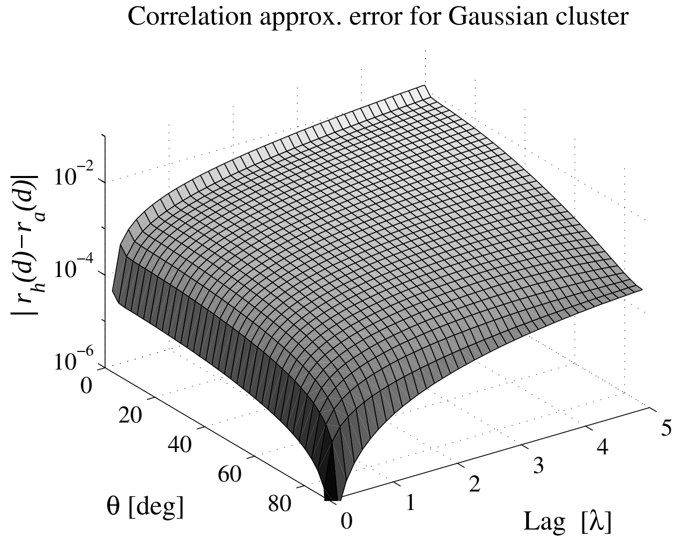

Figure 2.13: The magnitude of the approximation error as a function of angle of incidence θ and correlation lag d, using (2.104) denoted $r_a(d)$, instead of the true channel covariance $r_h(d)$, for a Gaussian cluster of scatterers. The angular distributions have 90% of the power within an angle of 5°.

The probability density function for the Cauchy distribution is given by

$$f_x(x) = \frac{\varsigma}{\pi(\varsigma^2 + x^2)}, \tag{2.106}$$

where $\varsigma > 0$ is a parameter governing the width of the distribution. The variance of a Cauchy distributed variable is infinite and, as seen in Figure 2.14, the mapping of this distribution to an angular distribution has fatter tails than the Gaussian distribution but it also has the desired narrow peak.

The corresponding angular distribution of scatterers, following from (2.106) in (2.97), is

$$f_\alpha(\alpha) = \frac{Ls}{\pi}\frac{1 + \tan^2\alpha}{\varsigma^2 + L^2\tan^2\alpha}, \tag{2.107}$$

where L is the distance to the cluster of scatterers. When $\varsigma \ll L$ this probability density function decays quite rapidly with increasing angle α and we can make an approximation of the angular PDF around $\alpha = 0$, setting $\tan\alpha = \alpha$,

$$f_\alpha(\alpha) \approx \frac{\varsigma}{\pi L} \cdot \frac{1}{\varsigma^2/L^2 + \alpha^2}, \tag{2.108}$$

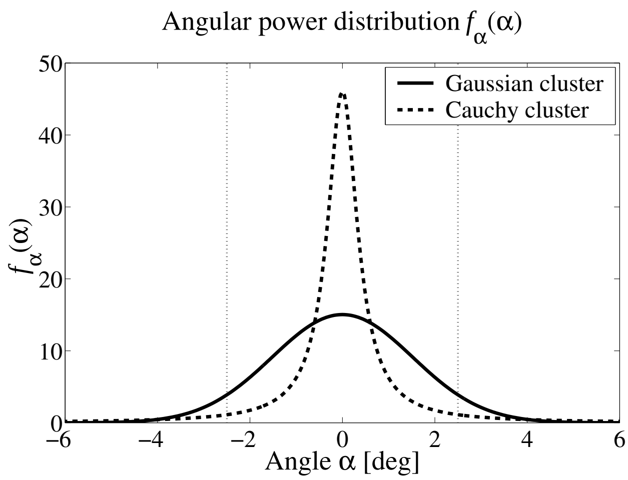

Figure 2.14: The angular distribution for a Gaussian and a Cauchy cluster. Both angular distributions have 90% of the power within an angle of 5°, but the angular spreads (the standard deviation for the angular distribution) are 1.5° and 5.6° for the Gaussian and Cauchy clusters, respectively. The large angular spread for the Cauchy cluster is due to the fat tails in the angular distribution.

which is also a Cauchy distribution. The characteristic function of this PDF (the Fourier transform), is an exponential damped function,

$$\begin{aligned} \Phi_\alpha(\omega) &= \int_{-\pi}^{\pi} f_\alpha(\alpha) \mathrm{e}^{-\jmath\omega\alpha} d\alpha \\ &\approx \int_{-\infty}^{\infty} \frac{\varsigma}{\pi L} \cdot \frac{1}{\varsigma^2/L^2 + \alpha^2} \mathrm{e}^{\jmath\omega\alpha} d\alpha = \mathrm{e}^{-|\omega|\varsigma/L}. \end{aligned} \tag{2.109}$$

The correlation function can be obtained using (2.100) as

$$r_h(d) \approx \mathrm{e}^{-|dk \sin\theta|\varsigma/L} \mathrm{e}^{\jmath dk \cos\theta} = \rho^{|dk|/2\pi} \mathrm{e}^{\jmath dk \cos\theta}, \tag{2.110}$$

where the damping factor ρ is given by

$$\rho = \mathrm{e}^{-2\pi|\sin\theta|\varsigma/L}. \tag{2.111}$$

The damping coefficient ρ is normalized so that if the distance d is measured in units of wavelengths, that is $\lambda = 1$ and $k = 2\pi$, then the damping is $\rho^{|d|}$.

The Cauchy cluster and the AR1-process

The correlation function for the channel due to a Cauchy distributed distant cluster of scatterers is recognized as the correlation function for a discrete time autoregressive system. A model for a sampled channel $h(t), t = [0, 1, \ldots[$, with a sampling rate resulting in m samples per traveled wavelength ($d = t/m$), is thus

$$h(t) = \left(\rho e^{\jmath 2\pi \cos\theta}\right)^{1/m} h(t-1) + w(t). \tag{2.112}$$

where $w(t)$ is an innovation noise with zero mean, that adds power to a flat fading channel at the same rate as it is dissipated due to the damping.[5] This AR1-process has a pole in $\left(\rho e^{\jmath 2\pi \cos\theta}\right)^{1/m}$ and if the noise $w(n)$ is white, then the correlation function of the channel is $r_h(n) = \rho^{|t/m|} e^{\jmath 2\pi (t/m)\cos\theta}$ which corresponds to (2.110) as the distance is measured in wavelengths. The AR1 process is thus pertinent for modeling the temporal/spatial effect of a clusters of scatterers with Cauchy like angular distribution on the channel.

The correlation due to the Gaussian cluster in equation (2.104) does not fit directly with the correlation of an AR1-process, as the damping is a function of the squared distance. Higher order ARMA-models can be used to approximate this behavior. If a model for a few wavelengths is sufficient, then an AR1-model can be used for the Gaussian cluster too, as seen in Figure 2.16.

Evaluation of approximation accuracy

To evaluate the accuracy of the approximation for the correlation function, it is compared to the true correlation for the channel, obtained by numerical calculations using the angular distribution (2.107) for a number of angles θ. This result is compared to the approximation given by (2.110). The angular distribution of (2.107) with ς/L =0.006 has 90% of the power within an angle of 5°, just as in the evaluation of the Gaussian cluster. The approximation for the Cauchy cluster is not as good as for the Gaussian cluster but it is still sufficient to motivate the use of AR-processes to model the effect of distant clusters of scatterers on the channel.

The damping is faster for the Gaussian cluster as compares to the Cauchy cluster, as seen in Figure 2.16 where the real part of the correlation function is plotted for a cluster at θ =60°. This is because the Cauchy cluster, in spite

[5] To ensure that the model of the channel is band-limited a colored noise with power only in the range given by plus/minus the maximum Doppler shift can be used.

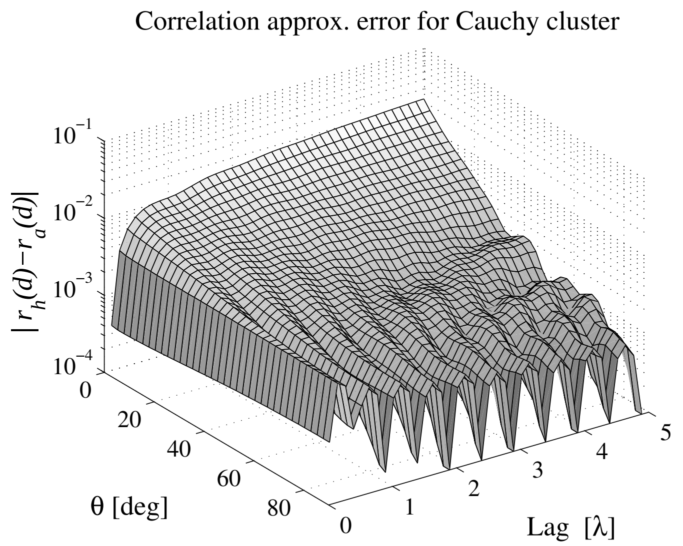

Figure 2.15: The magnitude of the approximation error as a function of angle of incidence θ and correlation lag d, using (2.110) denoted $r_a(d)$, instead of the true channel covariance $r_h(d)$, for a Cauchy cluster of scatterers. The angular distributions have 90% of the power within an angle of 5°.

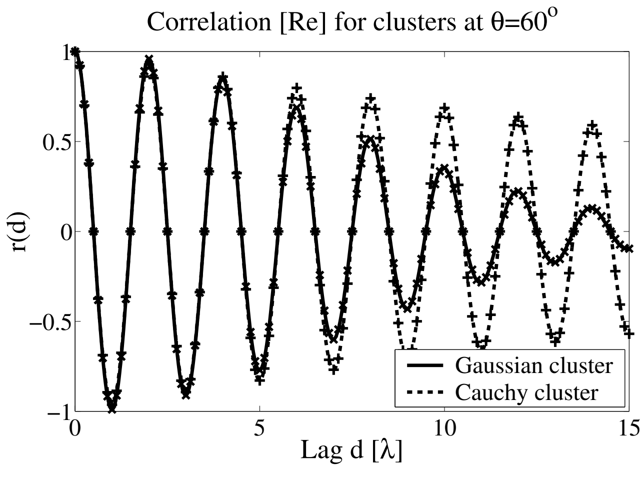

Figure 2.16: The real part of the channel correlation for a scenario with a Gaussian or a Cauchy cluster of scatterers at an angle of incidence of 60°. The $\times$ and $+$ are the correlation according to the approximations (2.104) and (2.110) respectively. Up to a few wavelengths both correlation functions can be approximated by the correlation for an AR1-process.

of the fat tails of the distribution, result in an angular distribution where the peak is narrower than for the Gaussian counterpart. The cluster is thus more like a point source, that does not have a damping in the correlation at all. For a few wavelengths there is hardly any difference between the correlation function for the channel due to the Gaussian and the Cauchy cluster. The approximations (2.104) and (2.110) follows the true correlations closely over the depicted range.

A number of clusters

There exists a mapping from the spatial distribution of scatterers, to the angular distribution of received power, as was exemplified above for the flat world scenario. In the environment there can be a large number of clusters, all contributing to the total received power. If we assume that the contributions from the different clusters of scatterers are uncorrelated, then we can write the correlation as a sum of the contributions from the different cluster, cf. (2.68),

$$r_h(\tau) = \sum_n p_n \int_\gamma f_{n,\gamma}(\gamma) \mathrm{e}^{\jmath \omega_D \cos\gamma \tau} d\gamma, \tag{2.113}$$

where p_k is the average power contribution from cluster k and $f_{k,\gamma}(\gamma)$ is the angular distribution of the cluster seen from the receiver. Each cluster can thus have an unique azimuthal and elevation angular spread. A reflector, or a direct path, that does not have any angular spread has a angular distribution which is a Dirac pulse at the angle of arrival.

The Doppler spectrum will thus also consist of a weighted sum of the individual Doppler spectra from the clusters (illustrated in Figure 2.17),

$$R_h(\omega) = \sum_n p_n \frac{2\pi}{\omega_D} f_{n,c}\left(\frac{\omega}{\omega_D}\right), \tag{2.114}$$

where $f_{n,c}(\cdot)$ is the distribution for $c = \cos\gamma$ for the n:th cluster.

2.4.6 The sampled channel

The continuous impulse response was discretized as a FIR-filter in Section 2.3, equation (2.28), where the coefficients of the discrete impulse response are changing continuously. We can not observe the channel directly but it can be estimated from the received data as will be described in Chapter 4. What is obtained from the channel estimation are snapshots of the

The angular distribution of power, $f_\alpha(\alpha)$ The Doppler spectrum, $R_h(\omega)$

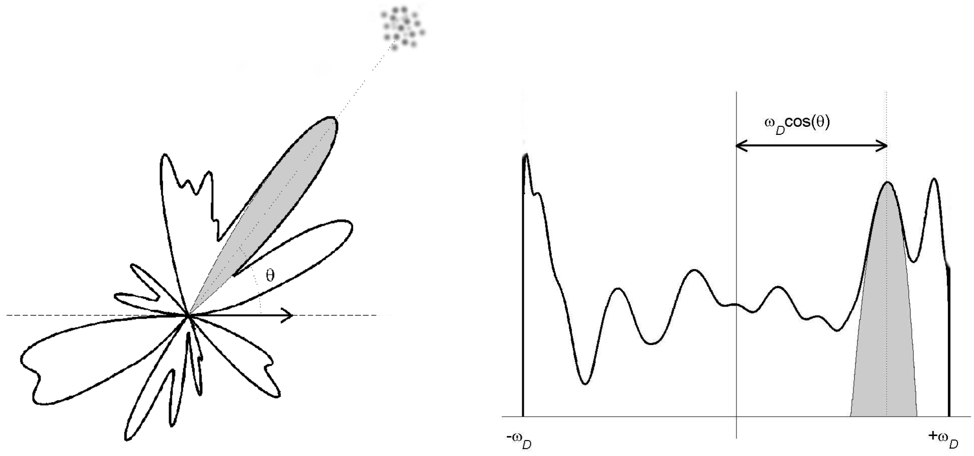

Figure 2.17: The angular distribution of power and the corresponding Doppler spectrum for 16 clusters with different powers. The grey dots is the cluster of scatterers, resulting in the power around angle θ in $f_\alpha(\alpha)$ (to the left) and the corresponding gray power in the Doppler spectrum (to the right).

impulse response at discrete time instances. Thus, the channel is sampled not only in delay of the impulse response, but also in time. If the dynamics of the channel is assumed to be band-limited, as in (2.76), then the channel only has to be sampled with twice the Doppler frequency. Due to the nonlinearities discussed in Section 2.3, causing a time varying angular power distribution, the channel will not be perfectly band-limited, which requires a higher sampling rate. Another problem, also calling for a higher sampling rate of the channel, is the estimation error arising in the channel identification. The reason is that noise reduction becomes simpler if the channel is oversampled with respect to the dynamics of the channel, as seen in Chapter 5.

In the rest of the thesis we will consider discrete time models for the channel, for which we shall exploit the findings on clusters properties presented earlier.

2.5 ARMA Modeling of the Channel

The Doppler spectrum that is observed in an environment where there are clusters of scatterers is continuous. When the number of scatterers and reflectors within the clusters is high, the contribution from each cluster can be modeled as a narrow-band filtered noise, as was discussed in the beginning of Section 2.4. In the Doppler spectrum there will be a peak with a certain width and falloff depending on the angular power distribution and on the direction of movement of the transceiver.

In this section we consider the feasibility to fit low order AR and ARMA models to give a similar spectrum as a cluster. The simplest possible model that results in a clear peak in the frequency domain is an AR filter with one complex pole. From the results above, e.g. equation (2.110), we know that the effect of one cluster of scatterers can be modeled by such an AR1-process.

A sum of damped oscillators

A conventional model for a fading tap is based on N point reflectors and scatterers, as in equation (2.65). Each channel tap may then be modeled by a weighted sum of the outputs from N complex oscillators, which have slowly varying frequencies (Doppler shifts) [18], [21]. A main result of Section 2.4 is to introduce damping into this description. A one pole AR filter per cluster, as in (2.112), corresponds to a linear time-invariant model that is valid for short intervals and that is well suited for an environment where the clusters consist of Cauchy distributed scatterers. Consequently, N clusters will thus correspond to a sum of N *damped* oscillators. Let $x_n(t)$ denote the contribution to the tap of the channel due to the n:th cluster. Using the model (2.112) for $x_n(t)$,

$$x_n(t) = \rho_n e^{\jmath 2\pi \cos\theta_n / m} x_n(t-1) + w_n(t). \tag{2.115}$$

Here, m is the number of samples per traveled wavelength, θ_n is an average angle of incidence for ray n and $0 \leq \rho_n \leq 1$ introduces damping. The damping is due to both averaging over slowly time-varying frequencies, as well as the presence of continuous scatterers. Uncertainties in the modeling are represented by $w_n(t)$, which are all band-limited and have zero means, since $x_n(t)$ and $h(t)$ are band-limited, but $w_n(t)$ are otherwise unknown. With the contribution from cluster n given as in (2.115), the tap is obtained as

$$h(t) = \sum_{n=1}^{N} \alpha_n x_n(t). \tag{2.116}$$

The complex weighting α_n includes the path loss and the phase. The sum over the contributions for the clusters can thus be expressed as a sum of N AR1-processes, which becomes

$$h(t) = \sum_{n=1}^{N} \frac{\alpha_n}{1 - q^{-1}\rho_n e^{\jmath 2\pi \cos\theta_n/m}} w_n(t), \tag{2.117}$$

Equation (2.117) can be expressed as an autoregressive part in $h(t)$ and a moving average part for the innovations,

$$\left[\prod_{n=1}^{N} \left(1 - q^{-1}\rho_n e^{\jmath\omega_n}\right)\right] h(t) = \sum_{n=1}^{N} \left[\prod_{l \neq n} \left(1 - q^{-1}\rho_l e^{\jmath\omega_l}\right)\right] \alpha_n w_n(t), \tag{2.118}$$

where $\omega_n = 2\pi\cos\theta_n/m$.

ARMA modeling

The sum of stochastic processes in (2.117) and (2.118) can be reformulated as *one* process, with the same second order moments. This operation is called spectral factorization, and it results in an ARMA innovation model of the form

$$h(t) = \frac{1 + c_1 q^{-1} + \ldots + c_{n_C} q^{-n_C}}{1 + d_1 q^{-1} + \ldots + d_{n_D} q^{-n_D}} w(t) \triangleq \frac{C(q^{-1})}{D(q^{-1})} w(t) \tag{2.119}$$

where $n_D = N$ and $n_C = N - 1$ and where $w(t)$ is a band-limited zero mean noise with variance σ_w^2. The filter $C(q^{-1})/D(q^{-1})$ is a stable minimum phase model of the dynamics of the channel. The AR part is given as in (2.118) with $D(q^{-1}) = \prod_{n=1}^{N} \left(1 - q^{-1}\rho_n e^{\jmath\omega_n}\right)$, with one pole per cluster. The polynomial $C(q^{-1})$ is a result of the combination of AR-processes.

If the damping factors are close to one, that is the poles are close to the unit circle, then the autoregressive part of the dynamics will dominate the model (2.119). Thus, even if the correct model is ARMA, an AR model might offer several advantages and a sufficiently good fit to the channel dynamics. As it generally is harder to estimate zeros as compared to poles, the AR model has been preferred in many channel predictors [22],[27]. In this thesis we will consider both AR and ARMA models for the channel, in the design of the channel predictors.

Time varying models

When the transceiver moves, the angle towards any given cluster changes. This result in a corresponding change of the position of the pole in the AR1 model for the corresponding contribution to the channel. This is a slow process, as compared to the channel dynamics, when the clusters are distant. However, for near by scatterers this effect can cause problems, as we saw in Section 2.3. In either case the ARMA model has to be updated, either iteratively or on a block basis. How often the model has to be re-estimated depends on how fast the angles towards the contributing scatterers change. In an environment where the power contribution from distant clusters dominates, the model can stay valid for several meters.

2.6 Conclusion and Implications for Predictor Design

- The linearized model (2.49) where a tap $h_m(t)$ in a mobile radio channel is described as a sum of weighted complex sinusoids,

  $$h_m(t) = \sum_{n=1}^{N} \alpha_{n,m} e^{\jmath \omega_{D_n} t}, \tag{2.120}$$

  is a valid approximation of the channel dynamics in an environment with plane waves and linear motion with constant velocity of the transceiver. The plane wave approximation is valid when the reflectors and scatterers are distant.

  According to this model, the tap is *perfectly predictable* by a linear FIR-predictor with N coefficients. Such predictor structures is thus of interest.

- The extended model (2.32)

  $$h_m(t) = \sum_{n=1}^{N} \alpha_{n,m}(t) e^{\jmath \phi_n(t)}, \tag{2.121}$$

  with a quadratic term in the phase (2.37)

  $$\phi_n(t) \approx \omega_{D_n} t + \frac{k r_n^{MS}}{2} \left(\frac{vt}{r_n^{MS}} \right)^2 \sin^2 \theta_n, \tag{2.122}$$

  should be used when the waves are spherical or cylindrical. This is the case when the contributing scatterers are close to the receiver.

  - This model includes the changing angle from the transceiver to the secondary source and thus increase the valid range for the model as compared to (2.120).
  - The quadratic term in the phase also appears when driving in curves and other situations where the assumptions of linear motion and constant velocity are invalid.
  - When the time interval $T_{\pi/3}$ defined by (2.57) is small, in relation to how often the channel model is updated, the linearized model (2.120) is insufficient. The time interval $T_{\pi/3}$ thus indicates if the quadratic term of the phase is needed or not.
  - This extended model motivates the use of adaptive predictors or block based predictors, that frequently update their coefficients, to follow the changing dynamics of the channel.

- When the number of contributing paths N is very large, it is convenient to make a statistical description of the channel instead of the deterministic description used in (2.120).

  - A group of buildings or other objects that reflect or scatter radio waves, can be seen as a cluster of reflectors or scatterers. The effect on the channel of such a cluster can be modeled as a narrowband filtered Gaussian noise.
  - Distant clusters of scatterers or reflectors seen under a small angle result in channels that can be approximated as sums of damped complex sinusoids with an added innovation. The frequency of the oscillation depends on the angle towards the center of the cluster whereas the damping, which is not necessarily exponential, depends on the shape of the cluster.
  - A Cauchy like cluster of scatterers result in a sharper peak in the angular distribution than a Gaussian cluster. The corresponding channel correlation is approximately the same as for an AR1-process.
  - When the elevation distribution of the clusters are taken into account the corresponding peaks in the Doppler spectrum generally becomes wider and lower, as compared to a model assuming a flat world.

- The statistical model, using the plane wave approximation, indicates that a distant cluster can be modeled as a damped complex sinusoid, instead as in (2.120) where the contribution from a source is modeled as a single complex sinusoid.

- Damped sinusoids can be modeled as stochastic AR1-processes if the damping is exponential, as for the distant Cauchy cluster. An AR1 process is a good model of the correlation function over several wavelengths also for Gaussian clusters.
- A sum of stochastic AR1-processes can, by spectral factorization, be modeled as a single ARMA-process with the same first and second order statistics. ARMA-models thus describe the dynamics of a mobile radio channel with contributions from many clusters.
- When the poles of the ARMA-model for the channel are close to the unit circle, an AR-model can render a good approximation to the dynamics.
- This motivates the use of predictors based on AR or ARMA-models for the dynamics of the taps.
- The linear ARMA or AR-models should be time-varying to encompass the changing directions towards the contributing clusters and other changes of the environment.

The statistical model with clusters of reflectors/scatterers resulting in an ARMA-model of the dynamics of a tap, is a more realistic description of the mobile radio environment than a sinusoidal model that is based on a finite number of scatterers/reflectors. Perfect predictability can not be expected for ARMA processes with slowly time-varying coefficients.

2.A The Required Size of a Reflector

To classify an object as a reflector or scatterer we need the concept of Fresnel zones. The first Fresnel zone is defined as the ellipse where the path length

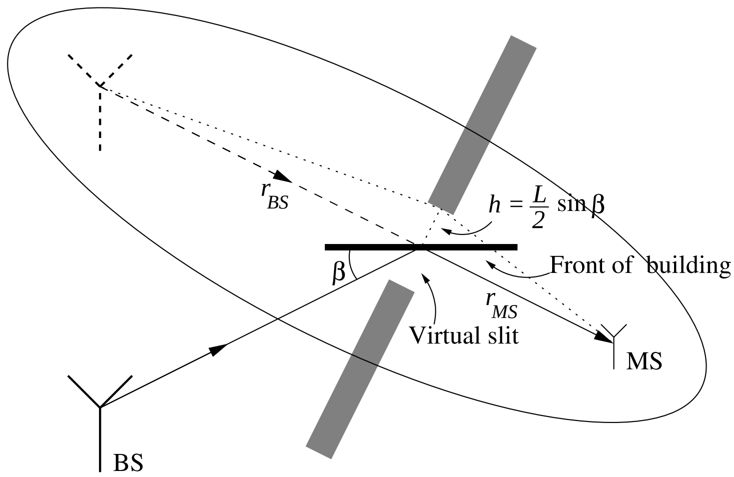

Figure 2.18: A building of width L reflects the waves from the antenna. The angle of incidence is β. The ellipse show the first Fresnel zone. The building acts as a slit of width $L\sin\beta$ for the virtual reflected source.

from the mirror image via a point on the ellipse exceeds the direct path from the mirrored image to the mobile station by $\lambda/2$. The base station is seen reflected in the building. A reflector can be viewed as a virtual slit for the mirror image as it is only inside the slit (the reflector) the mirror image is seen. If the slit opening allows at least the whole first Fresnel zone through, the object acts as a reflector (see Figure 2.18). The direct path is the path from the virtual mirror source to the receiver, which is equal the sum of the distance from the base station r_{BS} (or the secondary source) to the reflecting object and the distance from the object to the mobile station r_{MS}.

The difference in path length between the direct path and a path that passes the edge of the virtual slit is given by [36]

$$\Delta \approx \frac{h^2}{2}\left(\frac{r_{BS}+r_{MS}}{r_{BS}r_{MS}}\right). \tag{2.123}$$

With $\Delta = \lambda/2$ the slit let the whole first Fresnel zone through. The width of the object L, thus has to be large enough to let the first Fresnel zone through, for the building to act as a reflector. The slit is assumed to be tilted at an angle β, the angle of incidence, and the effective width of the slit is thus $L \sin \beta$. The smallest size for an object to be considered as a reflector is

$$L > \frac{2}{\sin \beta} \sqrt{\frac{r_{BS} + r_{MS}}{r_{BS} r_{MS}}} \lambda. \tag{2.124}$$

Objects smaller than this limit will cause noticeable diffraction, spreading energy not only in the angle of reflection. The diffraction thus causes a similar effect as a scatterer.

Example 2.7

Consider a scenario with an object acting as a reflector in the nth path. The carrier frequency is 1880 MHz (the wavelength λ is thus 16 cm) and the object is on a distance of 490 m from the base station and only 10 m from the mobile station. This object can be no smaller then 2.5 m to act as a reflector (that is for the most beneficial angle $\beta = \pi/2$ or $90°$). For an angle of reflection of $45°$ the object has to be at least 3.5 m wide to act as a reflector. An object further away from the mobile, 400 m from the base station and 100 m away from the mobile station, at an angle of $45°$ would have to be at least 10 m wide. Most objects along a road, as cars, are thus too small to act as reflectors. A building is often large enough to be considered as a reflector but most façades are not homogeneous reflecting surfaces. This causes a blend of reflection, diffraction and scattering from buildings.

2.B The Angular PDF from the Spatial PDF for a Linear Cluster

The mapping from the spatial distribution $f_x(x)$ of a one-dimensional cluster to the angular power distribution $f_\alpha(\alpha)$ for the cluster seen from the transceiver is derived in the following.

Here the scatterers are assumed to be spread on a line along a wall, following the distribution $f_x(x)$, where x denotes the coordinate along the wall. The wall is on the distance L from the transceiver and oriented as in Figure 2.9, with the wall perpendicular to a line from the antenna to the center of the cluster.

Let the attenuation depend only on the distance from the antenna to the center of the cluster L and not on the distance to the individual scatterers. This is a valid approximation if the standard deviation for the cluster is much smaller than the distance to the cluster ($\sigma_x \ll L$). Furthermore assume that all the scatterers reflect the same power and that the phases are random. The transformation from position to angle is

$$\alpha = g(x) = \arctan \frac{x}{L} \tag{2.125}$$

$$x = L \tan \alpha. \tag{2.126}$$

The Jacobian for the transformation is

$$g'(x) = \frac{\partial \alpha}{\partial x} = \frac{1}{L} \cdot \frac{1}{1 + x^2/L^2}. \tag{2.127}$$

Evaluate the Jacobian in $x = L \tan \alpha$,

$$g'(L \tan \alpha) = \frac{1}{L} \cdot \frac{1}{1 + \tan^2 \alpha} = \frac{\cos^2 \alpha}{L}. \tag{2.128}$$

The transformation from the spatial PDF to the angular PDF is obtained as

$$f_\alpha(\alpha) = \left. \frac{f_x(x)}{|g'(x)|} \right|_{x = L \tan \alpha} = \frac{L f_x(L \tan \alpha)}{\cos^2 \alpha}. \tag{2.129}$$

Distant cluster

For a distant cluster seen under a small angle, the main power contribution will be for angles where α is close to zero. Then $\tan \alpha \approx \alpha$ and $\cos^2 \alpha \approx 1$ which applied to (2.129) lead to an approximate expression for the angular PDF

$$f_\alpha(\alpha) \approx L f_x(L\alpha), \tag{2.130}$$

which is a scaled version of the spatial distribution.

Chapter 3

Measurement Data

To be able to verify the performance of the channel prediction algorithms proposed in this thesis, real measured channel impulse responses are needed. We have data from two measurement campaigns at our disposal. One was performed in urban Stockholm, which is a notorious mobile radio environment[1] and the other in the suburb Kista, just outside of Stockholm.

The data consist of snapshots of the channel impulse response for a few meters at 37 different locations. The snapshots are updated at 9.1kHz, which result in a fast sampled channel. These short data sets are ideal for evaluating the performance of block prediction methods, as a lot of different measurement locations and radio environments are represented in the material, while each set is too short to perform adaptive prediction. The high channel sampling rate is also a condition for the noise reduction presented in Chapter 5 to provide a large gain in predictability.

The data base consist of a number of channel-sounder measurements at different locations. The channel-sounder is a correlation channel-sounder, transmitting a sequence with good correlation properties to provide the best possible estimated channel impulse responses. The system is shown in Figure 3.1. The sequence $s(n)$ is sent through a transmitter over the mobile radio channel $\mathbf{h}(n)$ to the receiver, where the signal $y(n)$ is received. While no co-channel interference is present in the measurements, some noise $v(n)$ from the environment and the equipment corrupts the received signal.

[1] The city is situated on a large number of island with free radio propagation paths over the water, resulting in strong contributions to the channel impulse response at large delays.

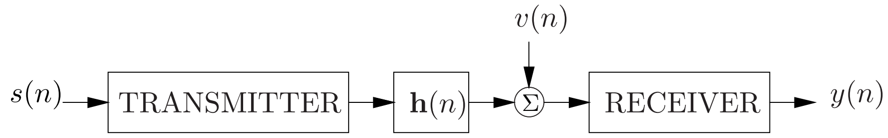

Figure 3.1: Channel sounder. The air interface, $\mathbf{h}(t)$*, is modeled as a time varying FIR-filter.*

Measurements

Wideband radio channel measurements were collected at 1880 MHz, at distances of 200 to 2000 m from the base station antenna placed on the roof of a high building. The mobile antenna was placed on a car driving in urban and suburban areas, mostly without line of sight. The vehicle velocity varied between 30 and 90 km/h. In total, 37 usable measurement runs were recorded at different positions. The measurements consist of 156.4 ms long recordings of the received signal at each measurement location. The transmitted signal consisted of 1430 repetitions of a sequence of length 109.4 μs. As the baseband sampling rate of the receiver was 6.4 MHz, each transmitted sequence of 109.4 μs resulted in 700 recorded samples.

Channel sampling

The impulse response of the channel is modeled by an FIR-filter, with parameters estimated by a block based least squares (LS) method. In Section 4.2 the procedure to identify FIR-channels from the measurements will be described in more detail. For each repetition of the transmitted sequence, that is for a block of 700 received samples, a new channel is identified, resulting in 1430 consecutive impulse responses at each measurement location. The *channel sampling frequency* is thus 6.4/700 MHz $\approx$ 9.1 kHz. Since the highest Doppler frequency in the measurements, that occurs for the maximal velocities around 90 km/h, is about 160 Hz ($f_D = f_c v/c$), the channel sampling frequency is sufficient to avoid aliasing of the Doppler spectrum. Compared to the maximal Doppler frequency the channel is oversampled by a factor of 28.

The block identification of the channel coefficients introduces an error, due to the time variation of the parameters of the channel during the identification interval, and an error due to the measurement noise. The size of the error due to time variation of the estimated coefficients, depends on the time-frequency (TF) product between the length of the identification interval (the block) and the Doppler frequency. This effect will be analyzed in more detail in Section 4.3. With the chosen block-length, the bias will be negligible for these measurements.

The time-span covered by the FIR-filter has to encompass all contributing paths. It was found that 120 taps, thus a time span of 18.75 μs, was sufficient. That is, only paths shorter than 5.6 km were found to contribute to the measurements. Note that 700 recorded samples are thus used to identify 120 complex parameters in the FIR-model.

An identification procedure using $s(n)$ and $y(n)$ in Figure 3.1 would result in an estimate of the channel convolved with the transmitter and receiver filters. Under the condition that the transmitter and receiver are linear and that the channel is slowly time varying, the order of the components can be interchanged as in Figure 3.2. To obtain a reference signal for identification

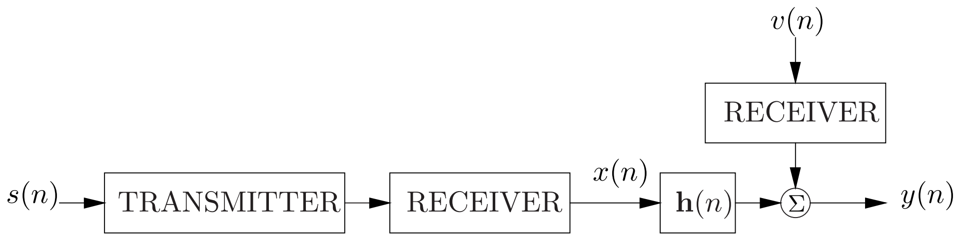

Figure 3.2: Equivalent channel sounding system. Used as a model in identification of the channel, $\mathbf{h}(n)$*.*

of the impulse responses, a back-to-back measurement is performed. The transmitted sequence, $s(n)$, is then sent through the transmitter directly connected to the receiver with a cable, as in Figure 3.3. On the receiver side the 700 samples of reference signal, $x(n)$, is obtained for the system without the air-interface. By using the back-to-back measurement for identification we avoid identifying the transmitter and receiver filters and obtain an estimate of the impulse response of the air interface.

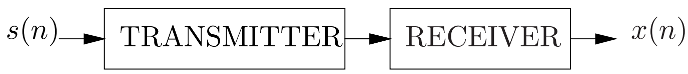

Figure 3.3: Back-to-back measurement on the channel sounder to obtain the reference signal $x(n)$*.*

Span of scales involved

The most important properties of the measurement are summarized in the Table 3.1. Note the large difference in scale. The entries of Table 3.1 are

Parameter	**Temporal measure**	**Spatial structure**
Carrier frequency	1880 MHz	15.95 cm
Baseband sampling rate	6.4 MHz	46.6 m
Impulse response length	18.75 μs	5.6 km
Max delay spread	3 μs	900 m
Channel update rate	109μs	0.9-2.7 mm
Length of measurement	156 ms	1.3-3.9 m

Table 3.1: Table of the parameters governing the measurements of the impulse responses. The vehicle velocity is 30-90 km/h (roughly 8-25 m/s). An interval in the spatial properties refers to the different vehicle velocities.

explained in the following.

- The carrier frequency is 1880 MHz and the wavelength is 15.9 cm. The waves will thus interact with objects larger than $\approx$10 cm.
- The baseband sampling rate is 6.4 MHz. That is, we obtain a sample every 0.156 μs, which corresponds to a spatial resolution of 46.6 m. Thus, the main contribution to each tap comes from reflections with path distances that differ less than 50 m. Thus, the geometry on the 10 m scale is involved.
- The impulse response length is 18.75 μs. This allows differences in path-lengths from the transmitter to the receiver of 5.6 km. Thus, the estimated impulse response accounts for reflections over the 10^3 m scale.
- The channel update rate is 109 μs which corresponds to a new estimate of the impulse response every 0.9-2.7 mm at vehicle velocities in the range 30-90 km/h. That is, we obtain at least 60 channel samples per wavelength. Thus we are working on the 10^{-3} m scale. With such small intervals, an assumption of stationarity over the channel estimation interval is well justified.

Chapter 4

Channel Estimation

4.1 Introduction

The mobile radio channel, as described in Chapter 2, has to be estimated in some manner to be made available for the receiver. In a communication system the estimation is generally performed using the received signal and applying some knowledge about the transmitted sequence of symbols. This chapter will mainly discuss least squares estimation, the most basic block method for channel estimation using training symbols. Block based estimation methods use a batch of received symbols to estimate the average channel in an interval where the channel is treated as time invariant. We shall call this a sample or snapshot of the channel obtained from each estimation interval. The continuous impulse response of the channel is thus sampled, not only in delay to form a discrete FIR-channel but also in time, resulting in snapshots of the time varying channel. The estimated samples of the channel will be corrupted by an estimation error that depends on the measurement noise, the transmitted symbols, the properties of the channel and the estimation algorithm.

In Section 4.2 some properties of the block based least squares method are recapitulated. Important for the noise reduction of Chapter 5, is that the estimation error results in a noise floor in the power delay profile, which can be used to obtain estimates of the variance of the estimation error. The LS channel estimation procedure is exemplified using the fast sampled measurement data from the database described in Chapter 3.

The error in the estimation, due to measurement noise and the deviation from a time invariant channel in the estimation interval, is analyzed in Section 4.3 for least squares estimation of a Rayleigh fading channel with Jakes spectrum. By a second order Taylor expansion the estimation error due to the time variation is approximated as the sum of a bias and an excess error, caused by the curvature and linear change of the channel, respectively. The resulting expressions can be used to obtain limits for how long estimation intervals that are reasonable to use under given conditions.

4.2 Identification Procedure

4.2.1 Channel model

In this work, the radio channel is described by a discrete time transfer function, i.e., a discrete time impulse response. This impulse response, denoted $\{h_k(n)\}_{k=0}^{\infty}$ or just $\{h_k(n)\}$, is time varying and the goal of the identification procedure is to estimate the time-dependent parameters in $\{h_k(n)\}$ as accurately as possible. The received signal is described as a convolution between the transmitted (pulse-shaped) signal, $x(n)$, and the impulse response $\{h_k(n)\}$ corrupted by an additive noise, $v(n)$:

$$y(n) = \sum_{k=0}^{\infty} h_k(n)x(n-k) + v(n). \tag{4.1}$$

Over short time intervals (a batch of data), the time-varying channel can be approximately described by a time invariant impulse response. A further simplification is to assume that the transfer function, $\{h_k(n)\}$, can be described by a time invariant finite impulse response (FIR) model of length M (following the result in (2.28)), in each time interval. The model (4.1) is then modified to

$$y(n) = \sum_{k=0}^{M-1} h_k x(n-k) + v(n), \tag{4.2}$$

where M has to be chosen large enough to encompass all significant contributing paths. By expressing (4.2) using in the unit delay operator q^{-1} ($q^{-1}x(n) = x(n-1)$), we obtain

$$y(n) = H(q^{-1})x(n) + v(n), \tag{4.3}$$

where $H(q^{-1}) = \sum_k h_k q^{-k}$. This model is a valid approximation for time segments that are short related to the channel variation. In Section 4.3 the effect of the time variation within the interval is investigated. In the following, batches of N data samples are used to obtain a description of the impulse response $\{h_k\}$ for each batch.

For a white data sequence $\{x(n)\}$ the variance of the received signal is

$$\sigma_y^2 = \mathrm{E}\{|y(n)|^2\} = \sigma_x^2 \sum_{k=0}^{M-1} |h_k|^2 + \sigma_v^2. \tag{4.4}$$

The instantaneous SNR for one batch of data is thus

$$\gamma = \frac{\sigma_x^2 \sum_{k=0}^{M-1} |h_k|^2}{\sigma_v^2}. \tag{4.5}$$

The average SNR is obtained as

$$\bar{\gamma} = \frac{\sigma_x^2 \sum_{k=0}^{M-1} \sigma_{h_k}^2}{\sigma_v^2}, \tag{4.6}$$

where $\sigma_{h_k}^2$ is the variance of tap k.

4.2.2 Empirical transfer function estimate

The convolution in (4.2) can be expressed as a multiplication in the frequency domain, i.e.,

$$Y(\omega) = H(\omega)X(\omega) + V(\omega). \tag{4.7}$$

The empirical transfer function estimate (ETFE) of $H(\omega)$ is then simply given as [49]

$$\hat{H}(\omega) = Y(\omega)/X(\omega) \tag{4.8}$$

and the impulse response estimate $\hat{h}_k$ is given by the inverse Fourier transform of $\hat{H}(\omega)$. This method for identification in the frequency domain usually gives results with unnecessarily high variance but due to its simplicity, it is still worth attention.

The variance in the frequency domain of an estimated transfer function will depend on signal to noise ratio in the frequency domain [50]. The power spectrum of the reference signal (which can be obtained from the back-to-back measurement), as seen in Figure 4.1, shows that the transmitted signal in the measurements does not excite frequencies beyond 2.5 MHz from the center frequency. This will result in low accuracy of the estimated channel impulse response for the highest frequencies, regardless of which identification method is chosen.

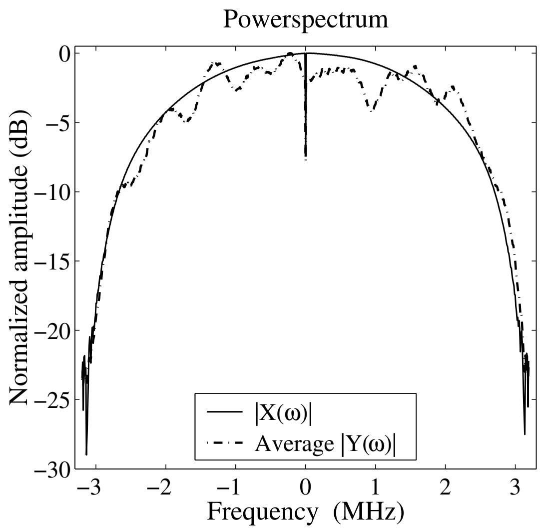

Figure 4.1: Normalized power spectrum of the transmitted signal $|X(\omega)|$ *and the average normalized power spectrum of the received signals at one measurement location (real data).*

4.2.3 The least squares method

The use of the empirical transfer function estimate would cause an unnecessarily high variance on the estimated parameters. The problem is particularly apparent for channel taps with low amplitudes. To obtain a better overall estimate of the channels, a least squares estimator [49] will be used.

Express equation (4.2) in vector form as

$$y(n) = \mathbf{x}^H(n)\mathbf{h} + v(n), \tag{4.9}$$

where

$$\mathbf{x}(n) = [x(n)\ x(n-1) \ldots\ x(n-M+1)]^H \tag{4.10}$$

is the vector of transmitted data and

$$\mathbf{h} = [h_0\ h_1 \ldots h_{M-1}]^T \tag{4.11}$$

is the parameter vector that is assumed time invariant in the estimation interval and $v(n)$ is a noise term assumed to have zero mean and variance σ_v^2. The total number of observations in one batch is N and the number of unknown parameters in $\mathbf{h}$ is M. Form the Toeplitz matrix for the transmitted data [1]

$$\mathbf{X} = [\mathbf{x}(1) \dots \mathbf{x}(N)]^H \tag{4.12}$$

and a received data vector

$$\mathbf{y} = [y(1) \dots y(N)]^T. \tag{4.13}$$

Similarly let $\mathbf{v}$ denote the vector of the N noise samples,

$$\mathbf{v} = [v(1) \dots v(N)]^T. \tag{4.14}$$

The Equation (4.9) can then for $n = 1 \dots N$ be formulated as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{v} \tag{4.15}$$

and the off-line LS estimate of the parameters of the channel is

$$\hat{\mathbf{h}} = \mathbf{X}^\dagger \mathbf{y} = \mathbf{h} + \mathbf{X}^\dagger \mathbf{v}, \tag{4.16}$$

or, by denoting the estimation error by

$$\mathbf{e} = \mathbf{X}^\dagger \mathbf{v}, \tag{4.17}$$

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{e}. \tag{4.18}$$

Here, $\mathbf{X}^\dagger = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$ denotes the Moore-Penrose inverse of the matrix $\mathbf{X}$ and $(\cdot)^H$ denotes complex conjugate transpose.[2]

Under the assumption that the noise $v(n)$ in (4.9) is zero mean, independent of the term $\mathbf{x}^H(n)\mathbf{h}$ and that $\mathbf{h}$ is time invariant, then $\mathrm{E}\{\hat{\mathbf{h}}\} = \mathbf{h}$. The covariance matrix of the estimation error $\mathbf{e}$ is then obtained as

$$\begin{aligned} \mathbf{R}_e &= \mathrm{E}\{\mathbf{e}\mathbf{e}^H\} = \mathrm{E}\{(\hat{\mathbf{h}} - \mathbf{h})(\hat{\mathbf{h}} - \mathbf{h})^H\} \\ &= \mathrm{E}\{\mathbf{X}^\dagger \mathbf{v}\mathbf{v}^H \mathbf{X}^{\dagger H}\} = \mathbf{X}^\dagger \mathbf{R}_v \mathbf{X}^{\dagger H}, \end{aligned} \tag{4.19}$$

where the noise covariance matrix, $\mathbf{R}_v$, is defined as

$$\mathbf{R}_v = \mathrm{E}\{\mathbf{v}\mathbf{v}^H\}. \tag{4.20}$$

[1] There will be samples with negative time indices in $\mathbf{X}$, samples that not necessarily are available. If these samples are unknown, then the problem can be circumvented by setting them to zero in $\mathbf{X}$. The matrix can also be truncated to $\mathbf{X} = [\mathbf{x}(M) \dots \mathbf{x}(N)]$, which only consist of known samples. Then $\mathbf{y}$ has to be truncated in the same manner.

[2] The vector $\mathbf{X}^H\mathbf{y}/N$ is an estimate of the cross-covariance vector, $\mathbf{r}_{yx}$, and $\mathbf{X}^H\mathbf{X}/N = \hat{\mathbf{R}}_x$ is the sample auto-covariance matrix of $\mathbf{x}(n)$. The LS solution in (4.16) is thus nothing but the Wiener-Hopf equations with this particular choice of estimators for the covariances.

The correlation matrix for the estimated channel, obtained as the expectation over noise realizations

$$\mathrm{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} = \mathbf{h}\mathbf{h}^H + \mathbf{X}^\dagger \mathbf{R}_v \mathbf{X}^{\dagger H}, \tag{4.21}$$

is of interest in the following discussion about the power delay profile.

When the noise is white, the noise covariance has the simple form $\mathbf{R}_v = \sigma_v^2 \mathbf{I}$ and the covariance of the estimation error in equation (4.19) reduces to

$$\mathbf{R}_e = \sigma_v^2 (\mathbf{X}^H \mathbf{X})^{-1}. \tag{4.22}$$

The estimation error will be studied further in Section 4.3.

4.2.4 Power delay profile estimates

In the previous section the channel $\mathbf{h}$ was treated as a deterministic time invariant vector in the estimation interval. When consecutive estimation intervals are considered, it is beneficial to treat each snapshot of the channel $\mathbf{h}(t)$, where t denotes the discrete time index for the snapshots of the channel, as a realization of a stochastic process. The process is assumed to be ergodic. Time averages and expectation over realizations will thus be the same.

The power delay profile (PDP) represents the distribution of the received power over delays. It is thus the expectation of the squared amplitude of the taps at one location. Through the PDP the effective length of the channel can be observed and it can also be used to measure the delay spread. Here we will use the PDP to obtain an estimate of the variance of the estimated parameters of the channel, where estimation errors will act as noise on the sequence of channel estimates.

The PDP is estimated as the time average over one measurement location of the power in the estimated taps $[|\hat{h}_0(t)|^2 \ldots |\hat{h}_{M-1}(t)|^2]$. Assume the channel tap to be time invariant over the estimation window of N received data samples and assume the noise to be independent of the regressors. The expected value for the power in one estimated tap, say tap k, estimated by the LS procedure from data corrupted by *white noise* is then obtained by using (4.22) in (4.21),

$$\mathrm{E}\{|\hat{h}_k(t)|^2\} = \mathrm{E}\{|h_k(t)|^2\} + \sigma_v^2[(\mathbf{X}^H\mathbf{X})^{-1}]_{kk}, \tag{4.23}$$

where the expectation is taken both over noise realizations and channel realizations. Here $(\cdot)_{kk}$ denotes the element on row and column k and the variance of the white noise is σ_v^2. As the estimated PDP is formed through time averaging of $\hat{\mathbf{h}}(t)$, it will have a bias given by the second term in (4.23).

This bias in the estimated PDF can be seen as a *noise floor* $\mathbf{n}_f$, at those channel taps that in reality are zero,

$$\mathbf{n}_f = \sigma_v^2 \,\mathrm{diag}[(\mathbf{X}^H\mathbf{X})^{-1}], \tag{4.24}$$

where diag$(\cdot)$ forms a column vector from the diagonal elements of a matrix. Thus, if $h_k(t) = 0, \forall t$, then $\mathrm{E}\{|\hat{h}_k(t)|^2\} = \sigma_v^2[(\mathbf{X}^H\mathbf{X})^{-1}]_{kk}$. The noise floor for tap k will be given by the kth element of the noise floor vector $\mathbf{n}_f$. We cannot expect to obtain any reliable estimates of the taps close to the noise floor. It could even be beneficial for the total accuracy to exclude those delays from the estimation procedure [51].

Verification of the noise floor for white measurement noise

To verify that the level of the noise floor, and thus the variance of the estimation error, is as expected, an estimated PDP from a measurement is compared to a Monte Carlo simulated profile. Here we use a dB scale with an arbitrary reference to display the measured data. This way of presenting measured data will be used throughout the thesis.

The simulated channel has a PDP formed after a pattern given by the estimated PDP. Some taps in the simulated channel should carry no signal power, so values in the PDP pattern under a certain threshold, which is chosen in an ad hoc way, is set to zero. The square root of the estimated PDP, with values under a threshold of -101.4 dB set to zero, is thus used as a pattern for the simulated channel parameters. The threshold level is chosen a little bit above the observed noise-floor of the estimated PDP in Figure 4.2, to include all the visible peaks from the estimated PDP into the pattern.

For each Monte Carlo trial, the taps in the channel are generated as the pattern PDP multiplied by normal distributed random complex numbers with zero mean and variance one. The generated channels will then have independent taps and on average the desired PDP. Each Monte Carlo trial will thus generate a snapshot of the channel with the desired statistical properties. The input signal used in the measurements are transmitted through the generated channel and the received signal is corrupted by an additive white Gaussian noise, the measurement noise. This noise has the same variance as the one estimated from the measurements. The simulated channel is then identified using equation (4.16), using 700 samples just as for the measurements.

A total of 100 Monte Carlo trials are performed with different noises and channel parameters. In Figure 4.2 the estimated PDP for one measurement

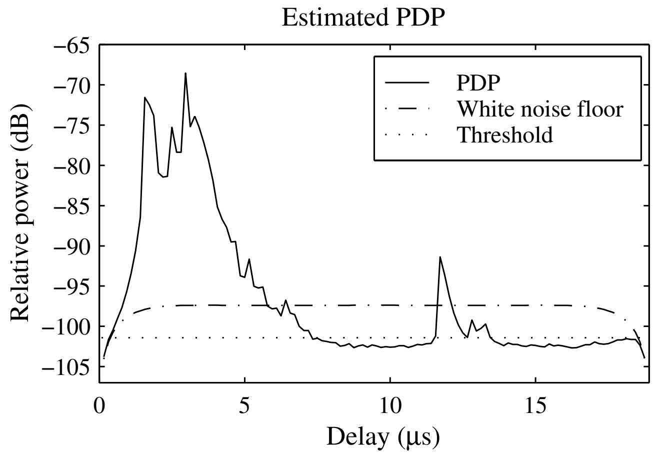

Figure 4.2: The estimated PDP for measurement 23 (cf. Chapter 3) and the calculated noise floor under the white noise assumption (4.24). The part of the PDP under the threshold is set to zero in the simulation.

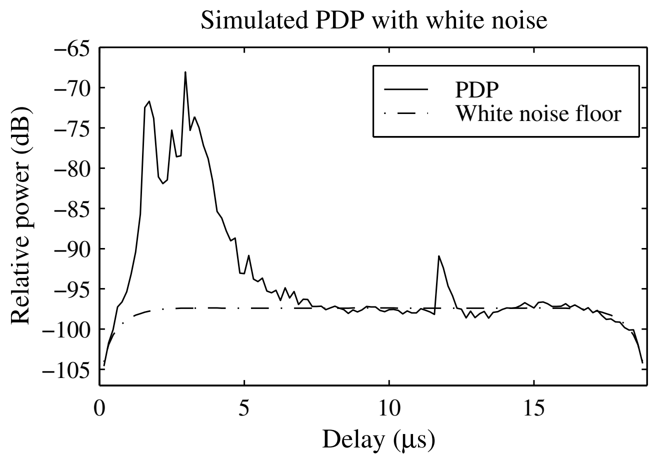

Figure 4.3: The simulated PDP with white noise. Measurement 23 (cf. Chapter 3) is used as a pattern for generating a set of channels used for Monte Carlo simulations.

location is displayed. I Figure 4.3 the average result from the 100 Monte Carlo trials generated with this PDP is depicted. In the measured data, Figure 4.2, there is significant deviation between the noise floor calculated as in (4.24) (dash-dotted curve) and the average power of the taps with the least power (dotted curve). This is not the case in the simulated PDP with white noise (Figure 4.3). There, the estimated noise floor for the simulated PDP coincides with the power level for the smallest estimated taps, those that in the pattern are zero and thus carry no signal power and only noise power, as expected from the theory. The difference between the result expected from theory and the measured PDP (in Figure 4.2) indicates that some assumptions about the data or the model are wrong. A possible error source could be the presence of colored measurement noise.

Noise floor for colored noise

To investigate the presence of colored noise we identify the channel using a prediction error method (PEM) with a more general model structure, including a noise-filter $C(q^{-1})$,

$$y(n) = H(q^{-1})x(n) + \underbrace{C(q^{-1})\nu(n)}_{v(n)} \tag{4.25}$$

where $\nu(n)$ is a white noise [50]. In this example we identify a moving average noise-filter $C(q^{-1})$, of degree five, for each batch of data in the measurement at the same location as before. Even though the identified channel parameters $h_k(t)$ showed large variations over the time t, the identified noise-filter parameters were rather stable for all the snapshots, as can be seen in Figure 4.4.[3] The average parameters for the noise coloring filter $C(q^{-1})$ are $[1.0, 0.20+0.03i, -0.2, 0.1, -0.08, 0.05]$, a rapidly decaying impulse response resulting in a weakly colored noise. The assumption of white noise is thus not valid.

The LS estimate $\hat{\mathbf{h}}$ remains unbiased even with colored noise, but the expected value for the power will change from what is given in equation (4.23) to

$$\mathrm{E}\{|\hat{h}_k(t)|^2\} = [\mathbf{X}^\dagger \mathbf{R}_v \mathbf{X}^{\dagger *}]_{kk} + \mathrm{E}\{|h_k|^2\}. \tag{4.26}$$

The noise floor is thus altered to

$$\mathbf{e}_f = \mathrm{diag}(\mathbf{X}^\dagger \mathbf{R}_v \mathbf{X}^{\dagger *}), \tag{4.27}$$

where $\mathbf{R}_v$ is the covariance matrix (4.14),(4.20) of $v(n) = C(q^{-1})\nu(n)$.[4] To calculate the noise floor as in (4.27) the noise covariance $\mathbf{R}_v$ must be estimated.

[3] This was somewhat unexpected, as the properties of environmental noise would change just as fast as the mobile radio channel when the mobile moves. If, on the other hand, the noise is produced internally in the receiver, then the noise filter could be close to time-invariant.

[4] The equation (4.23) is just a special case of (4.26) with $\mathbf{R}_v = \sigma_v^2 \mathbf{I}$.

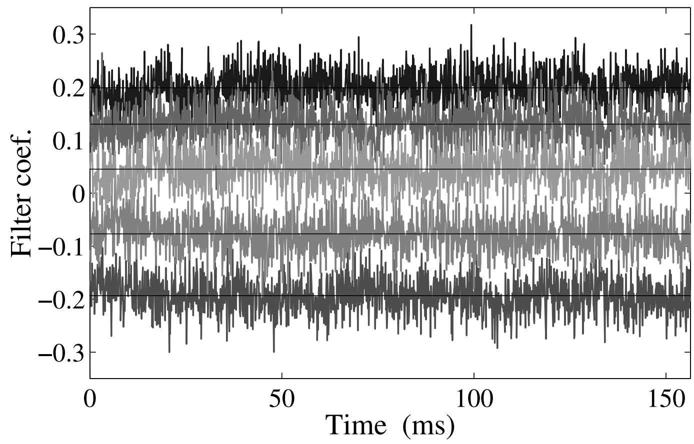

Figure 4.4: Real part of the identified noise-filter parameters in (4.25) from measurement 23. The darker the parameter the smaller the delay corresponding to the parameter in the $C(q^{-1})$ polynomial. The solid lines are the means.

When an additive colored noise is introduced in the simulations, a similar effect as for the measured data can be obtained, see Figure 4.5. The simulation is performed in the same manner as previously described, using LS for estimating the impulse responses, but the noise is now generated as a white noise filtered by a 40 tap long exponentially decaying noise filter. This is an ad hoc choice, built on the assumption that the echoes represented by the noise term decay rapidly. (If there is internal colored noise in the receiver, there are no echoes. Still the rapidly decaying impulse response is similar to that observed for the noise filter in measurements.) The taps in the noise-filter are selected as $c_k = (b_k + 1)e^{-k/2}/b$, where b_k is the absolute value of a normal distributed random variable with zero mean and variance one. The normalization factor b is selected so that the filter does not change the variance of the noise, i.e. $\sum |c_k|^2 = 1$. The variance of the colored noise is the same as for the white noise in the previous simulation. The colored measurement noise results in a lower noise floor than white noise with the same variance, when using LS-estimation for the impulse response.

The average parameters for the estimated noise-filter and the variance estimate from the PEM have been used to estimate the covariance matrix $\hat{\mathbf{R}}_v$ for the colored measurement noise on the measured data. The resulting noise floor, obtained as in (4.27), is in agreement with the observed noise floor in the estimated PDP from the measurement (the channel is estimated

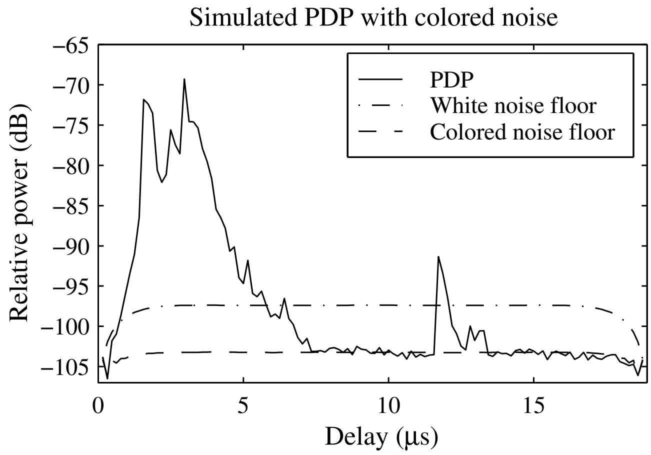

Figure 4.5: Simulated PDP, estimated by LS, with colored noise. The colored noise has the same variance as estimated from the measurement. The coloring noise channel impulse response is a rapidly decaying exponential function.

by LS).

From experience with measurements and simulations we can conclude that the observed lowest levels in the estimated PDP can be used as an estimate of the channel tap estimation error variance. Thus, an estimated tap can be modeled as the true value with an additive estimation error, as

$$\hat{h}_k(t) = h_k(t) + e_k(t), \tag{4.28}$$

where the variance of the estimation error $e_k(t)$ is given by the noise floor obtained from the estimated PDP. There is thus no need to estimate the variance using (4.27). We have also seen that calculating the noise floor using (4.24) can result in an overestimation of the noise floor.

The estimated error variance will be used in Chapter 5 in the design of a noise-reduction smoother, for filtering of the estimated channel parameters.

4.2.5 Choice of identification procedure

All of the evaluated methods, ETFE, LS and PEM give roughly the same results for the taps that contain the most energy. To find the best method we investigate the lowest level (the noise floor) in the estimated PDP, which is linked to the variance of the estimation error.

The ETFE produces estimated channel impulse responses in which the lowest level in the PDPs are a few dB over the lowest levels produced by the

LS estimation (this level varies from measurement to measurement though). The PEM, taking the colored noise into account, also results in higher lowest levels in the PDP as compared to the LS-method. This is because PEM estimates more parameters, which causes a higher variance in the estimates. In addition it is generally hard to estimate the noise-filter, $C(q^{-1})$ in (4.25), with good accuracy from a limited amount of data. Thus, we do not in this application gain anything in accuracy for the estimated impulse responses by using PEM instead of LS. As discussed in Section 4.2.4, the lowest level in the PDP can be linked to the variance of the estimation error. Since both the ETFE and PEM methods produce estimates with higher variance than LS for the taps with low amplitude, the LS-method is selected for identification of the channel.

4.3 Analysis of the LS Estimation Error on the Jakes Channel

A channel in a wireless communication link is often treated as time invariant over an estimation interval, during which a least squares estimate of the channel is calculated, using training symbols. By a second order Taylor expansion of the channel, the estimation error due to time variation can be approximated as the sum of a bias and an excess error, which are due to the curvature and linear change of the channel, respectively. In the following, approximations for the variances of the different contributions to the channel estimation error will be derived, for a Rayleigh fading channel with Jakes spectrum.

The LS-estimate

The LS estimate of the channel impulse response, $\hat{\mathbf{h}}$, for a time-varying channel, is obtained as in (4.16)

$$\begin{aligned} \hat{\mathbf{h}} &= \mathbf{X}^{\dagger}\mathbf{y} \stackrel{\triangle}{=} \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{y} \qquad (4.29) \\ &= \left(\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)\right)^{-1}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{h}(n) + \mathbf{X}^{\dagger}\mathbf{v}. \qquad (4.30) \end{aligned}$$

The expression (4.30) differs from (4.16) in that the unknown FIR channel is represented by a time-varying M-tap channel vector

$$\mathbf{h}(n) = [h_0(n)\ h_1(n) \ldots h_{M-1}(n)]^T. \qquad (4.31)$$

The channel estimate, $\hat{\mathbf{h}}$, can thus be seen as a *weighted average* of $\mathbf{h}(n)$ over the estimation interval $[1\ N]$, corrupted by an additive noise. The desired channel is the true channel in the middle of the estimation interval, as seen in Figure 4.6.

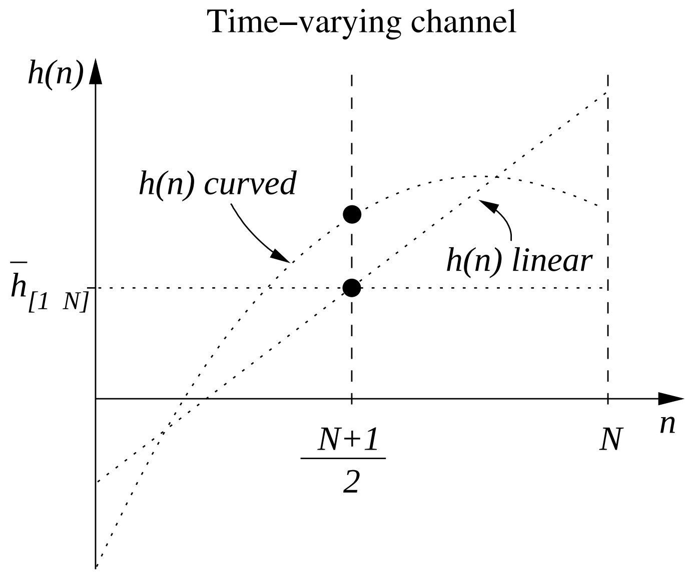

Figure 4.6: For a channel that changes linearly the average value coincides with the value in the middle of the interval. For a curved channel the value in the middle differs from the average value. The weighted average, as in the LS-estimate of the channel, differs from the true average mainly due to the linear change of the channel.

Continuous and discrete time

The discrete channel is denoted $\mathbf{h}(n)$, where n is the index at baseband sampling period. The continuous channel has a different time argument and is denoted $\mathbf{h}(t_c)$, where t_c denotes continuous time. The discrete and continuous channel are related as $\mathbf{h}(n) = \mathbf{h}(t_c)|_{t_c=nt_s}$, where t_s is the baseband sampling rate.

4.3.1 Noise-induced error

In (4.17) the error due to measurement noise is represented by

$$\mathbf{e}_{\mathrm{N}} = \mathbf{X}^{\dagger}\mathbf{v}. \tag{4.32}$$

For a white measurement noise that is independent of the data sequence $\{x(n)\}$, used for identification, the covariance matrix for this error is given by (4.22)

$$\mathbf{R}_{\mathrm{N}} = \mathrm{E}_v\{\mathbf{e}_{\mathrm{N}}\mathbf{e}_{\mathrm{N}}^H\} = \sigma_v^2 \mathbf{Q}, \tag{4.33}$$

where $\mathrm{E}_v\{\cdot\}$ denotes expectation over the noise and

$$\mathbf{Q} \stackrel{\triangle}{=} (\mathbf{X}^H\mathbf{X})^{-1}. \tag{4.34}$$

Assume $x(n)$ to be samples from a white complex circular sequence with zero mean and variance σ_x^2. The ensemble average with respect to the data sequence $\{x(n)\}$ will be denoted $\mathrm{E}_x\{\cdot\}$. In Appendix 4.A the expectation of the inverse of the unnormalized sample covariance matrix, $(\mathbf{X}^H\mathbf{X})^{-1}$, is approximated as (4.67)

$$\mathrm{E}_x\{\mathbf{Q}\} \approx \frac{1}{\sigma_x^2 N}\left(1 + \frac{M + \kappa_x - 2}{N}\right)\mathbf{I}. \tag{4.35}$$

Utilizing (4.35), the average $\mathrm{E}_x\{\mathbf{R}_{\mathrm{N}}\}$ can be approximated by

$$\bar{\mathbf{R}}_{\mathrm{N}} = \mathrm{E}_x\{\mathbf{R}_{\mathrm{N}}\} = \sigma_v^2 \mathrm{E}_x\{\mathbf{Q}\} \approx \frac{\sigma_v^2}{\sigma_x^2} \cdot \frac{1}{N} \cdot \left(1 + \frac{M + \kappa_x - 2}{N}\right)\mathbf{I}, \tag{4.36}$$

where κ_x is the Pearson kurtosis, which is defined as

$$\kappa_x \stackrel{\triangle}{=} \frac{\mathrm{E}_x\{|x(n)|^4\}}{(\mathrm{E}_x\{|x(n)|^2\})^2}. \tag{4.37}$$

For a complex Gaussian sequence $\kappa_x = 2$, for a complex constant modulus sequence $\kappa_x = 1$ and for a square M-QAM constellation $1 \leq \kappa_x \leq 1.4$. The estimation error due to measurement noise is thus evenly spread over all the estimated taps. The variance of the estimation error due to measurement noise is inversely proportional to the SNR and is approximately inversely proportional to the number of samples N used in the estimation.

The sum of variances of all parameter estimation errors, given as the trace of $\bar{\mathbf{R}}_{\mathrm{N}}$, can by using (4.36) be approximated as

$$\mathrm{tr}\bar{\mathbf{R}}_{\mathrm{N}} \approx \frac{\sigma_v^2}{\sigma_x^2} \cdot \frac{M}{N} \cdot \left(1 + \frac{M + \kappa_x - 2}{N}\right). \tag{4.38}$$

For large N, the kurtosis plays a minor role and when $N \gg M$, the variance of the noise induced error increases linearly with the number M of estimated coefficients (the number of taps).

4.3.2 Excess error

The weighted averaging of the channel in (4.30) can cause the LS-estimate to deviate from the true average of the channel even in the absence of measurement noise. To study this error we utilize a decomposition of the time varying channel into a sum of the average channel, $\bar{\mathbf{h}}_{[1\ N]}$, defined as

$$\bar{\mathbf{h}}_{[1\ N]} \triangleq \frac{1}{N}\sum_{n=1}^{N}\mathbf{h}(n), \tag{4.39}$$

and the time varying channel deviation from $\bar{\mathbf{h}}_{[1\ N]}$, $\boldsymbol{\vartheta}(n)$. Thus,

$$\mathbf{h}(n) = \bar{\mathbf{h}}_{[1\ N]} + \boldsymbol{\vartheta}(n). \tag{4.40}$$

Using (4.40), the LS solution (4.30) can be expressed as

$$\hat{\mathbf{h}}=\bar{\mathbf{h}}_{[1\ N]}+\mathbf{Q}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)\boldsymbol{\vartheta}(n) + \mathbf{e}_{\mathrm{N}}, \tag{4.41}$$

where the time invariant and the time varying terms are separated. We define the LS estimate's deviation from the average channel due to time variation, $\mathbf{e}_{\mathrm{E}}$, as the *excess error*

$$\mathbf{e}_{\mathrm{E}} = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)\boldsymbol{\vartheta}(n). \tag{4.42}$$

The LS estimate of the channel is thus the sum of the average channel in the interval, an excess error term due to the weighted averaging of the time varying channel and a noise term

$$\hat{\mathbf{h}} = \bar{\mathbf{h}}_{[1\ N]} + \mathbf{e}_{\mathrm{E}} + \mathbf{e}_{\mathrm{N}}. \tag{4.43}$$

Averaging over channel realizations

The covariance matrix for the excess error, $\mathbf{R}_{\mathrm{E}}$, is found by averaging over channel realizations

$$\begin{aligned}\mathbf{R}_{\mathrm{E}} &= \mathrm{E}_h\{\mathbf{e}_{\mathrm{E}}\mathbf{e}_{\mathrm{E}}^H\} & (4.44)\\ &= \mathrm{E}_h\left\{\mathbf{Q}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\mathbf{x}(n)\mathbf{x}^H(n)\boldsymbol{\vartheta}(n)\boldsymbol{\vartheta}^H(m)\mathbf{x}(m)\mathbf{x}^H(m)\right)\mathbf{Q}\right\}, & (4.45)\end{aligned}$$

where $\mathrm{E}_h\{\cdot\}$ is the averaging operator. The transmitted symbols are independent of the channel, so only $\boldsymbol{\vartheta}(n)\boldsymbol{\vartheta}^H(m)$ will be affected by the expectation. The cross-covariance matrix for the channel deviation from the average can be obtained as in Appendix 4.D, equation (4.84) and (4.85),

$$\mathbf{R}_\vartheta(n,m) = \mathrm{E}_h\{\boldsymbol{\vartheta}(n)\boldsymbol{\vartheta}^H(m)\} \approx f(n,m)\mathbf{R}_h, \tag{4.46}$$

where $\mathbf{R}_h$ is the channel correlation matrix and the scalar function $f(n,m)$ is defined as

$$f(n,m) \triangleq \frac{(t_s\omega_D)^2}{2}\left(n - \frac{N+1}{2}\right)\left(m - \frac{N+1}{2}\right). \tag{4.47}$$

The expression is obtained using a Taylor expansion of the channel (Appendix 4.B), combined with the covariance and cross-covariance matrices of the first and second derivatives of the channel for a *Rayleigh fading* channel with Jakes spectrum and maximal Doppler frequency ω_D (Appendix 4.C). The covariance matrix of the excess error (4.44) can by insertion of (4.46) and (4.47) be approximated as

$$\mathbf{R}_\mathrm{E} \approx \mathbf{Q}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(f(n,m)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{R}_h\mathbf{x}(m)\mathbf{x}^H(m)\right)\mathbf{Q}. \tag{4.48}$$

Thus, for any given data sequence $\{x(n)\}$, the covariance matrix of the LS excess error due to time variation of a Rayleigh fading channel with Jakes spectrum can be calculated using (4.48) and (4.47).

Averaging over data sequences

Given the distribution of the data sequences $\{x(n)\}$, an approximation for the average covariance matrix for an ensemble of training sequences

$$\bar{\mathbf{R}}_\mathrm{E} = \mathrm{E}_x\{\mathbf{R}_\mathrm{E}\} = \mathrm{E}_x\{\mathrm{E}_h\{\mathbf{e}_\mathrm{E}\mathbf{e}_\mathrm{E}^H\}\}, \tag{4.49}$$

can be calculated by some further manipulation of (4.48). The diagonal elements of $\bar{\mathbf{R}}_\mathrm{E}$ can be obtained as in Appendix 4.E, equation (4.100)

$$[\bar{\mathbf{R}}_\mathrm{E}]_{i,i} \approx \left[\mathrm{E}_x\left\{\mathbf{Q}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(f(n,m)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{R}_h\mathbf{x}(m)\mathbf{x}^H(m)\right)\mathbf{Q}\right\}\right]_{i,i} \tag{4.50}$$

$$\approx N\frac{(t_s\omega_D)^2}{24}\left((\kappa_x - 2)\sigma_{h_i}^2 + \sum_{k=0}^{M-1}\sigma_{h_k}^2\right), \tag{4.51}$$

where N is assumed to be large in relation to M. Here $\sigma^2_{h_k}$ denotes the variance of tap k. The variance of the excess error thus increases at the same rate as the number of samples used for the channel estimation. The kurtosis of the data sequence matters but does not dominate the behavior of the excess error variance. For a complex Gaussian sequence with $\kappa_x = 2$ all the taps would be corrupted by the same amount of excess error. For sequences with $\kappa_x < 2$, the strongest taps will actually be corrupted by slightly less excess error power than the smallest taps. As the Doppler frequency directly depends on the speed, the variance of the excess error increases with the square of the speed. A doubling of the vehicle velocity result in four times larger excess error variance.

The length of the estimation interval is $T = Nt_s$. To express the covariance as a function of the time-frequency product Tf_D, the Doppler frequency ω_D has to be expressed as $\omega_D = 2\pi f_D$. The time-frequency product Tf_D can be interpreted as the distance traveled during the estimation interval, measured in wavelengths. The sum of variances of the excess errors for all parameters, given as the trace of $\bar{\mathbf{R}}_{\mathrm{E}}$, using (4.51) with the substitution $t_s\omega_D = 2\pi Tf_D/N$, can be approximated as

$$\mathrm{tr}\bar{\mathbf{R}}_{\mathrm{E}} = \sum_{i=0}^{M-1} [\bar{\mathbf{R}}_{\mathrm{E}}]_{i,i} \approx \frac{\pi^2 (Tf_D)^2}{6N}(M+\kappa_x-2)\sum_{k=0}^{M-1}\sigma^2_{h_k}. \tag{4.52}$$

The variance of the excess error is, for a given N, proportional to the time-frequency product (the length of the estimation interval measured in wavelengths) to the power of two. For a given time-frequency product and large N, the variance is proportional to the inverse of N. The excess error is thus decreasing if the sampling interval, measured in wavelengths, decreases or if the number of samples within the interval increases.

4.3.3 Bias error

The average channel in the estimation interval, $\bar{\mathbf{h}}_{[1\ N]}$, deviates from the value in the middle of the interval if the trajectory of the channel deviates from a straight line. This deviation can be viewed as a *bias error* [15], that can be evaluated using the Taylor expansion of the average channel as in Appendix 4.B (4.71),

$$\mathbf{e}_{\mathrm{B}} = \bar{\mathbf{h}}_{[1\ N]} - \mathbf{h}(t_c)\big|_{t_c=\frac{N+1}{2}t_s} \approx -\frac{N^2-1}{24}t_s^2\,\frac{d^2\mathbf{h}(t_c)}{dt_c^2}\bigg|_{t_c=\frac{N+1}{2}t_s}. \tag{4.53}$$

The bias error thus depends on the curvature of the channel in the estimation interval. The LS estimate (4.30) also deviates from the value of the channel at the middle of the interval, which can be seen from (4.43) and (4.53)

$$\begin{aligned}\hat{\mathbf{h}} &= \bar{\mathbf{h}}_{[1\ N]} + \mathbf{e}_{\mathrm{E}} + \mathbf{e}_{\mathrm{N}} \\ &= \mathbf{h}(t_c)\big|_{t_c=\frac{N+1}{2}t_s} + \mathbf{e}_{\mathrm{B}} + \mathbf{e}_{\mathrm{E}} + \mathbf{e}_{\mathrm{N}},\end{aligned} \tag{4.54}$$

where we in the last approximation assumes $N \gg 1$. Averaging over realizations of the channel, the covariance matrix for $\mathbf{e}_{\mathrm{B}}$ is

$$\begin{aligned}\mathbf{R}_{\mathrm{B}} &= \mathrm{E}_h\{\mathbf{e}_{\mathrm{B}}\mathbf{e}_{\mathrm{B}}^H\} \approx \left(\frac{N^2-1}{24}t_s^2\right)^2 \mathrm{E}_h\left\{\frac{d^2\mathbf{h}(t_c)}{dt_c^2}\frac{d^2\mathbf{h}(t_c)}{dt_c^2}^H\right\}\Bigg|_{t_c=\frac{N+1}{2}t_s} \\ &= \frac{(N^2-1)^2}{1536}(t_s\omega_D)^4\mathbf{R}_h \approx \frac{\pi^4(Tf_D)^4}{96}\mathbf{R}_h.\end{aligned} \tag{4.55}$$

The variance of the bias error does thus solely depend on the length of the estimation interval and not on the statistics of the data sequence. The trace of $\mathbf{R}_{\mathrm{B}}$ can be approximated as

$$\mathrm{tr}\mathbf{R}_{\mathrm{B}} \approx \frac{\pi^4(Tf_D)^4}{96}\sum_{k=0}^{M-1}\sigma_{h_k}^2. \tag{4.56}$$

Note that the Taylor expansion (4.68) on which these expressions are based, is valid for no more than half a period of the fastest oscillation. Thus, the approximations are only valid for $Tf_D < 1/2$.

4.3.4 Total estimation error

The LS estimate of the impulse response, at the middle of the estimation interval, can now be expressed as

$$\hat{\mathbf{h}} = \mathbf{h}\left(\frac{N+1}{2}t_s\right) + \mathbf{e} \tag{4.57}$$

where the additive error term is approximated as

$$\mathbf{e} \approx \mathbf{e}_{\mathrm{N}} + \mathbf{e}_{\mathrm{E}} + \mathbf{e}_{\mathrm{B}}, \tag{4.58}$$

with the error terms given by (4.32), (4.42) and (4.53). The excess error and the bias error ($\mathbf{e}_{\mathrm{E}}$ and $\mathbf{e}_{\mathrm{B}}$) are approximately uncorrelated, since $\mathbf{e}_{\mathrm{E}}$ is related to $d\mathbf{h}(t_c)/dt_c$ and $\mathbf{e}_{\mathrm{B}}$ to $d\mathbf{h}^2(t_c)/d^2t_c$ and from Appendix 4.C, equation (4.77) we have that the expectation of the product between the first and second derivative of the channel is zero. Furthermore, under the assumption that the measurement noise is independent of the channel, $\mathbf{e}_\mathrm{N}$ is uncorrelated to both $\mathbf{e}_\mathrm{E}$ and $\mathbf{e}_\mathrm{B}$. The covariance matrix of the total estimation error can thus be modeled as

$$\mathbf{R}_e \approx \bar{\mathbf{R}}_\mathrm{N} + \bar{\mathbf{R}}_\mathrm{E} + \mathbf{R}_\mathrm{B}. \tag{4.59}$$

The sum of parameter error variance can be obtained as the trace of $\mathbf{R}_e$, as obtained from summing (4.38), (4.52) and (4.56), The normalized mean square error (NMSE) of the estimation error, obtained as $\mathrm{tr}\mathbf{R}_e$ normalized by the average gain of the channel, that is $\sum_{k=0}^{M-1} \sigma_{h_k}^2$, is a good measure for the estimation accuracy. Compiling the results we obtain

$$\begin{aligned}\frac{\mathrm{tr}\mathbf{R}_e}{\sum_{k=0}^{M-1} \sigma_{h_k}^2} &\approx \frac{\sigma_v^2}{\sigma_x^2 \sum_{k=0}^{M-1} \sigma_{h_k}^2} \cdot \frac{M}{N}\left(1 + \frac{M+\kappa_x-2}{N}\right) \\ &\quad + \frac{\pi^2 (Tf_D)^2}{6N}(M+\kappa_x-2) + \frac{\pi^4 (Tf_D)^4}{96}. \end{aligned} \tag{4.60}$$

The first term, the variance of the noise induced error, depends on the average SNR (4.6) and on the statistics of the transmitted sequence but not on the time-variation of the channel. The second term, the variance of the excess error, depends on the time variation of the channel and on the transmitted sequence but not on the transmit power, whereas the third term, the variance of the bias error, solely depend on the length of the estimation interval in relation to the time variation. If the length of the estimation interval is fixed, then T does not depend on the number of samples N used for the estimation. If, on the other hand, the sampling period t_s is fixed, then T in (4.60) should be exchanged with Nt_s, as the estimation interval then grows with N.

4.3.5 Simulation evaluations

Example 4.1 A GSM like channel

In our simulation example, we let each tap in an approximately Rayleigh fading channel with Jakes spectrum be simulated as a sum of 500 complex sinusoids with Gaussian distributed amplitude and frequency $f_D \cos(\theta_n)$, where θ_n is the angle of incidence, cf. equation (2.49) in Chapter 2. The angles are here assumed to be uniformly distributed between $[0\ 2\pi[$. The channel has four taps with exponentially decaying variances 1, 1/2, 1/4 and 1/8, respectively. The Doppler frequency is $f_D = 160$ Hz and the sampling period is $t_s = 5$ μs. The time-frequency product is thus $Tf_D = Nt_sf_D = N \cdot 8 \cdot 10^{-4}$. A white QPSK signal is transmitted over the time-varying channel and 10^5 samples are collected by the receiver. A measurement noise is added so that the received average SNR is 20 dB.

The channel is estimated block-wise with an LS estimator using different numbers of training symbols. The estimated channel is compared to the true channel and the error is calculated for the estimated taps. The variance of the error is estimated and summed for all the taps in the channel, to obtain $\mathrm{tr}\hat{\mathbf{R}}_e$ which then is normalized by $\sum_{k=0}^{M-1} \sigma_{h_k}^2$ to obtain the NMSE. The theoretical value for the normalized $\mathrm{tr}\mathbf{R}_e$ is obtained from (4.60). As seen in Figure 4.7, the theoretical approximation almost coincides with the result for the NMSE from the simulation for time-frequency products, Tf_D up to 1/2 (625 samples) and after that the NMSE is overestimated.

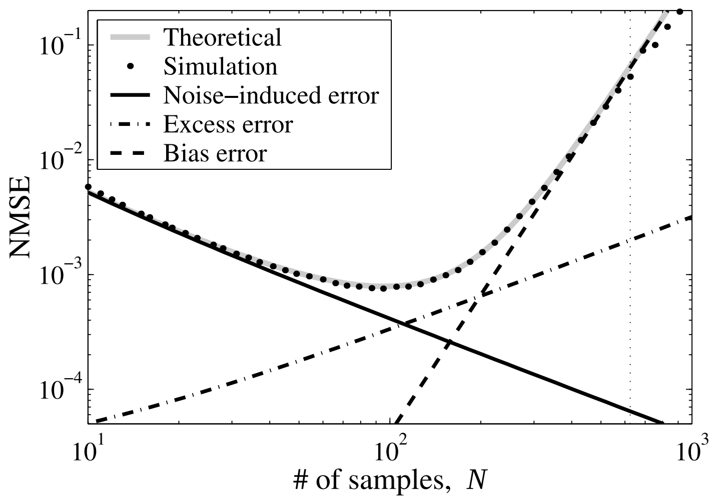

Figure 4.7: The variance of estimation error for the simulated channels is compared with the theoretical values obtained from (4.60). The contributions from (4.38), (4.52) and (4.56) are plotted separately. All variances are normalized by the total power of the channel. The vertical dotted line at N =625 marks that the estimation interval then is half a wavelength long.

This type of plot can be helpful to obtain the necessary amount of training data to achieve a certain channel NMSE at a specified maximum Doppler shift and signal SNR. To obtain a channel NMSE of less than 10^{-3} in this example, the number of training samples should be in the range of 53-153. If the SNR would increase from 20dB to 30dB then the line for the variance of the noise induced error would be drawn a factor 0.1 lower, with the same inclination, whereas the lines for the excess error and bias would remain as is.

Example 4.2 The appropriateness of the channel sampling rate for a fast sampled measured broadband channel

---

In the fast sampled measured channels of Chapter 3 the number of samples used for identification, and the number of estimated channel taps are fixed, but the time-frequency product varies due to different vehicle velocities. With the variance of the noise set to zero in (4.60) the expression can be used to obtain an idea of the contribution from the time variation to the estimation error in the measurements. The channel measurements in the data base have N =700, M =120 and the time-frequency products vary from about 0.002 to 0.02 wavelengths, depending on the speed of the receiver.[5] In Figure 4.8, the estimated channel estimation error NMSE for the measured channels are plotted together with the corresponding theoretical NMSE due to time variation, as functions of the time-frequency product. As can be seen in the plot the excess error dominates the theoretical NMSE. The estimated NMSE is generally much higher than the theoretical NMSE due to time variation in these measurements. It is thus the measurement noise that dominates the estimation error. The time variability due to the movement of the receiver during the estimation interval creates only a minor contribution to the total estimation error in these measurements.

---

[5] A velocity of 105 km/h result in a time-frequency product of $Tf_D = \frac{N}{f_s} \cdot \frac{f_c v}{c} = \frac{700}{6.4 \cdot 10^6} \cdot \frac{1880 \cdot 10^6}{3 \cdot 10^8 \cdot (105/3.6)}$ =0.02 wavelengths.

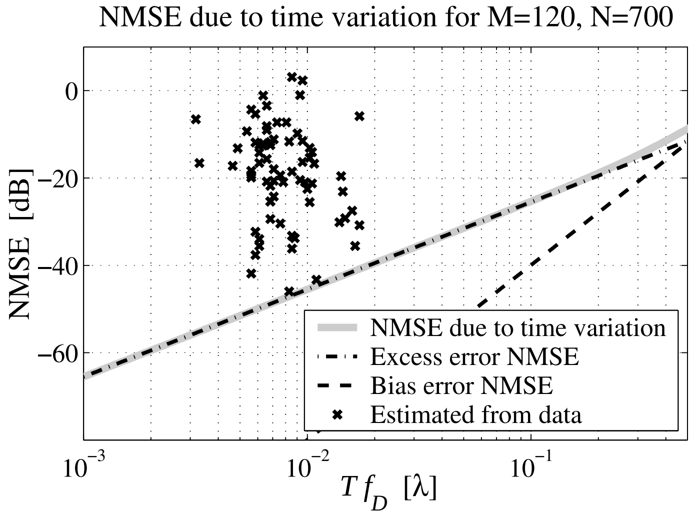

Figure 4.8: The NMSE for the estimation error due to time variation as a function of the time-frequency product. The number of samples used for estimation is N =700 and the number of estimated taps is M =120, just as for the fast sampled channel measurements in the data base. The NMSE for the the estimated channels are denoted by **x**. *The contributions from (4.52) and (4.56) are plotted separately. A time-frequency product of* $Tf_D = 10^{-2}$ *corresponds to a speed of 52.5km/h.*

4.4 Conclusion

- Block based LS-estimation is found to result in reasonable estimates of the impulse response for a measured time-varying broadband channel.

- The estimation error can be seen as a noise floor in the estimated power delay profile.

- The block based LS-estimation result in an estimate of the average channel impulse response in the estimation interval.

- In block based estimation of a time-varying impulse response, both the measurement noise and the time-variation contribute to the total estimation error.

- The estimation error due to measurement noise becomes proportional to the inverse of the number of samples used for identification.

- The error due to time-variation within the estimation interval can be parameterized using the time-frequency product Tf_D, which is the length of the estimation interval measured in wavelengths, and the number of symbols N used for identification.
- The variance of the estimation error due to weighted averaging of the time-varying channel becomes proportional to the time frequency product to the power of two and to the inverse of the number of samples used for identification.
- The average channel deviates from the true channel in the middle of the interval. This bias error is independent of the data sequence and the number of samples used for identification. The variance of the bias error is proportional to the time-frequency product to the power of four.

When the time-frequency product is small, the measurement noise will dominate the estimation error. Instead of increasing the estimation interval to average away measurement noise, the noise reduction can be performed on the estimated channel, using filters designed to fit the dynamics of the channel. The design and performance of such filters is the topic of Chapter 5.

4.A The Inverse of the Sample Covariance Matrix

In the expressions for the covariance matrices of the noise-induced error, (4.33), and the excess error, (4.44), the inverse of the sample covariance matrix plays an important role. We shall here derive a useful expression for this inverse. For a white data sequence, $\{x(n)\}$, the matrix product $\mathbf{X}^H\mathbf{X}$ can be decomposed as

$$\mathbf{X}^H\mathbf{X} = \sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}^H(n) = N\mathbf{R}_x + \sum_{n=1}^{N} \mathbf{Z}(n). \tag{4.61}$$

where

$$\mathbf{R}_x = \mathrm{E}_x\{x(n)x^H(n)\} = \sigma_x^2\mathbf{I} \tag{4.62}$$

is the covariance matrix for $x(n)$, and

$$\mathbf{Z}(n) = \mathbf{x}(n)\mathbf{x}^H(n) - \mathbf{R}_x \tag{4.63}$$

is the zero mean deviation from this covariance matrix. For a circular complex valued sequence, it holds [52]

$$\mathrm{E}_x\{\mathbf{Z}(n)\} = 0 \tag{4.64}$$

$$\mathrm{E}_x\{\mathbf{Z}(n)\mathbf{Z}(m)\} = \sigma_x^4(M + \kappa_x - 2)\mathbf{I}\delta_{n,m} \tag{4.65}$$

where κ_x is the Pearson kurtosis. To obtain an estimate of the inverse sample covariance matrix, we make a second order Taylor expansion around $\mathbf{I}$ as

$$\begin{aligned}\mathbf{Q} &\stackrel{\triangle}{=} \left(\mathbf{X}^H\mathbf{X}\right)^{-1} = \frac{1}{\sigma_x^2 N}\left(\mathbf{I} + \frac{1}{\sigma_x^2 N}\sum_{n=1}^{N}\mathbf{Z}(n)\right)^{-1} \\ &\approx \frac{1}{\sigma_x^2 N}\left(\mathbf{I} - \frac{1}{\sigma_x^2 N}\sum_{n=1}^{N}\mathbf{Z}(n) + \frac{1}{\sigma_x^4 N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\mathbf{Z}(n)\mathbf{Z}(m)\right).\end{aligned} \tag{4.66}$$

The expected value of (4.66), using (4.64) and (4.65), yields

$$\begin{aligned}\mathrm{E}_x\{\mathbf{Q}\} = \mathrm{E}_x\left\{\left(\mathbf{X}^H\mathbf{X}\right)^{-1}\right\} &\approx \frac{1}{\sigma_x^2 N}\left(\mathbf{I} + \frac{M + \kappa_x - 2}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\mathbf{I}\delta_{n,m}\right) \\ &= \frac{1}{\sigma_x^2 N}\left(1 + \frac{M + \kappa_x - 2}{N}\right)\mathbf{I}.\end{aligned} \tag{4.67}$$

4.B Taylor Expansion of the Channel

To obtain expressions for the average channel and the deviation, we perform a second order Taylor expansion of the continuous channel, $\mathbf{h}(t_c)$, around the middle of the estimation interval

$$\mathbf{h}(n) = \mathbf{h}(t_c)|_{t_c=nt_s} \approx \mathbf{h}(t_c)|_{t_c=\frac{N+1}{2}t_s} + \left(n - \frac{N+1}{2}\right) t_s \frac{d\mathbf{h}(t_c)}{dt_c}\Big|_{t_c=\frac{N+1}{2}t_s} + \left(n - \frac{N+1}{2}\right)^2 \frac{t_s^2}{2} \frac{d^2\mathbf{h}(t_c)}{dt_c^2}\Big|_{t_c=\frac{N+1}{2}t_s}, \tag{4.68}$$

where t_s denotes the sampling period. For oscillating channels, this Taylor expansion can be used for an estimation interval not longer than half a period of the fastest frequency component just as for sinusoids, as for longer intervals the second derivative is likely to change sign within the interval and a third term would be needed to accommodate this behavior.

To evaluate the average channel in the estimation interval, we need the sums,

$$\sum_{n=1}^{N} \left(n - \frac{N+1}{2}\right) = 0 \tag{4.69}$$

$$\sum_{n=1}^{N} \left(n - \frac{N+1}{2}\right)^2 = \frac{N(N^2-1)}{12}. \tag{4.70}$$

Using (4.68) and the sums (4.69), (4.70), the average channel can be expressed as

$$\bar{\mathbf{h}}_{[1\ N]} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{h}(n) \approx \mathbf{h}(t_c)_{t_c=\frac{N+1}{2}t_s} + \frac{N^2-1}{24} t_s^2 \frac{d^2\mathbf{h}(t_c)}{dt_c^2}\Big|_{t_c=\frac{N+1}{2}t_s} \tag{4.71}$$

The deviation $\boldsymbol{\vartheta}(n)$ can be obtained from (4.68) and (4.71) as

$$\boldsymbol{\vartheta}(n) = \mathbf{h}(n) - \bar{\mathbf{h}}_{[1\ N]} \approx \left(n - \frac{N+1}{2}\right) t_s \frac{d\mathbf{h}(t_c)}{dt_c}\Big|_{t_c=\frac{N+1}{2}t_s} + \left[\left(n - \frac{N+1}{2}\right)^2 - \frac{N^2-1}{12}\right] \frac{t_s^2}{2} \frac{d^2\mathbf{h}(t_c)}{dt_c^2}\Big|_{t_c=\frac{N+1}{2}t_s}. \tag{4.72}$$

As the $\boldsymbol{\vartheta}(n)$ is the deviation from the average channel in the estimation interval, we have that $\sum_{n=1}^{N} \boldsymbol{\vartheta}(n) = 0$. This expression for $\boldsymbol{\vartheta}(n)$ will be used to obtain approximate expressions for the covariance matrix for the excess error.

4.C Correlation for the Derivatives of the Channel

To compute $\mathbf{R}_{\mathrm{E}}$ in (4.44) and $\mathbf{R}_{\mathrm{B}}$ in (4.55), the covariance and cross-covariance matrices of the first and second derivatives of the channel are needed. For notational convenience $\mathbf{h}(t)$ denotes a continuous channel at time t in this section, as only continuous channels are considered in this section. For a Rayleigh fading channel the autocorrelation matrix is given by [11]

$$\mathrm{E}_h\{\mathbf{h}(t)\mathbf{h}^H(t+\tau)\} = \mathbf{R}_h \mathrm{J}_0(\omega_D\tau), \tag{4.73}$$

where the expectation is taken over realizations of the channel, not over time, and ω_D is the maximal Doppler frequency (in rad/s). The zero order Bessel function of the first kind $\mathrm{J}_0(\cdot)$, has the series expansion [53]

$$\mathrm{J}_0(\omega_D\tau) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(k!)^2}\left(\frac{\omega_D\tau}{2}\right)^{2k}. \tag{4.74}$$

Using (4.73) and (4.74), the covariance matrices of the derivatives of the channel can be derived.

Covariance for the first derivative

The covariance matrix of the first derivative is

$$\begin{aligned}\mathrm{E}_h\left\{\frac{d\mathbf{h}(t)}{dt}\frac{d\mathbf{h}(t)}{dt}^H\right\} &= \mathrm{E}_h\left\{\lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}} \frac{\mathbf{h}(t+\Delta t)-\mathbf{h}(t)}{\Delta t}\cdot\frac{\mathbf{h}^H(t+\Delta\tau)-\mathbf{h}^H(t)}{\Delta\tau}\right\}\\ &= \lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}} \frac{\mathbf{R}_h}{\Delta t\Delta\tau}(\mathrm{J}_0(\omega_D(\Delta t-\Delta\tau)) - \mathrm{J}_0(\omega_D\Delta t) - \mathrm{J}_0(\omega_D\Delta\tau)+1).\end{aligned} \tag{4.75}$$

To evaluate (4.75) only the first two terms in the series expansion (4.74) of $\mathrm{J}_0(\cdot)$ are needed, as higher orders will cancel or have terms that will approach zero in the limit.

$$\begin{aligned}&\mathrm{E}_h\left\{\frac{d\mathbf{h}(t)}{dt}\frac{d\mathbf{h}(t)}{dt}^H\right\} =\\ &\lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}} \frac{\mathbf{R}_h}{\Delta t\Delta\tau}\left(1-\frac{\omega_D^2(\Delta t-\Delta\tau)^2}{4}-1+\frac{\omega_D^2\Delta t^2}{4}-1+\frac{\omega_D^2\Delta\tau^2}{4}+1\right) = \frac{\omega_D^2}{2}\mathbf{R}_h.\end{aligned} \tag{4.76}$$

The correlation for the derivative of the channel is thus equal to the correlation of the channel times a factor proportional to the square of the maximal Doppler frequency.

Cross-covariance for the first and second derivative

The cross-covariance between the first and second derivative of the channel is

$$\mathrm{E}_h\left\{\frac{d\mathbf{h}(t)}{dt}\frac{d^2\mathbf{h}(t)}{dt^2}^H\right\}=$$

$$\mathrm{E}_h\left\{\lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}}\frac{\mathbf{h}(t+\Delta t)-\mathbf{h}(t-\Delta t)}{2\Delta t}\cdot\frac{\mathbf{h}^H(t+\Delta\tau)-2\mathbf{h}^H(t)+\mathbf{h}^H(t-\Delta\tau)}{\Delta\tau^2}\right\}$$
$$=\lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}}\frac{\mathbf{R}_h}{\Delta t\Delta\tau^2}\{\mathrm{J}_0(\omega_D(\Delta t-\Delta\tau))-2\mathrm{J}_0(\omega_D\Delta t)+\mathrm{J}_0(\omega_D(\Delta t+\Delta\tau))$$
$$-\mathrm{J}_0(\omega_D(\Delta t-\Delta\tau))+2\mathrm{J}_0(\omega_D\Delta t)-\mathrm{J}_0(\omega_D(\Delta t+\Delta\tau))\}=0. \quad (4.77)$$

The first and second derivative of the channel at time t are thus uncorrelated.

Covariance for the second derivative

The covariance matrix for the second derivative is

$$\mathrm{E}_h\left\{\frac{d^2\mathbf{h}(t)}{dt^2}\frac{d^2\mathbf{h}(t)}{dt^2}^H\right\}=$$

$$\mathrm{E}_h\left\{\lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}}\frac{\mathbf{h}^H(t+\Delta t)-2\mathbf{h}^H(t)+\mathbf{h}^H(t-\Delta t)}{\Delta t^2}\cdot\frac{\mathbf{h}^H(t+\Delta\tau)-2\mathbf{h}^H(t)+\mathbf{h}^H(t-\Delta\tau)}{\Delta\tau^2}\right\}$$
$$=\lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}}\frac{\mathbf{R}_h}{\Delta t^2\Delta\tau^2}\{4+2\mathrm{J}_0(\omega_D(\Delta t+\Delta\tau))+2\mathrm{J}_0(\omega_D(\Delta t-\Delta\tau))$$
$$-4\mathrm{J}_0(\omega_D\Delta t)-4\mathrm{J}_0(\omega_D\Delta\tau))\}. \quad (4.78)$$

To evaluate (4.78) terms up to the fourth order from the series expansion (4.74) of $\mathrm{J}_0(\cdot)$ is needed ($k=0,1,2$), higher order terms cancel or approach zero in the limit. The covariance for the the second derivative of the channel is then obtained as

$$\mathrm{E}_h\left\{\frac{d^2\mathbf{h}(t)}{dt^2}\frac{d^2\mathbf{h}(t)}{dt^2}^H\right\} = \tag{4.79}$$

$$\lim_{\substack{\Delta t\to 0\\ \Delta\tau\to 0}} \frac{\mathbf{R}_h}{\Delta t^2\Delta\tau^2}\left\{4+2\left(1-\frac{\omega_D^2(\Delta t+\Delta\tau)^2}{4}+\frac{\omega_D^4(\Delta t+\Delta\tau)^4}{64}\right)\right.$$
$$+2\left(1-\frac{\omega_D^2(\Delta t-\Delta\tau)^2}{4}+\frac{\omega_D^4(\Delta t-\Delta\tau)^4}{64}\right)$$
$$\left.-4\left(2-\frac{\omega_D^2(\Delta t^2+\Delta\tau^2)}{4}+\frac{\omega_D^4(\Delta t^4+\Delta\tau^4)}{64}\right)\right\}$$
$$=\frac{3\omega_D^4}{8}\mathbf{R}_h \tag{4.80}$$

4.D The Cross-covariance Matrix for the Deviation from the Average

The cross-covariance matrix for the deviation from the average can be approximated using the Taylor series (4.72)

$$\mathbf{R}_\vartheta(n,m) = \mathrm{E}_h\left\{\boldsymbol{\vartheta}(n)\boldsymbol{\vartheta}^H(m)\right\}$$
$$\approx\left(n-\frac{N+1}{2}\right)\left(m-\frac{N+1}{2}\right)t_s^2\,\mathrm{E}_h\left\{\frac{d\mathbf{h}(t_c)}{dt_c}\frac{d\mathbf{h}(t_c)}{dt_c}^H\right\}\Bigg|_{t_c=\frac{N+1}{2}t_s}$$
$$+\left(n-\frac{N+1}{2}\right)\left[\left(m-\frac{N+1}{2}\right)^2-\frac{N^2-1}{12}\right]\frac{t_s^3}{2}\,\mathrm{E}_h\left\{\frac{d\mathbf{h}(t_c)}{dt_c}\frac{d^2\mathbf{h}(t_c)}{dt_c^2}^H\right\}\Bigg|_{t_c=\frac{N+1}{2}t_s}$$
$$+\left(m-\frac{N+1}{2}\right)\left[\left(n-\frac{N+1}{2}\right)^2-\frac{N^2-1}{12}\right]\frac{t_s^3}{2}\,\mathrm{E}_h\left\{\frac{d^2\mathbf{h}(t_c)}{dt_c^2}\frac{d\mathbf{h}(t_c)}{dt_c}^H\right\}\Bigg|_{t_c=\frac{N+1}{2}t_s}$$
$$+\left[\left(n-\frac{N+1}{2}\right)^2-\frac{N^2-1}{12}\right]\left[\left(m-\frac{N+1}{2}\right)^2-\frac{N^2-1}{12}\right]\frac{t_s^4}{4}\times$$
$$\mathrm{E}_h\left\{\frac{d^2\mathbf{h}(t_c)}{dt_c^2}\frac{d^2\mathbf{h}(t_c)}{dt_c^2}^H\right\}\Bigg|_{t_c=\frac{N+1}{2}t_s}. \tag{4.81}$$

The expectations of the derivatives in the expression above are evaluated for the Jakes channel in Appendix 4.C. For the Jakes channel we thus obtain

$$\begin{aligned}\mathbf{R}_\vartheta(n,m) \approx & \left(n-\frac{N+1}{2}\right)\left(m-\frac{N+1}{2}\right)\frac{(t_s\omega_D)^2}{2}\mathbf{R}_h \\ & + \left[\left(n-\frac{N+1}{2}\right)^2-\frac{N^2-1}{12}\right]\left[\left(m-\frac{N+1}{2}\right)^2-\frac{N^2-1}{12}\right]\frac{3(t_s\omega_D)^4}{32}\mathbf{R}_h.\end{aligned} \tag{4.82}$$

This can also be expressed using the time frequency product Tf_D, where $T = Nt_s$ is the length of the estimation interval and $f_D = \omega_D/2\pi$ is the Doppler frequency in Hz. Substituting $t_s\omega_D$ in (4.82) with $2\pi Tf_D/N$, we obtain

$$\begin{aligned}\mathbf{R}_\vartheta(n,m) \approx & \underbrace{\left(1-\frac{2n-1}{N}\right)}_{-1<(\cdot)<1}\left(1-\frac{2m-1}{N}\right)\frac{\pi^2(Tf_D)^2}{2}\mathbf{R}_h+ \\ & \underbrace{\left(\frac{2n(n-(N+1))}{N^2}+\frac{N^2+3N+2}{3N^2}\right)}_{-\frac{1}{3}<(\cdot)<\frac{2}{3}}\left(\frac{2m(m-(N+1))}{N^2}+\frac{N^2+3N+2}{3N^2}\right)\frac{3\pi^4(Tf_D)^4}{8}\mathbf{R}_h.\end{aligned} \tag{4.83}$$

To obtain reasonable results in the channel estimation the time-frequency product has to be small, that is $Tf_D \ll 1$. Under that condition we see that the second term in (4.83) can be dropped, as the expressions within parentheses are limited and the time-frequency product to the power of four is small compared to the time-frequency to the power of two, in the relevant region. It is thus sufficient to only consider the first derivative of the channel, that is the linear change, in the approximate expression for the cross-covariance for the channel deviation from the average.

In (4.82) we see that the cross-covariance for the time-varying channel deviation from the average can be approximated as a scalar function of the indices n and m, times the correlation matrix for the channel

$$\mathbf{R}_\vartheta(n,m) = f(n,m)\mathbf{R}_h. \tag{4.84}$$

When only the linear change of the channel is considered $f(n,m)$ is obtained from (4.82) as

$$f(n,m) = \left(n-\frac{N+1}{2}\right)\left(m-\frac{N+1}{2}\right)\frac{(t_s\omega_D)^2}{2}. \tag{4.85}$$

with the property that

$$\sum_{n=1}^{N}\sum_{m=1}^{N} f(n,m)g(n) = \sum_{m=1}^{N}\sum_{m=1}^{N} f(n,m)g(m) = 0, \tag{4.86}$$

where $g(\cdot)$ may be any function, but when $n = m$

$$\sum_{n=1}^{N}\sum_{m=1}^{N} f(n,m)\delta_{n,m} = \sum_{n=1}^{N} f(n,n) = \frac{(t_s\omega_D)^2}{2}\frac{N(N^2-1)}{12}, \tag{4.87}$$

where we have used (4.70).

4.E Averaging over Data Sequences

The covariance matrix, after averaging over channel realizations, for the excess error can be approximated as (4.48)

$$\mathbf{R}_{\mathrm{E}} \approx \mathbf{Q}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(f(n,m)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{R}_h\mathbf{x}(m)\mathbf{x}^H(m)\right)\mathbf{Q}. \tag{4.88}$$

To obtain the average over data sequences too, we substitute $x(n)x^H(n) = \mathbf{Z}(n) + \sigma_x^2\mathbf{I}$ and use the approximation (4.66) for $\mathbf{Q}$ in (4.88)

$$\begin{aligned}\mathbf{R}_{\mathrm{E}} \approx & \frac{1}{\sigma_x^2 N}\left(\mathbf{I} - \frac{1}{\sigma_x^2 N}\sum_{n=1}^{N}\mathbf{Z}(n) + \frac{1}{N^2\sigma_x^4}\sum_{n=1}^{N}\sum_{m=1}^{N}\mathbf{Z}(n)\mathbf{Z}(m)\right)\times \\ & \sum_{k=1}^{N}\sum_{l=1}^{N}\left(f(k,l)\left(\mathbf{Z}(k)+\sigma_x^2\mathbf{I}\right)\mathbf{R}_h\left(\mathbf{Z}(l)+\sigma_x^2\mathbf{I}\right)\right)\times \\ & \frac{1}{\sigma_x^2 N}\left(\mathbf{I} - \frac{1}{\sigma_x^2 N}\sum_{i=1}^{N}\mathbf{Z}(i) + \frac{1}{\sigma_x^4 N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{Z}(i)\mathbf{Z}(j)\right).\end{aligned} \tag{4.89}$$

When taking the expectation over circular complex data sequences the first order moment of $\mathbf{Z}(n)$ is zero (4.64). The second order terms will be kept but third order, and higher terms, will be omitted as they will be comparably small. As $\sum_k\sum_l f(k,l)g(l) = \sum_k\sum_l f(k,l)g(k) = 0$, from (4.86), only terms containing both $\mathbf{Z}(k)$ and $\mathbf{Z}(l)$ will give a contribution. As third and forth order terms are omitted, only one term will contribute to the approximate expression for the covariance matrix. The covariance matrix for the excess error, averaged over data sequences, can thus be obtained as

$$\begin{aligned} \mathrm{E}_x\{\mathbf{R}_\mathrm{E}\} &\approx \frac{1}{\sigma_x^4 N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} f(n,m) \mathrm{E}_x\{\mathbf{Z}(n)\mathbf{R}_h\mathbf{Z}(m)\} && (4.90) \\ &= \frac{1}{\sigma_x^4 N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} f(n,m) \mathrm{E}_x\{x(n)x^H(n)\mathbf{R}_h x(m) x^H(m)\}, && (4.91) \end{aligned}$$

where we in the last equality exploited (4.86). To evaluate this expression further we define the matrix $\boldsymbol{\Xi}(n,m)$ as

$$\boldsymbol{\Xi}(n,m) = \mathbf{x}(n)\mathbf{x}^H(n)\mathbf{R}_h\mathbf{x}(m)\mathbf{x}^H(m). \tag{4.92}$$

Note that the three terms in the middle form a scalar

$$\mathbf{x}^H(n)\mathbf{R}_h\mathbf{x}(m) = \sum_{k=1}^{M}\sum_{l=1}^{M} r_{k,l}\, x(n-k+1)x^*(m-l+1), \tag{4.93}$$

where $r_{k,l} = [\mathbf{R}_h]_{k,l}$ is the k,l:th element of the channel covariance matrix. As this part is scalar, it can be moved to the beginning of the product. The matrix can thus be expressed as

$$\boldsymbol{\Xi}(n,m) = \sum_{k=1}^{M}\sum_{l=1}^{M} r_{k,l}\, x(n-k+1)x^*(m-l+1)\mathbf{x}(n)\mathbf{x}^H(m). \tag{4.94}$$

The i,j:th element of the product $\mathbf{x}(n)\mathbf{x}^H(m)$ is

$$[\mathbf{x}(n)\mathbf{x}^H(m)]_{i,j} = x^*(n-i+1)x(m-j+1). \tag{4.95}$$

The average E_x over training sequences of the i,j:th element of $\boldsymbol{\Xi}(n,m)$ is [52]

$$\begin{aligned} &\mathrm{E}_x\{[\boldsymbol{\Xi}(n,m)]_{i,j}\} \\ &\quad = \mathrm{E}_x\left\{ \sum_{k=1}^{M}\sum_{l=1}^{M} r_{k,l}\, x(n-k+1)x^*(m-l+1)x^*(n-i+1)x(m-j+1) \right\} \\ &\quad = \sigma_x^4 \sum_{k=1}^{M}\sum_{l=1}^{M} r_{k,l} \left(\delta_{i,k}\delta_{j,l} + \delta_{n-i,m-j}\delta_{n-k,m-l} + (\kappa_x - 2)\delta_{i,k}\delta_{j,l}\delta_{n-i,m-j} \right). \end{aligned} \tag{4.96}$$

For a channel with uncorrelated taps, that is $r_{k,l} = r_{k,k}\delta_{k,l}$, this result in a diagonal matrix. An approximate expression for the variance of the excess error on the individual taps can be derived for a general channel covariance matrix. To find the variance of the excess error on the individual taps we need only to calculate the diagonal elements of the covariance matrix $\bar{\mathbf{R}}_{\mathrm{E}}$ and thus only $\mathrm{E}_x\{[\mathbf{\Xi}(n,m)]_{i,j}\}$ for $i = j$, which according to (4.96) renders

$$\mathrm{E}_x\{[\mathbf{\Xi}(n,m)]_{i,i}\} = \sigma_x^4 \left(r_{i,i} + \delta_{n,m}(\kappa_x - 2) r_{i,i} + \delta_{n,m} \sum_{k=1}^{M}\sum_{l=1}^{M} r_{k,l}\delta_{n-k,m-l} \right). \tag{4.97}$$

The expectation E_x of the summation over n, m in (4.48) can now by (4.91), (4.96), (4.97) be calculated as

$$\mathrm{E}_x \left\{ \sum_{n=1}^{N}\sum_{m=1}^{N} f(n,m)[\mathbf{\Xi}(n,m)]_{i,i} \right\} = \sigma_x^4 N(N^2-1)\frac{(t_s\omega_D)^2}{24}\left((\kappa_x - 2)\sigma_{h_i}^2 + \sum_{k=1}^{M} \sigma_{h_k}^2 \right). \tag{4.98}$$

where $\sigma_{h_k}^2 = r_{k,k}$ is the variance of tap k. In (4.98) we have exploited that the sum over $f(n,m)$ is zero unless $n = m$, when it is obtained as in (4.87). Only the terms multiplied with $\delta_{n,m}$ in (4.97) will thus contribute to the sum. The diagonal elements of the covariance for the excess error averaged over data sequences can thus by inserting (4.98) in (4.91) be approximated as

$$[\bar{\mathbf{R}}_{\mathrm{E}}]_{i,i} = [\mathrm{E}_x\{\mathbf{R}_{\mathrm{E}}\}]_{i,i} \approx \frac{1}{\sigma_x^4 N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} f(n,m)\mathrm{E}_x\{[\mathbf{\Xi}(n,m)]_{i,i}\} \tag{4.99}$$

$$\approx N\frac{(t_s\omega_D)^2}{24}\left(\sum_{k=1}^{M}\sigma_{h_k}^2 + (\kappa_x - 2)\sigma_{h_i}^2 \right), \tag{4.100}$$

where in the last approximation $N^2 \approx N^2 - 1$.

Chapter 5

Noise Reduction of Estimated Channels

5.1 Introduction

When considering consecutive snapshots of the channel, each tap of an estimated channel impulse response forms a time series, sampled at the channel sampling rate. As outlined in Section 4.2.4, this time series will be corrupted by noise. From the models of the channel presented in Chapter 2, we know that the frequency content of a tap of a channel will be approximately band-limited, whereas the channel estimation error will be close to white[1].

In Section 4.3.1 we saw that a longer estimation interval reduce the estimation error due to measurement noise. This averaging over the channel can be regarded as a low-pass filter with the impulse response of a rectangular window. Such a filter does not exploit the properties of either the noise or the signal. To obtain efficient noise reduction that exploits the dynamics of the channel, it is beneficial to apply Wiener-smoothers to the time series of sampled channel. Especially when the channel is highly oversampled in relation to the Doppler frequency, simple linear smoothers can be designed to reduce the influence of the estimation error significantly. This is highly beneficial for the prediction performance, especially when sub-sampled predictors are used.

An ideal low-pass filter cutting off all frequencies outside $[-f_D \; f_D]$ would

[1]When LS-estimation of the channel is used, the measurement noise depending estimation error will be independent from snapshot to snapshot. The estimation due to the time variation of the channel will be correlated over different channel snapshots but these contributions will be small with the choices of estimation intervals used here.

give a large noise reduction [16] but it would not include the part of the tap energy outside the Doppler band. It would also be hard to implement without long delays. The best noise reduction should optimize amplification with regard to the signal power and the noise power at each frequency and still be able to retain a reasonably low delay. In the following we will present and evaluate such methods.

We will study the Wiener-design of both IIR and FIR-smoothers. The IIR-smoother is designed based on an ARMA-model for the dynamics of the tap and offers a compact smoother with few coefficients. The FIR-smoother is designed based on a model for the correlation of the tap and need many more coefficients than the IIR-smoother to achieve the same performance, but it is not as numericly sensitive. Both smoothers need an estimate of the SNR which can be obtained either from the power delay profile or from the Doppler spectrum.

5.2 Estimated SNR for a Tap

Under the assumption that the average over time of the tap $h_k(t)$ and of the corresponding estimation error $e_k(t)$ are both zero, an estimate of the variance of the estimated tap $\hat{h}_k(t) = h_k(t) + e_k(t)$ can be obtained as the magnitude of the k:th tap in the PDP, as was demonstrated in Section 4.2.4. The time index t denotes the discrete time index for the snapshots of the channel.

All the taps of an LS-estimated snapshot of the channel impulse response will be corrupted by roughly the same amount of noise due to the estimation error, under the assumption that the training sequence is approximately white. The variance of $e_k(t)$ is thus approximately the same for all taps and is in the following denoted σ_e^2.

The variance for the estimation error, that is the power, is indicated by the noise floor in the PDP, as seen in Figure 5.1. The same noise floor as in the PDP can be observed in the average Doppler spectrum depicted in Figure 5.2.[2] It is the average over the Doppler spectra for all the taps, for the same measurement as for the PDP in Figure 5.1. Outside a narrow frequency band where most of the power is located, the main contribution to the power comes from noise. At these frequencies the average Doppler spectrum is close to the level of the noise floor observed in the PDP. As the

[2] The Doppler spectrum is estimated using a Capon method, (MASC) [54], to avoid the high side-lobes and low resolution in ordinary Fourier transform estimation of the spectrum.

noise floor has a flat spectrum, the channel estimation error is white.

The channel can for most practical purposes be considered as band limited to $[-f_D\ f_D]$, as we have no way of telling what happens under the noise floor. Even though we can not estimate the noise level inside the band-limits, it is reasonable to extrapolate the noise floor as flat over this interval too.

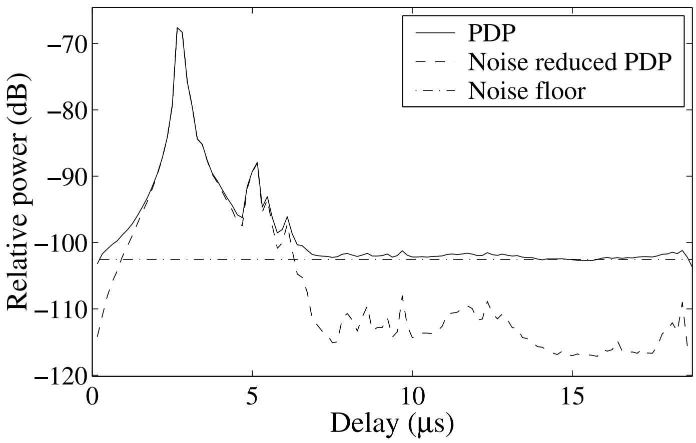

Figure 5.1: Power delay profile (PDP) and noise floor for one measurement location. The noise reduced PDP is estimated from the taps, after the proposed noise reduction.

With an estimate $\hat{\sigma}_e^2$ of the variance for the estimation error obtained from the noise floor in either the PDP or the average Doppler spectrum, the average SNR $\bar{\gamma}_k$ for the k:th tap can be estimated. Under the assumption that $h_k(t)$ and $e_k(t)$ are uncorrelated, the variance for the estimated tap is $\sigma^2_{\hat{h}_k} = \sigma^2_{h_k} + \sigma^2_e$ and the SNR can thus be obtained as

$$\bar{\gamma}_k = \frac{\sigma^2_{h_k}}{\sigma^2_e} \approx \frac{\hat{\sigma}^2_{\hat{h}_k} - \hat{\sigma}^2_e}{\hat{\sigma}^2_e}. \tag{5.1}$$

In practise the estimated variance $\hat{\sigma}^2_{\hat{h}_k}$ for the observed tap may be smaller than the estimated noise floor $\hat{\sigma}^2_e$ for some small taps. To avoid this problem we limit the SNR downwards by setting

$$\hat{\gamma}_k = \frac{\max(\hat{\sigma}^2_{\hat{h}_k} - \hat{\sigma}^2_e, \epsilon\hat{\sigma}^2_e)}{\hat{\sigma}^2_e}, \tag{5.2}$$

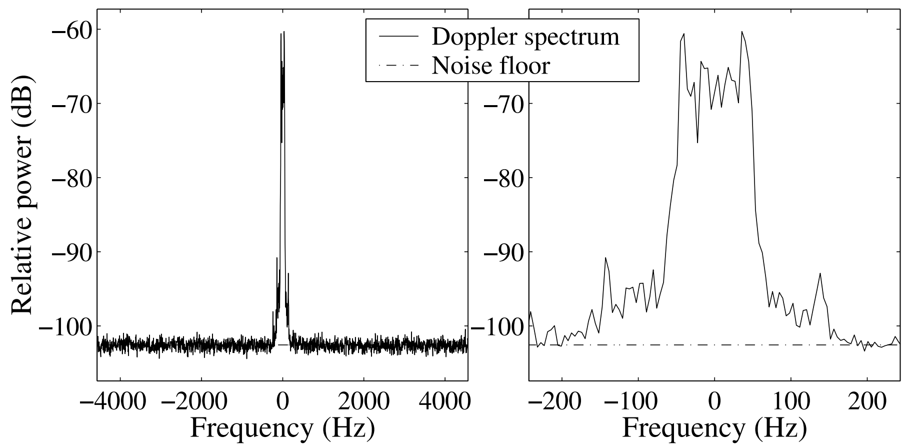

Figure 5.2: Average Doppler spectrum (zoomed to the right) and noise floor for the same measurement as used in Figure 5.1.

where ϵ is a small positive number. The estimated SNR will then be limited downwards by $\min \hat{\gamma}_k = \epsilon$.

The Doppler frequency

An estimate of the Doppler frequency can also be obtained from the average Doppler spectrum. The frequency region of interest $[-f_D \ f_D]$, is simply taken as the region where most of the power above the noise floor is concentrated. This will be sufficiently accurate for use in the smoother design.

5.3 IIR-smoothers

To be able to design a smoothing filter that suppresses the noise and lets the signal through with as little distortion as possible, we need to know how the power of the signal is distributed in the frequency domain. By using low order ARMA-models for the channel dynamics, a robust smoother can be derived [27],[55].

In Section 2.5 we proposed to model a channel tap, $h_k(t)$, as an ARMA process,

$$h_k(t) = \frac{C(q^{-1})}{D(q^{-1})} w_k(t). \tag{5.3}$$

Here the time index t denotes samples taken at the channel sampling rate and the driving noise (innovation) $w_k(t)$ is assumed to be white with zero mean. When the channel is highly oversampled, most of its power will be located at low frequencies around zero and the ARMA-model will thus have a low-pass characteristic. Even though a low order ARMA-model cannot reproduce the fine structure of the power-spectrum of an oversampled tap, the filter is still sufficient for indicating what levels of signal power that can be expected at different frequencies, on average. Thus, the frequency response of the low-pass filter $C(q^{-1})D^{-1}(q^{-1})$ scaled by the variance of the driving noise $w_k(t)$ serves as a model for the Doppler spectrum of the tap $h_k(t)$. Given the filter $C(q^{-1})D^{-1}(q^{-1})$ and the variance of the tap $h_k(t)$, the variance of the driving noise $w_k(t)$ can be calculated.

An additive estimation error, $e_k(t)$, corrupts the tap. The Wiener fixed-lag smoother, represented as a stable IIR-filter

$$\tilde{h}_k(t-m|t) = \frac{Q_0(q^{-1})}{R_0(q^{-1})}(h_k(t) + e_k(t)), \tag{5.4}$$

is designed to minimize $\mathrm{E}\{|\tilde{h}_k(t|t+m) - h_k(t)|^2\}$ at a given SNR [56]. As the Doppler spectrum is flat outside the Doppler frequency, the noise, $e_k(t)$ is modeled as white. The estimated level of the noise floor is used as the variance of the measurement noise $e_k(t)$ in (5.4). The procedure results in an estimate of the tap, delayed with the smoothing lag m, see Figure 5.3.

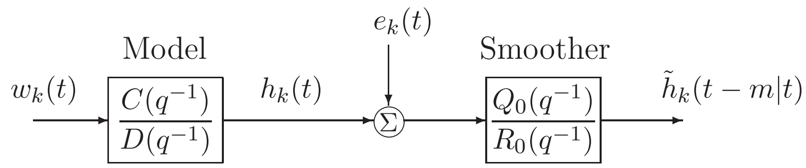

Figure 5.3: The model for smoothing design for noise reduction of a tap.

If the model $C(q^{-1})D^{-1}(q^{-1})$ is a low-pass filter with cut-off frequency at f_D and the noise is white, then the smoother $Q_0(q^{-1})R_0^{-1}(q^{-1})$ will be a low-pass filter that has an approximately flat frequency response with amplification one, and a linear phase in the passband $[-f_D \ f_D]$. Thus all frequencies within the passband will be left unchanged whereas power outside the passband will be attenuated. When the design model $C(q^{-1})D^{-1}(q^{-1})$ is ARMA(n_d, n_c), the smoother is ARMA$(\max(n_c, n_d), n_d + m)$. The higher the smoothing lag, the more coefficients are required in the moving average part, $Q_0(q^{-1})$, of the smoother.

Example 5.1 Noise reduction on a measured tap

The tap corresponding to the second peak of the PDP in Figure 5.1 has an estimated SNR of 14 dB and is approximately band-limited to ±63 Hz. As the channel sampling rate for this measurement is 9.14 kHz, the channel has an over sampling ratio (OSR) of 73. As a model filter $C(q^{-1})D^{-1}(q^{-1})$ in (5.3), for the dynamics of the tap, a Chebyshev type one low-pass filter of degree 4 with 4 dB passband ripple is chosen. The cut-off frequency of the filter is set to the Doppler frequency for a vehicle velocity of 105 km/h which for this measurement correspond to a Doppler frequency of 183 Hz. This corresponds to just 4% of the total bandwidth in the spectrum. A Chebyshev filter with high passband ripple has a faster roll-off than a Butterworth filter of the same order. As we will see, the ripple shows no significant effect on the final smoother.

Using the Chebyshev filter together with the knowledge about the noise level (from the PDP), a Wiener-smoother, with smoothing lag of 5 samples, is designed. In the left part of Figure 5.4 the theoretical power spectrum of the signal $C(q^{-1})D^{-1}(q^{-1})w_k(t)$, that is the frequency response of the Chebyshev filter scaled by the variance of the driving noise, is plotted together with the Doppler spectrum for the tap. The tap is scaled to have unit variance. The right-hand part of Figure 5.4 shows the Doppler spectrum after noise reduction using the smoother. The smoother has 0 dB amplification (and close to linear phase) in the passband, even though the Chebyshev filter has a significant amount of passband ripple, and it attenuates the out of band power. Thus, the relevant part of the tap is left virtually unaltered.

The smoother in Example 5.1 has a slightly broader passband than the Chebyshev filter used in the design. This is due to a trade-off between bandwidth and linearity of the phase in the passband. For a lower noise level the bandwidth would increase, to obtain a more linear phase in the passband whereas for a higher noise level the bandwidth of the smoother would decrease, to diminish the effect of the noise power. This reduction of bandwidth will however lead to a less linear phase response in the passband. The use of a longer smoothing-lag result in a better compromise and increases the performance, as we will see in Section 5.5.

A problem with numerical accuracy of the smoother occurs at very high oversampling of the channel, as both the poles and the zeros of $C(q^{-1})D^{-1}(q^{-1})$

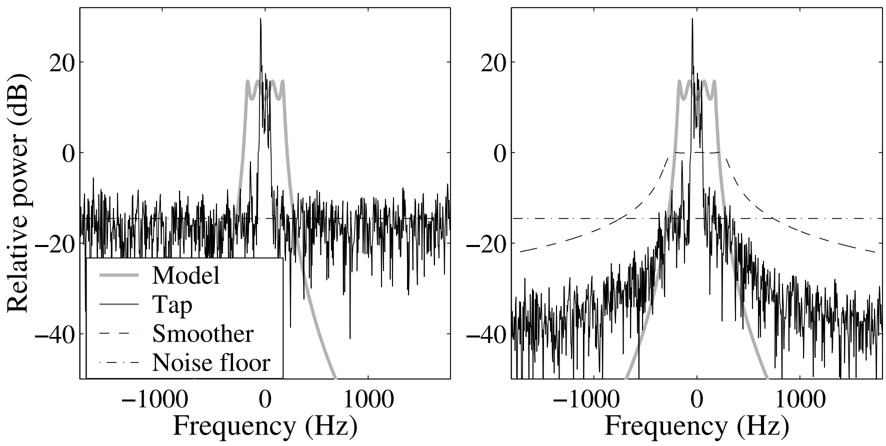

Figure 5.4: Power spectrum of the model and the Doppler spectrum for the tap corresponding to the second peak of the PDP in Figure 5.1, to the left. To the right the corresponding Doppler spectrum of the tap after noise reduction using the smoother (with dashed frequency response). The tap is normalized to have unit variance and the SNR is estimated to be 14 dB. The full bandwidth is ±4.6 kHz out of which the Figure shows ±1.8 kHz.

then will be clustered close to one in the imaginary plane. It is hard to create sufficiently narrow-band IIR-low-pass-filters and the smoother design becomes very sensitive to numerical errors on the coefficients.

5.4 FIR-smoothers

To completely avoid the non-linear phase in the passband of the IIR-smoother, low-pass FIR-filters can be used as in [17], where Parks-McClellan filters are applied to noise reduction on channels. The use of such FIR-filters is however suboptimal as compared to the FIR-Wiener-smoother, designed given a model for the correlation and an estimate of the SNR. The FIR-Wiener-smoother does not necessarily have a linear phase in the passband region, but it is designed to minimize the smoothing error, just as the IIR-Weiner-smoother.

The covariance functions of the taps in the channel are taken to be a zero order Bessel-function of the first kind (Jakes model) and the noise is assumed to be white. The Jakes model is a good description of the channels on average, even though the different realizations of the channel can deviate significantly from it.

For a given smoothing-lag and length of the smoother, the Wiener-smoother is parameterized by the SNR and the Doppler frequency. These entities can be estimated from the power delay profile and the average Doppler spectrum of the channel, as we saw in Section 5.2.

The FIR-Wiener-smoother

The correlation function for the channel in the Jakes model is given by the zero order Bessel-function of the first kind,

$$r_J(\tau) = \mathrm{J}_0(2\pi f_D \tau), \tag{5.5}$$

where f_D is the Doppler frequency.

A Wiener smoother with N coefficients and smoothing-lag m will then have the coefficient vector

$$\mathbf{w}^{[m]} = (\mathbf{R}_J + \mathbf{I}\sigma_e^2)^{-1}\mathbf{r}_J^{[m]}, \tag{5.6}$$

where $\mathbf{R}_J$ is the covariance matrix for a Jakes model channel,

$$\mathbf{r}_J^{[m]} = [r_J(-m)\dots r_J(N-1-m)]^T \tag{5.7}$$

and σ_e^2 is the variance of the white estimation error normalized by the channel tap variance. The FIR-Wiener-smoother with lag m is thus

$$\tilde{h}_k(t-m|t) = \hat{\boldsymbol{\varphi}}_k^H(t)\mathbf{w}^{[m]}, \tag{5.8}$$

where $\hat{\boldsymbol{\varphi}}_k^H(t)$ is a vector with current and delayed samples of the estimated channel

$$\hat{\boldsymbol{\varphi}}_k(t) = \left[\hat{h}_k(t)\ \hat{h}_k(t-1)\dots\hat{h}_k(t-N+1))\right]^H. \tag{5.9}$$

The remaining error after smoothing is

$$\varepsilon(t) = h_k(t) - \tilde{h}_k(t|t-m). \tag{5.10}$$

The FIR-smoother is similar to the FIR-predictor treated in Section 6.2. The variance of the smoothing error can be obtained as in (6.13) under the assumption that the tap really can be described by the Jakes model,

$$\mathrm{E}\{|\varepsilon(t)|^2\} = \sigma_\varepsilon^2 = r_J(0) - \mathbf{r}_J^{[m]H}\mathbf{R}_J\mathbf{r}_J^{[m]}. \tag{5.11}$$

This expression will be used to investigate the performance of the FIR-Wiener-smoother on Rayleigh fading tap with a Jakes spectrum.

5.5 Noise Reduction Performance

An ideal bandpass-filter, blocking all the energy outside the frequency band given by $[-f_D \ f_D]$ would reduce the energy of a white noise by a factor $2f_D/f_s$, where f_s is the channel sampling frequency. The properties of a band-limited tap inside $[-f_D \ f_D]$ would remain unaffected by such an idealized filter. The improvement in SNR measured in dB would then be $10\log_{10}(f_s/2f_D)$, independent of how noisy the tap is. The performance depends only on the amount of oversampling, with the OSR given by $f_s/2f_D$. For a highly oversampled channel there is thus a great potential in the noise reduction. The properties of a Wiener smoother approaches those of an ideal low-pass filter when $m \to \infty$ and when the design is performed for high SNR's. The question at hand is how much the IIR and FIR-smoothing operations can realize of this potential when both the smoothing-lag and the number of coefficients are limited.

5.5.1 FIR-Wiener-smoother

The theoretical performance of the FIR-Wiener-smoother can be obtained using (5.11). The performance of the noise reduction is measured as the *relative* improvement, that is the ratio between the variance σ_e^2 of the tap estimation error and the variance σ_ε^2 of the remaining error after smoothing. The improvement in SNR measured in dB is thus $10\log_{10}(\sigma_e^2/\sigma_\varepsilon^2)$. The theoretical results for the performance are obtained under the assumptions that the tap follows the Jakes channel model and that the estimation error is white.

The improvement in SNR, as compared to the noisy signal before smoothing, as a function of SNR and smoothing lag is seen in Figure 5.5. A shorter FIR-smoother with only 64 coefficients is compared to a longer FIR-smoother with 512 coefficient for the OSR of 25 and 100. The SNR improvement is highest for low SNRs and decreases as the SNR increases. This is due to that it is easier to reduce a high noise level, as compared to when there is only a small amount of noise. This is a good property, as it is mainly the noisy taps that need the noise reduction. The improvement in SNR saturates when the smoothing lag increases. The saturation is faster for lower SNRs than for higher and for a smoother using fewer coefficients as compared to using more. For both cases most of the noise reduction capacity is achieved already at a smoothing-lag of 20 samples. In Figure 5.6 the difference in performance between the smoother with 64 coefficients and the one with 512 is seen. For the OSR of 25, the difference is quite small, especially for low

SNRs and also small smoothing lags. The difference is larger for an OSR of 100. With an OSR of 100 there are 200 samples per traveled wavelength. A smoother using 64 coefficients then covers roughly one third of a wavelength. This is too short to achieve a competitive noise reduction performance. The smoother should cover something like one traveled wavelength to obtain a good noise reduction performance.

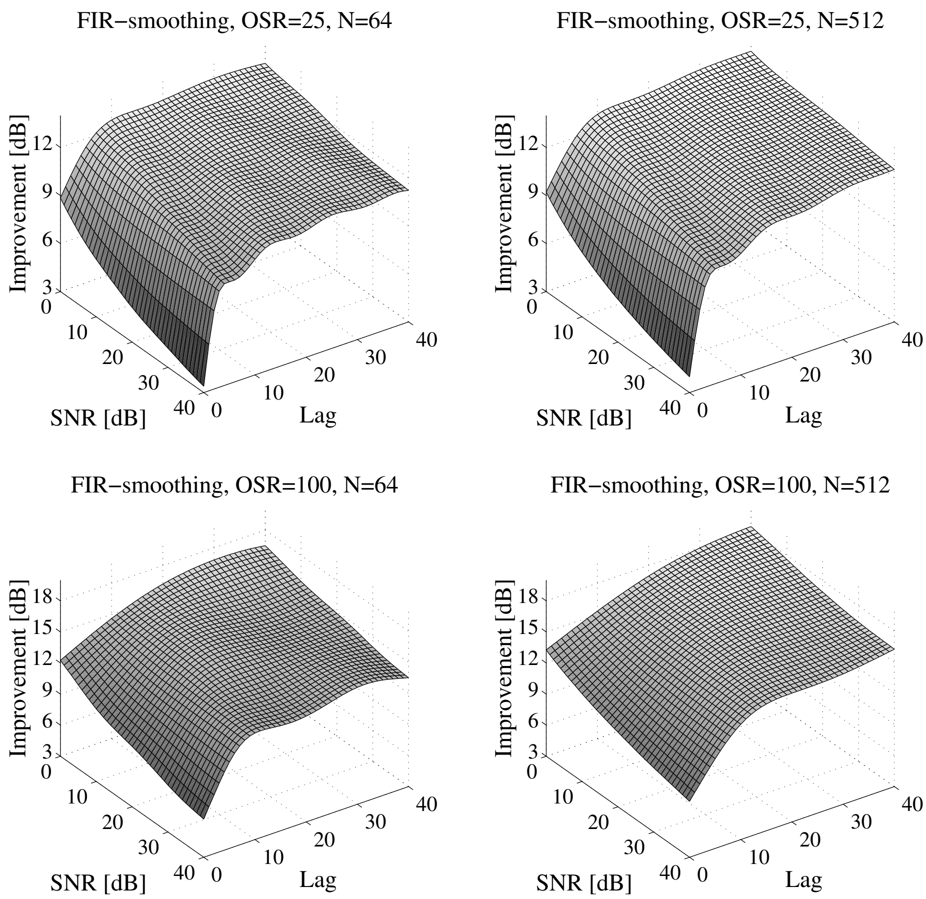

Figure 5.5: The improvement in SNR using a FIR-smoother on a Jakes channel with OSR of 25 (top row) and 100 (second row), using a smoother with 64 (left column) or 512 (right column) coefficients. An ideal bandpass filter would render 14 dB and 20 dB improvement, for the OSR of 25 and 100 respectively. The improvement saturates for high smoothing lags.

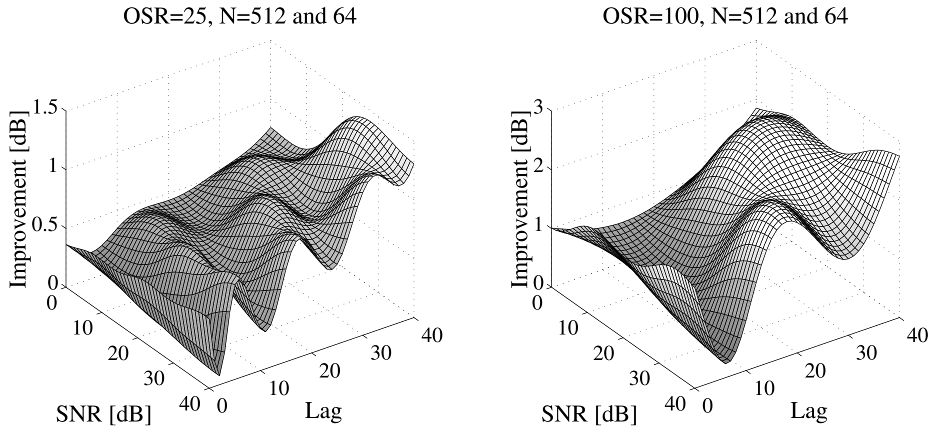

Figure 5.6: The difference in noise reduction performance that is achieved with 512 coefficients as compared to only 64 coefficients. The difference increases with the OSR, which is 25 to the left and 100 to the right.

5.5.2 Simulated Jakes channel

The noise reduction performance of the IIR and FIR-smoothers is studied on a simulated Jakes channel at different amounts of oversampling. The simulated data consist of an approximately Rayleigh distributed time series formed through the weighted summation of 50 complex sinusoids with frequencies taken as $f_D \cos\theta_n$ and where the angle θ_n is a stochastic variable drawn from the uniform distribution $[0\ 2\pi[$. The weights in the summation are complex and Gaussian distributed. Noise is added to give different SNR conditions and the SNR improvement after noise reduction is measured for a range of smoothing-lags. The simulated channel consist of 10000 data points and 100 Monte Carlo trials for each SNR and smoothing lag are performed, with different noise and channel realizations.

The model filter $C(q^{-1})D^{-1}(q^{-1})$ in (5.3) is chosen as a Chebyshev type one low-pass filter of degree 4 with 4 dB passband ripple and the cut-off frequency at the Doppler frequency f_D. The filter serves as the rough approximate model for the dynamics of the tap. For each SNR and smoothing-lag, an IIR-Wiener-smoother is designed according to [56]. The performance of the IIR-smoother is compared to a FIR-Wiener-smoother with 128 coefficients which is designed as in Section 5.5.1. In Table 5.1 the average gain in SNR over the Monte Carlo trials is presented.

As can be seen from Table 5.1, most of the gain is already achieved with

SNR improvement (max 14dB) OSR=25, IIR/FIR						
	Smoothing-lag					
SNR	0	2	5	10	20	30
0	8.9/9.0	10.1/10.2	11.6/ 11.6	12.8/12.8	13.0/13.0	13.4/13.4
5	7.6/7.8	9.3/9.4	11.3/11.3	12.4/12.5	12.8/12.8	13.1/13.3
10	6.6/6.8	8.7/8.8	11.0/11.1	12.0/12.2	12.4/12.	12.5/12.8
20	5.0/5.4	8.1/8.2	10.6/10.9	10.8/11.3	11.7/12.4	12.0/12.6
30	3.7/4.5	7.9/8.0	9.6/10.5	10.2/10.8	10.7/11.9	10.9/12.3
SNR improvement (max 17dB) OSR=50, IIR/FIR						
	Smoothing-lag					
SNR	0	2	5	10	20	30
0	10.9/10.9	11.7/11.7	12.7/12.7	14.2/14.2	15.6/15.4	15.7/15.5
5	9.8/9.7	10.8/10.8	12.3/12.3	14.0/14.0	15.3/15.2	15.3/15.2
10	8.8/8.9	10.1/10.1	11.8/11.8	13.8/13.8	14.6/14.6	14.7/14.7
20	7.1/7.4	9.0/9.2	11.4/11.5	13.3/13.6	13.6/13.8	14.3/14.6
30	5.8/6.5	8.4/8.7	11.2/11.3	12.2/13.0	12.9/13.4	13.4/14.3
SNR improvement (max 20dB) OSR=100, IIR/FIR						
	Smoothing-lag					
SNR	0	2	5	10	20	30
0	12.5/12.7	13.1/13.3	13.7/13.8	15.0/15.1	16.8/16.9	17.7/17.8
5	11.3/11.7	12.0/12.4	13.1/13.3	14.6/14.8	16.7/16.8	17.1/17.5
10	10.2/10.7	11.2/11.6	12.4/12.7	14.3/14.5	16.0/16.4	16.1/16.7
20	8.4/9.5	9.7/10.5	11.7/12.1	14.0/14.2	14.8/16.1	15.0/15.9
30	6.8/8.3	8.8/9.6	11.4/11.6	13.2/14.2	13.4/15.0	14.2/15.3

Table 5.1: Table over the average gain in SNR (dB) for the IIR and the FIR-smoother for different smoothing-lags and SNR prior to noise reduction. The IIR smoother is based on a Chebyshev type one low-pass filter of degree 4 with 4 dB passband ripple and the cut-off at the Doppler frequency as a model for the dynamics, except for OSR=100, where a fourth order Butterworth filter is used instead, due to numerical reasons. The FIR-smoother is based on the correlation for a Jakes channel. The IIR and FIR-smoothers have similar performance.

a smoothing-lag of 2 samples. To increase the smoothing-lag from 2 to 5 samples increase the gains in SNR about 1-1.5 dB. A further increase from 5 to 30 samples, result in about 3 dB improvement for an OSR of 50, which is relatively small when the SNR is around 20 dB. The improvement in SNR saturates when the smoothing lag is further increased, since then most of the noise outside the band-limits is suppressed and the smoother can not suppress the remaining noise within the band-limits. To reduce the noise within the band-limits more involved signal models, taking the fine structure of the power-spectrum into account, have to be used instead of the simple low-pass model or Jakes model.

We can assume that the gains in SNR will be close to those indicated in the table for most mobile radio channels with corresponding OSRs. The gain in SNR decreases when the OSR decreases as then there is less noise outside the frequency band occupied by the signal, that can be suppressed. The smoothing lag where the improvement in SNR saturates also decreases as the OSR decreases due to the same reason.

The performance of the FIR-smoother is similar to the IIR-smoother on the simulated channels. However, when the FIR-smoother needs on the order of hundred coefficients, the IIR-smoother needs only on the order of ten coefficients. Still the FIR-smoother is here preferred, due to numerical sensitivity of the coefficients in the IIR-smoother.

5.6 Noise Reduction and Prediction

Noise reduction could alternatively be performed by the predictor directly. The optimization of the predictor would then include the attainment of sufficient noise reduction. However, there is a clear advantage in separating the noise reduction from the prediction. In the noise reduction is it possible to exploit the high oversampling of the channel and to use a simple model for the dynamics of the taps, to obtain a pre-designed low complexity smoother. As the amount of data available for training the predictor is limited, it is beneficial if the predictor has as few coefficients to train as possible. If the noise already has been dealt with, then the predictor can use all the available degrees of freedom for prediction.

To circumvent the problem of the extra delay imposed by the smoothing-lag, a bank of smoothing-filters can be applied, with smoothing-lags from zero up to a lag that is sufficient to obtain close to optimal smoothing performance. The channel predictor can then use the latest available channel samples with the, at that time, best available noise reduction.

As we will see in Chapter 6, a predictor for a highly over-sampled channel is preferably sub-sampled, to keep complexity low and still attaining high performance. Such a sub-sampled structure can not exploit the over-sampling for noise reduction. Thus, by separating the noise reduction from the prediction we keep the complexity low while we still can use predictors with structures directly designed to suite the dynamics of the taps. The sub-sampling of a linear predictor leads to periodicity of the frequency response of the predictor, which in turn leads to aliasing in the prediction as we will see in Section 6.5. The noise reduction acts as an anti-aliasing filter, greatly improving the performance of the sub-sampled linear predictors. Noise reduction is thus a key issue when designing channel predictors.

5.7 Conclusion

- Noise reduction of over-sampled mobile radio channels can be performed using linear Wiener-smoothers.
  - The noise reduction performance is roughly proportional to the over sampling ratio (OSR).
  - Taps with low SNR benefits the most from the smoothing.
  - The use of larger smoothing-lags improves the noise reduction but the performance saturates as the lag is increased.
- An IIR-Wiener-smoother designed based on an low-pass ARMA-model of low order, for the dynamics of the tap, offers a compact smoother with few coefficients.
- An FIR-smoother, designed based on a model (preferably the Jakes model) for the correlation of the tap need significantly more coefficients than the IIR-smoother to achieve the same performance, but it is less numericly sensitive.
- The FIR-smoother should at least be long enough to cover observed channel samples from a traveled length corresponding to about one wavelength.
- Both the IIR and FIR-smoothers need estimates of the SNR and the Doppler frequency, which can be obtained either from the power delay profile or from the Doppler spectrum.
- The noise reduction acts as an anti-aliasing filter, facilitating for sub-sampled predictors.

Chapter 6

Channel Tap Prediction

6.1 Introduction

Linear prediction is a powerful tool for prediction of time series. In this work we mainly consider power prediction, that is to predict the absolute square of a complex valued signal, and some of the methods presented in Chapter 7 for that purpose will be based on linear regressions of the observations of the complex valued time series. In this Chapter the fundamental properties of the linear prediction of a complex valued channel tap are discussed. The prediction of the complex valued impulse response has in it self a great importance. It can be used to obtain information about the frequency response of the channel and in antenna array systems the predictions can be used to reduce the multiuser interference [57].

The results of complex valued linear prediction is recapitulated in Section 6.2. When the channel is oversampled, noise reduction can be applied to the regressors, as described in Chapter 5. The noise reduction results in the use of filtered regressors, treated in Section 6.2.3. Section 6.3 covers the use of iterative predictors to reduce complexity. The importance of the sub-sampling or delay spacing is demonstrated in Section 6.4.

There are two main approaches in linear prediction of a time series. The correlation based sub-sampled direct FIR-predictor and the model-based predictor, using (sub-sampled) AR or ARMA models for the dynamics of the time series. We will study both of these approaches in this Chapter.

Throughout this and the next Chapter, we assume that the correlation of the signal and noise are known and time invariant. In practice these correlations change over time and have to be estimated from the observed signal using a limited amount of data. This limits the number of predictor, or model, coefficients that can be estimated with meaningful accuracy. When working on real measurements as in Section 6.7, the sample covariances are as estimates of the covariances. The predictor coefficients are then given by the LS-solution.

A tap in a mobile radio channel

The complex valued time series under consideration constitutes a tap in a Rayleigh fading mobile radio channel, which can be modeled as a correlated complex valued circular Gaussian stochastic variable (as was motivated in Chapter 2). In Chapter 2 we modeled the contribution of a cluster of reflectors to the channel as a narrow band filtered complex Gaussian noise. Thus, even if the number of contributing clusters is low, as long as they have a angular distribution, the statistics of the channel is likely to be Gaussian and the taps are thus Rayleigh fading.

The performance of linear complex valued prediction is limited by the correlation properties of the tap and the additive noise. The same correlation functions also determine the performance of the power predictors, see Chapter 7. For omni-directional scattering, that is the Jakes channel model described in Section 2.4.2, it is easy to obtain theoretical limits of performance for the prediction of a tap. This channel model will act as an example for what results that can be anticipated, throughout this and the following Chapters. The prediction is performed on the estimated snapshots of the channel and the sampling rate of the channel time series (indexed by t) is thus given by how often the channel is estimated. The noise on the observed channel samples consist of the channel estimation error, described in Chapter 4.

The model

A complex valued scalar signal $x(t)$ is observed in noise as

$$y(t) = h(t) + e(t), \tag{6.1}$$

where $y(t)$ is the observation and $e(t)$ is a noise with zero mean and variance σ_e^2. The present Chapter treats the problem of predicting the signal $h(t+L)$, given the observations $y(t)$ up to time t. The user-defined prediction range L is measured in channel samples. In Chapter 7 the problem of predicting the power $|h(t+L)|^2$ will be analyzed, with the corresponding notation.

Assume $h(t)$ and $e(t)$ to be zero mean stationary Gaussian, complex circular independent random variables[1] with covariances

$$r_h(\tau) = \mathrm{E}\{h(t)h^*(t-\tau)\}, \tag{6.2}$$

and

$$r_e(\tau) = \mathrm{E}\{e(t)e^*(t-\tau)\}, \tag{6.3}$$

respectively. The variance for $h(t)$ is $r_h(0)$ and will be denoted σ_h^2 in the following.

6.2 Linear FIR-Prediction of a Complex Valued Signal

The FIR-predictor structure is highly suited for signals generated by AR-processes, which according to Chapter 2 is a suitable model for the dynamics of the taps in a mobile radio channel (even though an ARMA model describe the channel even better). As the FIR-predictor also offers a robust low complexity predictor that can fully exploit the correlation in the signal over the lags it uses, it is a natural choice for prediction of the taps of a mobile radio channel. The optimal linear FIR-predictor using past noisy observations to predict a signal is given by the Wiener-Hopf equations. Derivations of the optimal linear predictor are found in most textbooks in statistical signal processing, i.e. [42], [58], [59]. We will in this section recapitulate these results and introduce the corresponding notation.

6.2.1 The FIR-predictor

The goal is to predict a signal L time instances ahead using a FIR-predictor with M coefficients. The complexity is thus limited by the choice of M. In a vector formulation of the FIR-predictor, the signal can be predicted from past noisy observations as

$$\hat{h}(t+L|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\theta}, \tag{6.4}$$

where

$$\boldsymbol{\varphi}(t) = [y(t) \;\; y(t-\Delta t)\ldots y(t-(M-1)\Delta t)]^H, \tag{6.5}$$

[1]The signal in mind is a tap in a mobile radio channel which generally has zero mean. The Rayleigh fading channel fulfills these conditions. The Rice fading channel has, in addition to the **CN** distributed component, a strong deterministic oscillating component. The complex valued samples are in that case not Gaussian and can have a mean different from zero.

is the regressor and the column vector $\boldsymbol{\theta} = [\theta_1 \ldots \theta_M]^T$ contains the complex valued predictor coefficients. It is here assumed that the taps of the FIR-predictor are equally spaced[2] with delay spacing Δt. The delay spacing is commonly set to one, which for a given M is a suboptimal choice in many cases. The choice of M and Δt should depend on the statistics of $h(t)$ and $e(t)$ and in the common case where the statistics has to be estimated, the choice also should depend on the number of training samples. In Section 6.4 we will see the importance of the delay spacing for the prediction performance of a tap described by the Jakes model.

The vector product in (6.4) can also be expressed as a filtering operation with a sub-sampled FIR-filter (as seen in Figure 6.1),

$$\hat{h}(t+L|t) = \theta(q^{-\Delta t})y(t), \tag{6.6}$$

where $\theta(q^{-\Delta t}) = \theta_1 + \theta_2 q^{-\Delta t} + \ldots \theta_M q^{-(M-1)\Delta t}$ is a polynomial in the unit delay operator q^{-1}. This is a convenient representation when the frequency response of the predictor is of interest.

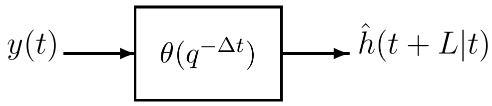

Figure 6.1: The predictor (6.6) is a sub-sampled FIR-filter.

The complex valued prediction error is given by

$$\varepsilon_c(t) = h(t) - \hat{h}(t|t-L) = h(t) - \boldsymbol{\varphi}^H(t-L)\boldsymbol{\theta}. \tag{6.7}$$

As $h(t)$ and $y(t)$ have zero mean, the same holds for the prediction error. The mean square error (MSE) will thus be equal to the variance, which is

$$\begin{aligned} \sigma^2_{\varepsilon_c} &= \mathrm{E}\{|\varepsilon_c(t)|^2\} = \mathrm{E}\{(h(t) - \boldsymbol{\varphi}^H(t-L)\boldsymbol{\theta})(h^H(t) - \boldsymbol{\theta}^H\boldsymbol{\varphi}(t-L))\} \\ &= \sigma^2_h - \mathbf{r}^H_{h\varphi}\boldsymbol{\theta} - \boldsymbol{\theta}^H\mathbf{r}_{h\varphi} + \boldsymbol{\theta}^H\mathbf{R}_{\varphi}\boldsymbol{\theta}, \end{aligned} \tag{6.8}$$

where $\mathbf{r}_{h\varphi}$ is the cross-covariance between the signal and the regressor

$$\begin{aligned} \mathbf{r}_{h\varphi} &= \mathrm{E}\{h(t)\boldsymbol{\varphi}(t-L)\} \\ &= \mathrm{E}\{h(t)[y(t-L) \ldots y(t-L-(M-1)\Delta t)]^H\} \\ &= [r_h(L) \quad r_h(L+\Delta t) \ldots r_h(L+(M-1)\Delta t)]^T \end{aligned} \tag{6.9}$$

[2] It is common to use equally spaced taps in FIR-filters. Though, for some signals it can be beneficial to use non-equal delay spacing as then the spacing between the taps can differ and the filter can be tailored to exploit the correlation further. A signal generated by a noise fed through a filter with an impulse response that has short stretches of non-zero elements in between many lags that are zero, is a typical example of a signal for which a predictor with a limited number of coefficients benefits from non-equal delay spacing.

and $\mathbf{R}_\varphi$ is the covariance matrix for the regressors

$$\mathbf{R}_\varphi = \mathrm{E}\{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\} = \mathbf{R}_h + \mathbf{R}_e \tag{6.10}$$

where

$$\mathbf{R}_h = \begin{pmatrix} \sigma_h^2 & r_h^*(\Delta t) & \dots & r_h((M-1)\Delta t) \\ r_h(\Delta t) & \sigma_h^2 & & \\ \vdots & & \ddots & \\ r_h((M-1)\Delta t) & & & \sigma_h^2 \end{pmatrix} \tag{6.11}$$

is the covariance matrix for $h(t)$ and $\mathbf{R}_e$ is the corresponding correlation matrix for the noise $e(t)$.

From the Wiener-Hopf equations, we know that the MSE in (6.8) is minimized by $\boldsymbol{\theta} = \boldsymbol{\theta}_c$ with

$$\boldsymbol{\theta}_c = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi} \tag{6.12}$$

and that the minimum MSE is given by [58]

$$\min_{\boldsymbol{\theta}} \sigma_{\varepsilon_c}^2 = \sigma_h^2 - \mathbf{r}_{h\varphi}^H\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}. \tag{6.13}$$

In the case of perfect prediction, the error will be zero and in case of no prediction, that is $\boldsymbol{\theta} = 0$, the error will equal the signal $h(t)$. The variance of the prediction error will thus be bounded as $0 \leq \sigma_{\varepsilon_c}^2 \leq \sigma_h^2$ when the optimal coefficients are used. (In practice, with imperfectly tuned predictors, we may of course have $\sigma_{\varepsilon_c}^2 > \sigma_h^2$.)

The correlation between the signal $h(t)$ and the prediction $\hat{h}(t|t-L)$ is

$$r_{h\hat{h}} = \mathrm{E}\{h(t)\hat{h}^*(t|t-L)\} = \mathrm{E}\{\boldsymbol{\theta}^H\boldsymbol{\varphi}(t-L)h(t)\} = \boldsymbol{\theta}^H\mathbf{r}_{h\varphi}, \tag{6.14}$$

which for the optimal coefficients $\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$ is

$$\begin{aligned} r_{h\hat{h}} &= \mathbf{r}_{h\varphi}^H\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi} & (6.15)\\ &= \sigma_h^2 - \sigma_{\varepsilon_c}^2, & (6.16) \end{aligned}$$

where we have used (6.13) in the last equality.

The variance of the prediction depends on the predictor coefficients and the covariance matrix for the regressor as

$$\sigma_{\hat{h}}^2 = \mathrm{E}\{|\hat{h}(t+L|t)|^2\} = \boldsymbol{\theta}^H\mathrm{E}\{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\}\boldsymbol{\theta} = \boldsymbol{\theta}^H\mathbf{R}_\varphi\boldsymbol{\theta}. \tag{6.17}$$

Using the optimal coefficients we obtain

$$\sigma_{\hat{h}}^2 = \mathbf{r}_{h\varphi}^H \mathbf{R}_{\varphi}^{-1} \mathbf{r}_{h\varphi} = \sigma_h^2 - \sigma_{\varepsilon_c}^2. \tag{6.18}$$

The variance of the prediction $\hat{h}(t|t-L)$ will thus be smaller than the variance of $h(t)$, when the optimal prediction coefficients are used.

The correlation between the prediction error ε_c and the prediction $\hat{h}(t|t-L)$ can be obtained using (6.7) and (6.14)

$$r_{\varepsilon_c\hat{h}} = \mathrm{E}\{\varepsilon_c(t)\hat{h}^*(t|t-L)\} = \mathrm{E}\{\big(h(t) - \hat{h}(t|t-L)\big)\hat{h}^*(t|t-L)\} = r_{h\hat{h}} - \sigma_{\hat{h}}^2. \tag{6.19}$$

When using the MSE-optimal predictor coefficients, the correlation is obtained by inserting (6.16) and (6.18) into (6.19),

$$r_{\varepsilon_c\hat{h}} = r_{h\hat{h}} - \sigma_{\hat{h}}^2 = \sigma_h^2 - \sigma_{\varepsilon_c}^2 - \sigma_h^2 + \sigma_{\varepsilon_c}^2 = 0. \tag{6.20}$$

The prediction $\hat{h}(t|t-L)$ and prediction error $\varepsilon_c(t)$ are thus uncorrelated at time t, when the MSE-optimal linear predictor is used.

6.2.2 Correlated signals

If instead of the flat fading channel with just one tap, we consider a frequency selective channel with many taps, then the scalar model (6.1) is changed to the vector model

$$\mathbf{y}(t) = \mathbf{h}(t) + \mathbf{e}(t), \tag{6.21}$$

where $\mathbf{y}(t)$ is a column vector of observations of the signals $\mathbf{h}(t)$ corrupted by the noise vector $\mathbf{e}(t)$. The prediction can be performed separately for each tap, just as in equation (6.4),

$$\hat{h}_k(t+L|t) = \boldsymbol{\varphi}_k^H(t)\boldsymbol{\theta}_k, \tag{6.22}$$

where the regressor $\boldsymbol{\varphi}_k^H(t)$ contain the observations of tap k, as in (6.5). However, if the taps are correlated, then this is a suboptimal predictor, as the additional information in the other taps is unexploited. Pulse shaping of transmitted symbols in the communication link, as described in Section 2.3, cause the taps of the estimated channel to be correlated. Generally the adjacent taps are most correlated and then the correlation between the taps decays with increasing distance in lag.

Let the regressor consist, not only of the observations of on tap, but of observations of all the taps

$$\boldsymbol{\varphi}(t) = [\mathbf{y}^T(t) \;\; \mathbf{y}^T(t-\Delta t) \ldots \mathbf{y}^T(t-(M-1)\Delta t)]^H. \tag{6.23}$$

Then a linear prediction of tap k can benefit from the information in the other taps,

$$\hat{h}_k(t+L|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\theta}_k. \tag{6.24}$$

The covariance matrix for the regressor, $\mathbf{R}_\varphi = \mathrm{E}\{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\}$, and the cross-covariance between tap k and the regressor, $\mathbf{r}_{h_k\varphi} = \mathrm{E}\{h_k(t)\boldsymbol{\varphi}(t-L)\}$, are similar to (6.10) and (6.9) but will in addition contain elements of the correlation between the taps. The optimal predictor coefficients for tap k then follows from (6.12),

$$\boldsymbol{\theta}_k = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h_k\varphi}. \tag{6.25}$$

The drawback using all the observed taps in the regressor is that many more predictor coefficients are needed. This is a problem when the statistics are unknown, and the coefficients have to be estimated from a limited number of observations. To keep the number of coefficients low, only the nearest neighboring taps to tap k could be included in the regressor, when predicting tap k. Most of the correlation between the taps is then still exploited, at a smaller increase in the number of predictor coefficients.

6.2.3 Filtered regressors

To increase the performance of the linear prediction it is possible to use filtered regressors, that are less affected by the noise (or tap estimation error) $e(t)$. A full knowledge about the correlation properties of signal and noise is not needed to perform the noise reduction. It is sufficient to use a model of the dynamics and an estimate of the noise variance to design a Wiener-smoother, as in Chapter 5, if the signal is highly oversampled.

The smoothing lag introduced by the noise reduction operation introduces a corresponding delay when the smoothed signals are used in the regressor for prediction of future values. The predictor has to compensate for the delay by a longer prediction range. We thus want the smoothing-lag to be small. But as a larger smoothing-lag also results in better noise reduction, we also want to use as large smoothing-lags as possible. To circumvent this problem, a bank of noise reduction smoothers with smoothing lags from zero up to as long as needed, can be used. For a regressor with M elements and a delay spacing of Δt the largest lag is $(M-1)\Delta t$.[3] The regressor containing smoothed observations has to be available without delay. Under this

[3] In practice it is not necessary to have a smoother for each lag from zero to $(M-1)\Delta t$. It is sufficient to use a subset of lags where the largest lag obtain close to optimal noise reduction.

condition, each element in the regressor consist of the smoothed observation with an as large smoothing-lag as possible.

In Chapter 5 both FIR and IIR-smoothers are considered. The IIR-smoother offers a more compact description with fewer coefficients than an FIR-smoother with the corresponding noise-reduction capacity. The drawback with the IIR-smoother is that it is very numerically sensitive and thus hard to implement. In the following we will therefore use an FIR-smoother of length N instead. The use of the FIR-smoother also simplifies the calculation of the correlation below.

Smoothed observations in the regressors

The smoothed observations, using a FIR-smoother of length N, with lag m is obtained from (5.8) as

$$\tilde{y}(t-m|t) = \sum_{k=0}^{N-1} w_k^{[m]} y(t-k) = \boldsymbol{\varphi}_y^H(t)\mathbf{w}^{[m]}, \tag{6.26}$$

where

$$\boldsymbol{\varphi}_y(t) = [y(t)\ y(t-1)\dots y(t-N+1)]^H \tag{6.27}$$

is a vector with N old and present observations, $w_k^{[m]}$ is the kth coefficient of the FIR-smoother with lag m and

$$\mathbf{w}^{[m]} = [w_0^{[m]} \dots w_{N-1}^{[m]}]^T \tag{6.28}$$

is the corresponding filter vector. The smoothing operation can also be expressed using the polynomial filter

$$W^{[m]}(q^{-1}) = w_0^{[m]} + w_1^{[m]} q^{-1} + \dots + w_{N-1}^{[m]} q^{-(M-1)}. \tag{6.29}$$

Then (6.26) is expressed as

$$\tilde{y}(t-m|t) = W^{[m]}(q^{-1})y(t). \tag{6.30}$$

The regressor with delay spacing Δt, using appropriately smoothed observations, is

$$\boldsymbol{\varphi}(t) = [\tilde{y}(t|t)\,\tilde{y}(t-\Delta t|t)\dots\tilde{y}(t-(M-1)\Delta t|t)]^H\,. \tag{6.31}$$

The prediction is then obtained as in (6.4) using the smoothed regressor given by (6.31) instead of the unfiltered vector (6.5)

$$\hat{h}(t+L|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\theta}. \tag{6.32}$$

If a desired smoothing lag $k\Delta t$ in the regressor (6.31), is larger than the largest available smoothing lag $m_{\max}$ provided by the bank of smoothers, then $\varphi_{k+1}(t) = \tilde{y}(t - k\Delta t|t)$ in (6.31) is substituted with

$$\begin{aligned}
\varphi_{k+1}(t) &= \tilde{y}(t - k\Delta t | t - k\Delta t + m_{\max}) && (6.33)\\
&= \boldsymbol{\varphi}_y^H(t - k\Delta t + m_{\max})\mathbf{w}^{[m_{\max}]} && (6.34)\\
&= W^{[m_{\max}]}(q^{-1})q^{-k\Delta t + m_{\max}}y(t), && (6.35)
\end{aligned}$$

which renders a smoothed estimate of the signal with the desired delay but using a shorter smoothing lag.

A FIR-predictor using the filtered regressor (6.31) from a bank of FIR-smoothers is depicted in Figure 6.2. As both the noise-reduction and the

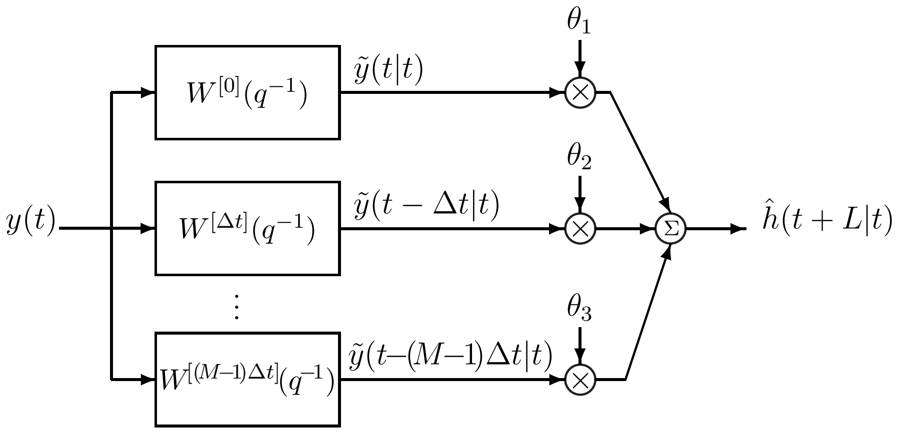

Figure 6.2: A predictor using smoothed regressors (6.31) from a bank of smoothers.

predictor are linear filters, they can be combined into one single predictor as shown in Figure 6.4.

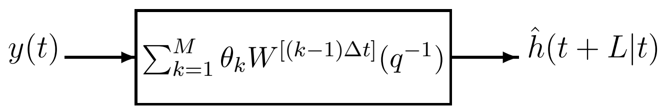

Figure 6.3: The combined smoother and predictor system in Figure 6.2 represented as one single linear filter.

The correlation matrix for the smoothed regressors

The element $[n, m]$ of the covariance matrix for the regressor (6.31), that is $\mathbf{R}_\varphi = \mathrm{E}\{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\}$, consist of the correlation between two smoothed observations with lags $(m-1)\Delta t$ and $(n-1)\Delta t$ respectively,

$$\begin{aligned}
[\mathbf{R}_\varphi]_{n,m} &= \mathrm{E}\{\tilde{y}(t-(m-1)\Delta t|t)\tilde{y}^*(t-(n-1)\Delta t|t)\} \\
&= \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} w_k^{[(m-1)\Delta t]} w_l^{[(n-1)\Delta t]*} \underbrace{\mathrm{E}\{y(t-k)y^*(t-l)\}}_{r_y(l-k)} \qquad (6.36)\\
&= \left(\mathbf{w}^{[(n-1)\Delta t]}\right)^H \mathbf{R}_y \mathbf{w}^{[(m-1)\Delta t]}, \qquad (6.37)
\end{aligned}$$

where $\mathbf{R}_y = \mathrm{E}\{\boldsymbol{\varphi}_y(t)\boldsymbol{\varphi}_y^H(t)\}$ is the covariance matrix, of size $N \times N$, of the observations. The covariance matrix for the regressor $\mathbf{R}_\varphi$, which is $M \times M$, can thus be obtained from the matrix multiplications

$$\mathbf{R}_\varphi = \mathbf{W}^H \mathbf{R}_y \mathbf{W}, \qquad (6.38)$$

where $\mathbf{W}$ is an $N \times M$ matrix with all the smoothers for the different lags used in the regressor

$$\mathbf{W} = [\mathbf{w}^{[0]}\ \mathbf{w}^{[\Delta t]} \ldots \mathbf{w}^{[(M-1)\Delta t]}]. \qquad (6.39)$$

The elements of the cross-correlation $\mathbf{r}_{h\varphi}$ can be obtained in the same manner,

$$\begin{aligned}
[\mathbf{r}_{h\varphi}]_n = \mathrm{E}\{h(t)\tilde{y}(t-(n-1)\Delta t|t)\} &= \left(\mathbf{w}^{[(n-1)\Delta t]}\right)^H \mathrm{E}\{h(t)\varphi_y(t-L)\} \\
&= \left(\mathbf{w}^{[(n-1)\Delta t]}\right)^H \mathbf{r}_{h\varphi_y}. \qquad (6.40)
\end{aligned}$$

The cross-correlation for the signal and filtered regressor is thus

$$\mathbf{r}_{h\varphi} = \mathbf{W}^H \mathbf{r}_{h\varphi_y}. \qquad (6.41)$$

The relations (6.38) and (6.41) can be used when the covariance matrix for the observations are known. They will be used to obtain the limits of prediction performance for the Jakes channel in Section 6.7.1.

Whenever the signal $h(t)$ is unknown, we have to rely on the observations $y(t)$ and the noise reduction to obtain a smoothed estimate of the signal that can be used instead of the true signal, e.g. in expressions for the prediction error. The approximation $h(t) \approx \tilde{y}(t|t-m\Delta t)$ is then used to substitute

$h(t)$ with the smoothed estimate. This is done whenever real measurements are used.

Only when there occurs abrupt changes in the radio environment, causing discontinuities in the observed taps of the channel, the smoothed observation $\tilde{y}(t|t-m\Delta t)$ will deviate significantly from $h(t)$ and also $y(t)$. In that case a reset of the prediction coefficients has to be done, as the statics change. The difference between $y(t)$ and $\tilde{y}(t|t-m\Delta t)$ can therefore be used to detect abrupt changes of the statistics.

6.2.4 Estimation of predictor coefficients from data

When the signal $h(t)$ is unknown it is substituted by $\tilde{y}(t|t-m\Delta t)$. In the same manner the true covariances are substituted by estimated covariances when the statistics of the signal and noise are unknown.

The L-step prediction problem can be formulated as

$$\tilde{y}(t|t+m_{\max}) = \boldsymbol{\varphi}^H(t-L)\boldsymbol{\theta} + \hat{\varepsilon}_c(t), \tag{6.42}$$

where $\tilde{y}(t|t+m_{\max})$ and $\boldsymbol{\varphi}^H(t-L)$ consist of smoothed observations. This is a problem similar to the channel estimation problem (4.9) that was solved with the least squares method. The $\boldsymbol{\theta}$ that minimize the variance of $\hat{\varepsilon}_c(t)$ over a set of observations can thus be obtained using the least squares method.

Form the matrix and vector of smoothed observations

$$\boldsymbol{\Phi} = [\boldsymbol{\varphi}(i)\dots\boldsymbol{\varphi}(j)]^H \tag{6.43}$$

$$\tilde{\mathbf{y}} = [\tilde{y}(i+L|i+L+m_{\max})\dots\tilde{y}(j+L|j+L+m_{\max})]^H \tag{6.44}$$

where $i<j$ and both i and j are chosen so that all the elements of $\boldsymbol{\Phi}$ and $\tilde{\mathbf{y}}$ are available. The least squares estimate of $\boldsymbol{\theta}$ is obtained as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Phi}^\dagger\tilde{\mathbf{y}}. \tag{6.45}$$

The sample correlation estimates of the correlation matrix for the regressors and the cross-correlation between signal and regressor can be obtained as

$$\hat{\mathbf{R}}_\varphi = \frac{1}{j-i+1}\boldsymbol{\Phi}^H\boldsymbol{\Phi} \tag{6.46}$$

$$\hat{\mathbf{r}}_{h\varphi} = \frac{1}{j-i+1}\boldsymbol{\Phi}^H\tilde{\mathbf{y}}. \tag{6.47}$$

These estimates will be used whenever prediction on measured channels is performed. The tap prediction is obtained as

$$\hat{h}(t+L|t) = \boldsymbol{\varphi}^H(t)\hat{\boldsymbol{\theta}} \tag{6.48}$$

The structure of the predictor is illustrated by Figure 6.4.

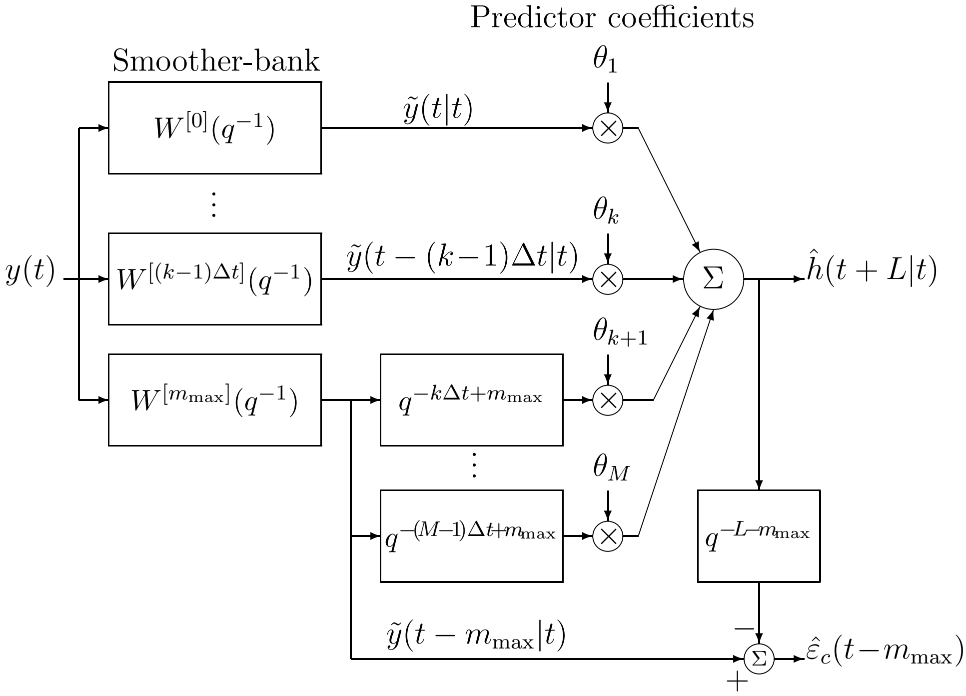

Figure 6.4: The bank of smoothers and the predictor. Here $(k-1)\Delta t < m_{\max} \leq k\Delta t$ where m_{max} is the largest smoothing lag provided by the bank of smoothers. In the evaluation of the prediction error $\varepsilon_c(t)$, the signal $h(t)$ is assumed unknown and is substituted by $\tilde{y}(t|t-m_{\max})$. The true prediction error is thus unknown but an estimate $\hat{\varepsilon}_c(t)$ is obtained.

6.3 Iterative Prediction

In iterative prediction, the predictor is tuned to predict the time series a short step into the future. The prediction is then used as an observation of the time series that is included in the regressor. A prediction of the future regressor is thus performed and this predicted regressor is used in the prediction of yet another step ahead. This is repeated until the desired prediction range is achieved. Iterative prediction can be used to save complexity when predictions for many ranges have to be obtained, as only one predictor is trained instead of one for each range. The reduction in complexity is traded for accuracy, as the iterative predictor is not optimal. The iterated FIR-predictor is closely related to an AR-model based predictor, as we will see in an example.

Predicted regressor

The observations are predicted a Δt step into the future. Analogous to (6.4) we obtain

$$\hat{y}(t+\Delta t|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\theta}_y \tag{6.49}$$

where $\boldsymbol{\theta}_y = \mathbf{R}_\varphi^{-1}\mathbf{r}_{y\varphi}$, as in (6.12), renders optimal prediction of the observation. Here $\mathbf{r}_{y\varphi} = \mathrm{E}\{y(t)\boldsymbol{\varphi}(t-\Delta t)\}$ is the cross-covariance between the observation and the regressor. A regressor that includes a predicted observation as its first element can thus be expressed as an update matrix times the regressor

$$\begin{aligned} \hat{\boldsymbol{\varphi}}^H(t+\Delta t|t) &= [\hat{y}(t+\Delta t|t) \;\; y(t) \;\; y(t-\Delta t)\ldots y(t-(M-2)\Delta t)] \\ &= \boldsymbol{\varphi}^H(t)\boldsymbol{\Phi} \ , \end{aligned} \tag{6.50}$$

where the update matrix is

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\theta}_y & \begin{matrix} \mathbf{I_{M-1}} \\ 0\ldots 0 \end{matrix} \end{pmatrix}. \tag{6.51}$$

Here $\mathbf{I_{M-1}}$ denotes the $M-1 \times M-1$ identity matrix. A prediction of the observation two Δt steps ahead can thus be obtained using the one Δt step predictor parameters and the one Δt step predicted regressor,

$$\hat{y}(t+2\Delta t|t) = \hat{\boldsymbol{\varphi}}^H(t+\Delta t|t)\boldsymbol{\theta}_y = \boldsymbol{\varphi}^H(t)\boldsymbol{\Phi}\boldsymbol{\theta}_y. \tag{6.52}$$

By successive inclusion of $l-1$ predicted observations, and exclusion of $l-1$ outdated observations, in the regressor we obtain the predicted regressor

$$\hat{\boldsymbol{\varphi}}^H(t+(l-1)\Delta t|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\Phi}^{l-1} \ . \tag{6.53}$$

The $l\Delta t$ predictor for the observations are obtained from equation (6.53) as

$$\hat{y}(t+l\Delta t|t) = \hat{\boldsymbol{\varphi}}^H(t+(l-1)\Delta t|t)\boldsymbol{\theta}_y = \boldsymbol{\varphi}^H(t)\boldsymbol{\Phi}^{l-1}\boldsymbol{\theta}_y. \tag{6.54}$$

The iterated prediction of the observations thus corresponds to the predictor $\hat{y}(t+l\Delta t|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\theta}$ with the coefficients $\boldsymbol{\theta} = \boldsymbol{\Phi}^{l-1}\boldsymbol{\theta}_y$. To make a comparison to a state space representation easier we can express (6.54) as

$$\hat{y}(t+l\Delta t|t) = \hat{\boldsymbol{\varphi}}^H(t+l\Delta t|t)\mathbf{c} = \boldsymbol{\varphi}^H(t)\boldsymbol{\Phi}^l\mathbf{c}. \tag{6.55}$$

where $\mathbf{c} = [1 \;\; 0\ldots 0]^T$.

Predicted signal

As we now have a predicted regressor we can use it to predict the signal, just as well as the observation. If the Δt step predictor for the time series $h(t)$ is $\hat{h}(t+\Delta t|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\theta}_h$ then the iterative $L = l\Delta t$ step predictor of the time series $h(t)$ thus is

$$\hat{h}(t+l\Delta t|t) = \hat{\boldsymbol{\varphi}}^H(t+(l-1)\Delta t|t)\boldsymbol{\theta}_h = \boldsymbol{\varphi}^H(t)\begin{pmatrix}\boldsymbol{\theta}_y & \begin{matrix}\mathbf{I_{M-1}} \\ 0\ldots 0\end{matrix}\end{pmatrix}^{l-1}\boldsymbol{\theta}_h. \quad (6.56)$$

If the optimal Δt step predictors for $y(t)$ and $h(t)$ are used, that is the coefficients are $\boldsymbol{\theta}_y = \mathbf{R}_\varphi^{-1}\mathbf{r}_{y\varphi}$ and $\boldsymbol{\theta}_h = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$ where $\mathbf{r}_{h\varphi} = \mathrm{E}\{h(t)\boldsymbol{\varphi}(t-\Delta t)\}$, then the iterative predictor can be interpreted as the predictor in (6.4) with the coefficients

$$\boldsymbol{\theta} = \begin{pmatrix}\boldsymbol{\theta}_y & \begin{matrix}\mathbf{I_{M-1}} \\ 0\ldots 0\end{matrix}\end{pmatrix}^{l-1}\boldsymbol{\theta}_h = \begin{pmatrix}\mathbf{R}_\varphi^{-1}\mathbf{r}_{y\varphi} & \begin{matrix}\mathbf{I_{M-1}} \\ 0\ldots 0\end{matrix}\end{pmatrix}^{l-1}\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}. \quad (6.57)$$

This choice of predictor parameters, obtained by iteration, coincides with the optimal predictor parameters when the signal is generated by an AR-process and is observed without noise, as shown in the example below.

Example 6.1 Iterative prediction of an AR-process

The iterative predictor coincides with the optimal predictor for the case shown in this example. Assume that there is no noise on the observations and that an AR-process generates the time series. In polynomial form this is expressed as

$$A(q^{-1})h(t) = w(t-1) \quad , \quad y(t) = h(t) \quad (6.58)$$

where $A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$ is an n:th order polynomial in the delay operator q^{-1} and $w(t)$ is the innovation sequence. The observation can thus be expressed as a weighted sum of old observations and an innovation

$$y(t) = -a_1y(t-1) - \cdots - a_ny(t-n) + w(t-1). \quad (6.59)$$

The Wiener-Hopf solution for the predictor is equivalent to the steady state Kalman predictor [60]. It is thus convenient to use the state space formulation of (6.58),

$$\begin{aligned}\mathbf{z}(t+1) &= \mathbf{A}\mathbf{z}(t) + \mathbf{c}w(t) \\ y(t) &= \mathbf{c}^H\mathbf{z}(t)\end{aligned} \quad (6.60)$$

with

$$\mathbf{A} = \begin{pmatrix} -a_1 & \dots & -a_{n-1} & -a_n \\ 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{pmatrix} \tag{6.61}$$

and $\mathbf{c} = [1 \quad 0 \dots 0]^H$. The states in (6.60) consists of old observations, $\mathbf{z}(t) = [y(t) \quad y(t-1) \dots y(t-n+1)]^T$. The Kalman one-step predictor is

$$\hat{\mathbf{z}}(t+1|t) = \mathbf{A}\hat{\mathbf{z}}(t|t-1) + \mathbf{k}[y(t) - \mathbf{c}^H \hat{\mathbf{z}}(t|t-1)] \tag{6.62}$$

and in steady state the Kalman prediction gain is

$$\mathbf{k} = [-a_1 \quad 1 \quad 0 \dots 0]^T \tag{6.63}$$

as then only the first state has an error due to the innovation. The estimated states in (6.60) thus consist of old observations except the first, that is a predicted observation. The predicted states in equation (6.62) are thus $\hat{\mathbf{z}}(t+1|t) = \mathbf{A}\mathbf{z}(t)$. The steady state l-step Kalman predictor for the AR-process is

$$\hat{\mathbf{z}}(t+l|t) = \mathbf{A}^{l-1}(\mathbf{A}\hat{\mathbf{z}}(t|t-1) + \mathbf{k}[y(t) - \mathbf{c}^H \hat{\mathbf{z}}(t|t-1)]) = \mathbf{A}^l \mathbf{z}(t) \tag{6.64}$$

The optimal l-step prediction for the observations is thus obtained as

$$\hat{y}(t+l) = \mathbf{c}^T \mathbf{A}^l \mathbf{z}(t) = (\mathbf{z}^*(t))^H (\mathbf{A}^T)^l \mathbf{c} \tag{6.65}$$

which is a predictor of the same form as in equation (6.55) with $\mathbf{z}^*(t)$ corresponding to the regressor $\boldsymbol{\varphi}(t)$. We only need to show that the update matrix $\mathbf{\Phi}$ defined in (6.51) is the same as $\mathbf{A}^T$ in this example to show that the iterative prediction is optimal in this case.

The covariance matrix for the states, with old observations, is obtained as $\mathrm{E}\{\mathbf{z}(t)\mathbf{z}^H(t)\} = \mathbf{R}_z$ and the cross-covariance is $\mathrm{E}\{y(t)\mathbf{z}^H(t-l)\} = \mathbf{r}_{yz}$. The coefficients for the optimal one-step predictor are given by

$$\boldsymbol{\theta}_y = \mathbf{R}_z^{-1}\mathbf{r}_{yz} = [-a_1 \quad -a_2 \cdots - a_n]^T \tag{6.66}$$

which can be seen directly in equation (6.59), since $w(t-1)$ has zero mean. Thus the update matrix $\mathbf{\Phi}$ defined in (6.51) is

$$\mathbf{\Phi} = \begin{pmatrix} \boldsymbol{\theta}_y & \begin{matrix} \mathbf{I_{M-1}} \\ 0 \dots 0 \end{matrix} \end{pmatrix} = \begin{pmatrix} -a_1 & 1 & & 0 \\ \vdots & & \ddots & \\ -a_{n-1} & 0 & & 1 \\ -a_n & 0 & \dots & 0 \end{pmatrix} = \mathbf{A}^T. \tag{6.67}$$

The iterative predictor is thus optimal in the case where we observe an AR-process without noise and let $\Delta t = 1$ (one time step).

6.4 The Delay Spacing

When designing an L-step FIR-predictor there are two parameters that remain to be studied, that is the number of predictor coefficients M, and the delay spacing Δt. When the statistics of the signal and the noise are unknown, the predictor coefficients have to be estimated from the observed taps of the channel. As the statistics in practice are time varying, the number of samples (the adaption interval) of the channel that can be used for estimation of the predictor coefficients are limited. The accuracy of the estimated coefficients is also dependent on the channel estimation error and the dynamics of the channel. All these factors set a practical limit on how many predictor coefficients that can/should be estimated from the observed channel. The gain in prediction performance obtained by using more predictor coefficients, can be lost due to the estimation error on those coefficients. It is thus beneficial if the number of predictor coefficients M, can be kept low at a small loss in performance. This puts an emphasis on the proper choice of delay spacing and on the noise reduction pre-processing step.

We will in the following section see that the choice of the delay spacing Δt (see Figure 6.5) plays a crucial role for the obtained prediction performance.

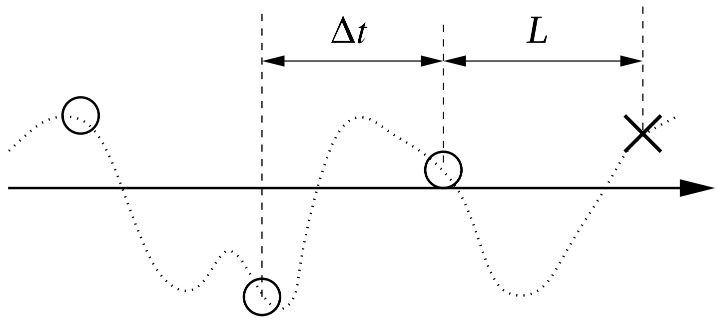

Figure 6.5: How should the spacing, Δt, between the samples used for prediction be chosen for a given prediction range L and a given number M of predictor coefficients? The rings $\circ$ denote the samples in the regressor and the cross $\times$ denotes the sample to be predicted.

The performance criterion to be used for the complex prediction is the NMSE

$$V(\boldsymbol{\theta}, \Delta t, M, L, \gamma) = \frac{\mathrm{E}\{|h(t) - \hat{h}(t|t-L)|^2\}}{\mathrm{E}\{|h(t)|^2\}}. \tag{6.68}$$

Assume a Rayleigh fading signal with the SNR γ predicted L-steps ahead by an FIR-predictor with M coefficients and delay spacing Δt. Then the best prediction coefficients $\boldsymbol{\theta}$ are given by $\boldsymbol{\theta} = \mathbf{R}_{\varphi}^{-1}\mathbf{r}_{h\varphi}$ from equation (6.12). For a given prediction range and length of the predictor (a complexity constraint), the delay spacing is thus the only remaining parameter to be tuned, to minimize the NMSE.

In Section 6.1 a vector formulation of the FIR-predictor (6.4) for a complex signal was introduced. With the optimal coefficients $\boldsymbol{\theta}_c = \mathbf{R}_{\varphi}^{-1}\mathbf{r}_{h\varphi}$ from (6.13) the FIR-predictor is given by

$$\hat{h}(t+L|t) = \boldsymbol{\varphi}^H(t)\boldsymbol{\theta}_c \tag{6.69}$$

where the regressor vector consists of noisy observations

$$\boldsymbol{\varphi}(t) = [y(t) \;\; y(t-\Delta t) \ldots y(t-(M-1)\Delta t)]^H. \tag{6.70}$$

The regressor vector can of course consist of smoothed noisy observations as in (6.31) and can also include other correlated signals (6.23), i.e. neighboring taps in the channel. The delay spacing Δt, adjusts how densely the samples in the regressor are taken from the time series. A change of Δt thus changes the elements of $\mathbf{r}_{h\varphi}$ and $\mathbf{R}_{\varphi}$ in (6.9) and (6.11), respectively.

To illustrate how Δt and M can be chosen we will consider prediction of a tap of channel, following Jakes channel model, in the following section.

6.4.1 The Jakes channel

Assume that a flat fading mobile radio channel is Rayleigh fading and has a Jakes spectrum. Noisy observations of the complex channel $h(t)$, will be used to predict the complex tap $h(t)$. For a few different prediction ranges L, we will find the Δt as a function of the SNR that minimizes $V(\boldsymbol{\theta}, \Delta t, M, L, \gamma)$ in (6.68).

Covariance

The auto-covariance for a tap represented by the Jakes model for a mobile radio channel [11] is given by the zero order Bessel function of the first kind,

$$r_h(\tau) = \sigma_h^2 \mathrm{J}_0(2\pi f_D \tau), \tag{6.71}$$

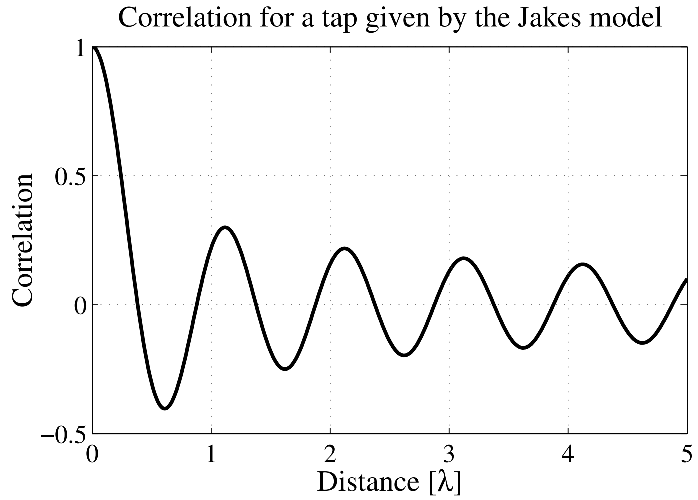

Figure 6.6: The correlation for a complex channel with Jakes spectrum as a function of distance. The correlation is $\mathrm{J}_0(2\pi l)$, where l is measured in wavelengths.

where f_D is the Doppler frequency. It thus depends on the product of Doppler frequency and delay

$$f_D\tau = \frac{f_c}{c}v\tau = \frac{v\tau}{\lambda}, \tag{6.72}$$

which is the distance measured in wavelengths. Here f_c is the carrier frequency, λ is the corresponding wavelength, c is the speed of light and v is the speed of the mobile antenna. If the delay spacing is measured in wavelengths instead of time, as in Figure 6.6, then the problem becomes independent of the speed of the mobile antenna. The correlation has zero-crossings with roughly half a wavelengths spacing. To exploit the correlation fully the delay spacing Δt thus should be smaller than about half a wavelength.

Delay spacing

Assume that the observed channel is contaminated by an additive white Gaussian noise. Then the covariance matrix (6.10) for the regressor has an additional diagonal loading term, $\mathbf{R}_\varphi = \mathbf{R}_h + \sigma_e^2\mathbf{I}$. For a given Δt in the regressor (6.70), we can use the covariance from (6.71) to form $\mathbf{r}_{h\varphi}$ and $\mathbf{R}_\varphi$ as in (6.9) and (6.12). The NMSE is then obtained using (6.13). To find the close to optimal delay spacing we calculate the NMSE (6.68) for all Δt_λ, where Δt_λ represent wavelengths instead of samples, from $10^{-5}\lambda$ up to 0.6λ in steps of $10^{-5}\lambda$, which is the channel sampling distance. A very small channel sampling distance is chosen to come close to a continuous choice of Δt. The influence of a specific channel sampling rate on the result is thus diminished. The Δt_λ that result in the smallest MSE is then selected. Figure 6.7 presents the best delay spacing for different numbers of coefficients M, as a function of the tap-to-estimation-error SNR. The prediction ranges are one eighth of a wavelength and one quarter of a wavelength. The length of the FIR filter is $(M-1)\Delta t_\lambda$, and the optimal lengths are shown in Figure 6.8.

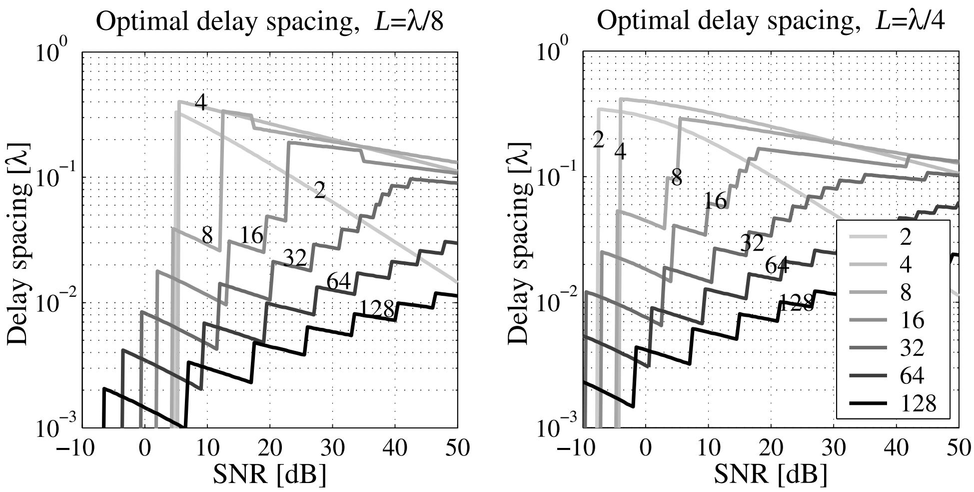

Figure 6.7: The optimal delay spacing for prediction of a channel tap, with Jakes spectrum, as a function of SNR, for a predictor with $M = \{2, 4, \ldots 128\}$ *coefficients.*

The curves for the optimal Δt_λ as a function of SNR are discontinuous with abrupt changes of Δt_λ. As seen in Figure 6.7, the optimal delay spacing is highly dependent on the SNR and on the number of coefficients M. The discontinuities can be explained by considering the MSE as a function of Δt_λ for a selection of SNRs, as in Figure 6.9, where $M = 8$ and $M = 32$. There are many shallow local minima in the MSE curve for a given SNR. Which of the local minima that is the global minimum changes when the SNR varies. The sudden jump from one minimum to another causes the abrupt changes in the optimal delay spacing that we see in Figure 6.7. There are more local minima in the curves corresponding to $M = 32$ than $M = 8$. This cause

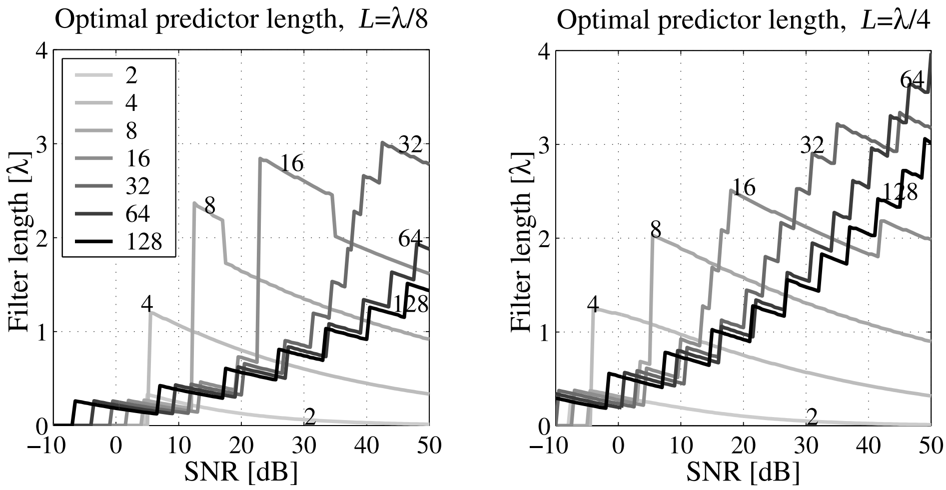

Figure 6.8: The optimal filter length $(M-1)\Delta t_\lambda$, for prediction of a channel tap, with Jakes spectrum, as a function of SNR, for a predictor with $M = \{2, 4, \ldots 128\}$ coefficients.

the discontinuities in Figure 6.7 to occur more frequently for predictors with many coefficients as compared to predictors with fewer coefficients. We also see from Figure 6.9 that the predictors are rather robust to the choice of Δt_λ, that is the bottom of the MSE curves is rather flat. This property is exploited in the following section.

A robust choice of Δt

The performance of the predictor is quite robust with respect to the choice of delay spacing, as long as it is of the right order of magnitude. The delay spacing doesn't need to be adjusted at the slightest change of SNR, to keep a close to optimal performance. We will here find delay spacings that will give reasonable prediction performance in the SNR range 10-40 dB, for the Jakes channel model corrupted by white noise.

The delay spacing that minimizes the NMSE on average, over an interval of SNR can be used when the SNR is only approximately known. To avoid that the performance in the domain of low SNR dominates, we consider the NMSE (6.68) in the dB scale,

$$\Delta t_a = \max_{\Delta t} \int_{\gamma_{\min}}^{\gamma_{\max}} 10 \log_{10}(V(\Delta t, \gamma)) d\gamma, \tag{6.73}$$

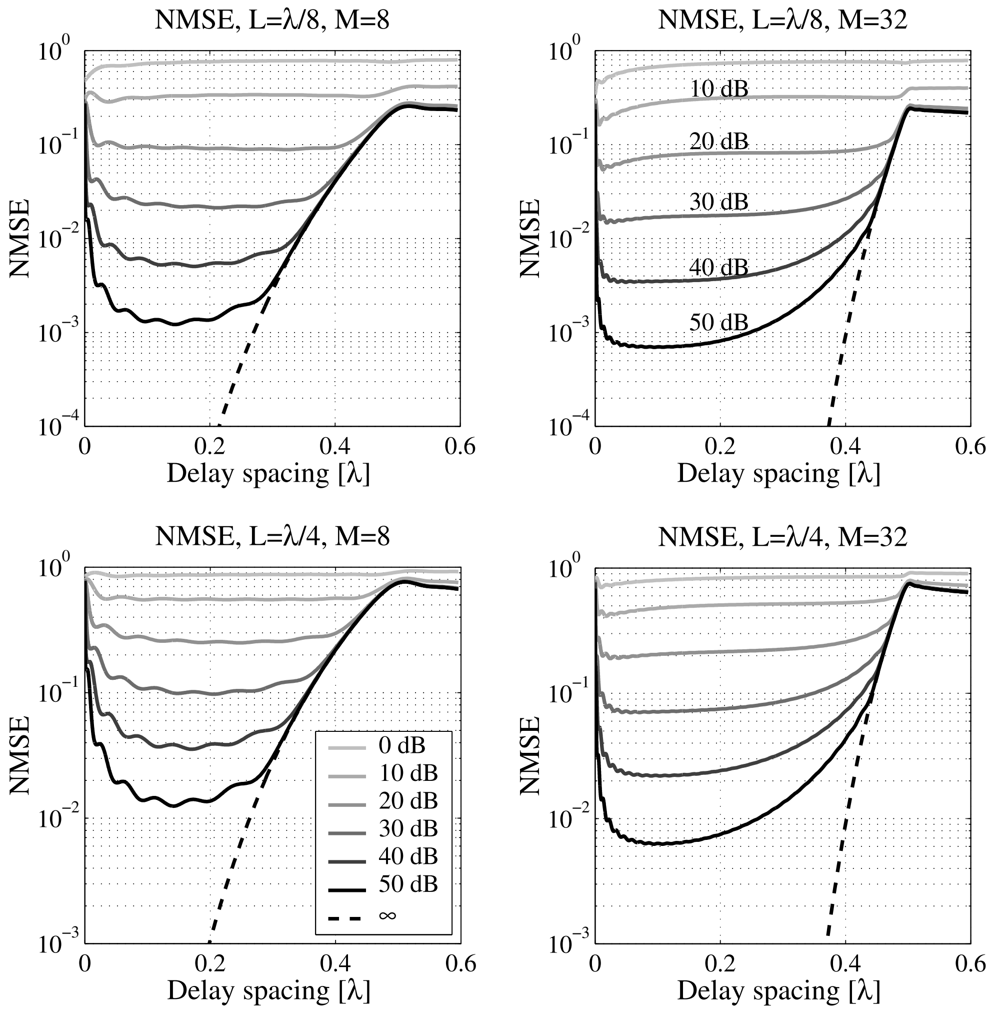

Figure 6.9: The NMSE as a function of Δt_λ for a selection of SNRs. The predictor has 8 or 32 coefficients and the prediction ranges are one eight and one quarter of a wavelength. There are many local minima in the NMSE curves. The position of the global minimum can change, when changing the SNR.

where Δt_a is the delay spacing that on average renders the lowest NMSE when the SNR is in the region $[\gamma_{\min}\ \gamma_{\max}]$. Table 6.1 lists the delay spacing that on average renders the best prediction performance for a Rayleigh fading channel with Jakes spectrum and additive white noise with SNR in the range of 10 to 40 dB for prediction ranges of one eighth, one quarter and one half wavelength.

Prediction range = 0.125λ								
No. coef. M	2	4	8	16	32	64	128	256
Δt_a [λ]	0.087	0.18	0.18	0.13	0.024	0.011	0.0053	0.0025
Av. dev. [%]	22	8	3	6	6	6	5	6
Max dev. [%]	53	19	10	29	17	13	11	14
Prediction range = 0.25λ								
No. coef. M	2	4	8	16	32	64	128	256
Δt_a [λ]	0.080	0.17	0.17	0.15	0.072	0.023	0.011	0.005
Av. dev. [%]	12	7	2	1	3	3	3	3
Max dev. [%]	30	17	6	3	8	8	7	8
Prediction range = 0.5λ								
No. coef. M	2	4	8	16	32	64	128	256
Δt_a [λ]	0.048	0.16	0.17	0.15	0.11	0.055	0.023	0.011
Av. dev. [%]	1	4	2	0.6	0.4	0.6	0.8	0.9
Max dev. [%]	3	10	4	2	0.8	1	2	2

Table 6.1: The spacing that on average renders the best NMSE (measured in dB) with SNR in the region 10-40 dB for the prediction ranges one eight, one quarter and one half of a wavelength. Max deviation denotes the largest deviation from the optimal NMSE curve, whereas average deviation is the average deviation from the optimal NMSE curve over the whole region.

A general trend is that the more coefficients in the filter, the shorter the delay spacing should be. The predictor needs to cover a sufficient interval of the measured channel, to achieve the best performance. With more coefficients, the samples can be taken denser from this interval and the delay spacing thus decreases. The opposite holds for the dependence on the prediction range: The further ahead we want to predict, the longer interval of the measured channel the predictor has to cover. This increases the delay spacing.

The robustness towards variation of the delay spacing increases with the prediction range. This is seen in Table 6.1, where both the average and the maximum deviation from the optimal performance, when using the robust delay spacing, decreases when the range increases. In Figure 6.7 we see that the optimal delay spacing for 4 and 8 coefficients, remain close to the robust choice over the range of SNRs (10-40 dB), for the prediction ranges of one eighth and one quarter of a wavelength. This stands in contrast to the optimal delay spacing, using more coefficients, which changes about one order of magnitude in the same SNR range. The robust delay spacing therefore renders quite close to optimal NMSE for the predictors using 4 and 8 coefficients. The predictors using more coefficients also show good robustness, in spite of the large spread in optimal delay. This is because the performance is less sensitive to the position of each individual sample, when using many densely placed coefficients in the predictor.

The delay spacing Δt_λ can be seen as the preferable sample rate of the channel to achieve the best prediction performance. Even though the dynamics of the channel in this example is band-limited and a sampling rate of half a wavelength would be sufficient in theory, the additive noise forces us to sample at a higher rate. If the channel is sampled at a fixed rate t_s (measured in wavelengths), then the robust delay spacing Δt_a can give the sub-sampling factor for the predictor as $\Delta t_a/t_s$, rounded to the closest integer. This is a good choice for the sub-sampling measured in samples.

Note that no noise-reduction has been applied. There is thus a clear advantage for the predictors using many coefficients, that have a large number of degrees of freedom which can be used for noise suppression. The assumption of white noise is also questionable, at least when the delay spacing is on the order of a thousandth of a wavelength or smaller. The channel samples are then taken so densely that it is unrealistic to assume that the noise on adjacent channel samples would be uncorrelated. In the performance evaluation in Section 6.7 the smallest delay spacing will be limited and noise reduction will be applied. The results in this section are still important as they show how crucial the choice of a sensible delay spacing is.

6.5 The Sub-sampled Predictor and Aliasing

In the previous section we saw that it can be beneficial to space the samples in the predictor with a certain delay, that depend on the prediction range, the SNR and the number of predictor coefficients. If the sampling rate of the channel is higher than that delay, the predictor will be sub-sampled. The frequency response of a sub-sampled linear predictor like (6.6)

$$\hat{h}(t+L|t) = \theta(q^{-\Delta t})y(t), \tag{6.74}$$

is periodic, as seen in Figure 6.10 of Example 6.2.

Example 6.2 A sub-sampled predictor of a Jakes channel

Consider FIR-prediction of a Jakes model channel that is corrupted by additive white noise resulting in an SNR of 10 dB. The channel is sampled at a rate of 100 samples per traveled wavelength and is thus highly over-sampled with an OSR of 50. A FIR-Wiener-predictor with 16 coefficients

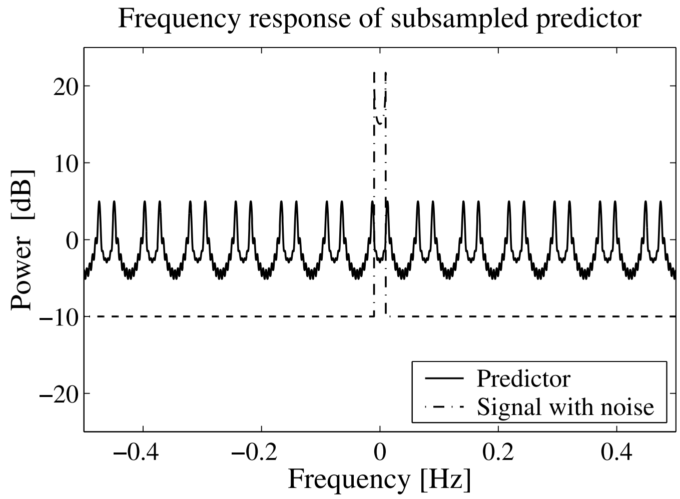

Figure 6.10: The frequency response of a sub-sampled FIR-Wiener predictor with 16 coefficients for a Rayleigh fading tap with Jakes spectrum. The SNR is 10 dB, the Doppler frequency is 0.01 Hz and the sub-sampling rate is 13.

is designed for a prediction range of a tenth of a wavelength. The delay spacing is taken from Table 6.1 as $\Delta t_a = 0.13\lambda$. The sub-sampling factor is thus 13, with an OSR of almost 4. The predictor achieves a prediction NMSE of 0.14. The frequency response in Figure 6.10, for the optimal FIR-predictor is periodic.

To avoid the periodic amplification of noise in the prediction, an FIR-Wiener smoother with 128 coefficients and a smoothing lag of zero samples is applied in a first noise reduction step. Then a sub-sampled FIR-Wiener predictor is designed to use the noise reduced signal. The combined smoother and predictor system has a frequency response with a low-pass character, avoiding amplification of the noise outside the relevant frequency band, as seen in Figure 6.11. This predictor achieves a prediction NMSE of 0.045, almost 5 dB better than the predictor working

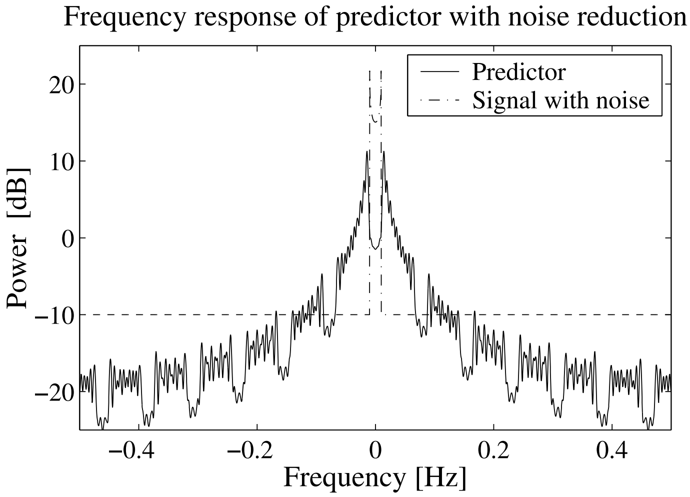

Figure 6.11: The frequency response of a sub-sampled FIR-predictor with 16 coefficients combined with a Wiener-smoother with 128 coefficients and zero smoothing lag.

directly on the noisy signal.

The use of the sub-sampled predictor thus leads to aliasing. The predictor amplifies the observed signal in frequency regions where the noise dominates. The noise-reduction pre-processing step solves this problem. The smoother reduces the noise level outside the frequency band of interest and thus works as an anti-aliasing filter with close to linear phase and 0 dB amplification in the passband region combined with a low delay. The optimization of the predictor coefficients should be performed using the statistics of the noise reduced signal to obtain the best possible performance.

6.6 Model Based Prediction using AR or ARMA Models

The direct prediction approach, designing L-step FIR-predictors with a least squares algorithm, using filtered regressors, is a very robust prediction method that is based on the FIR-Wiener filter approach. An alternative is the model based, also called indirect, predictor. The model based predictor relies on an accurate estimate of a model for the process generating the signal, instead of estimation of predictor coefficients. This model can then be used in a Kalman predictor to obtain predictions at any range that is a multiple of the chosen delay spacing [55]. The steady state solution of the Kalman predictor is also given by the causal IIR Wiener predictor. The iterative predictor, presented in Section 6.3, can be interpreted as a model based method, under the assumption that the channel tap that is to be predicted can be measured without estimation error.

From the discussion about channel modeling, Section 2.5, we know that the taps of a sampled channel can be approximately modeled by an ARMA-process. It should be noted that the channel is not generated by an ARMA-process but that the dynamics can be approximately described by such a process. In that sense there exists no true ARMA-model, only approximations that fit more or less well to the dynamics of the observed channel. Wiener-predictors for the channel can be designed, once ARMA-models for the dynamics of the taps are estimated. The model-system and the predictor as seen in Figure 6.12 is similar to the model used for noise reduction in Section 5.3. For $t = s\Delta t$, $s = 1, 2, \dots$, the sub-sampled signal is $y_s(s) = y(s\Delta t)$.

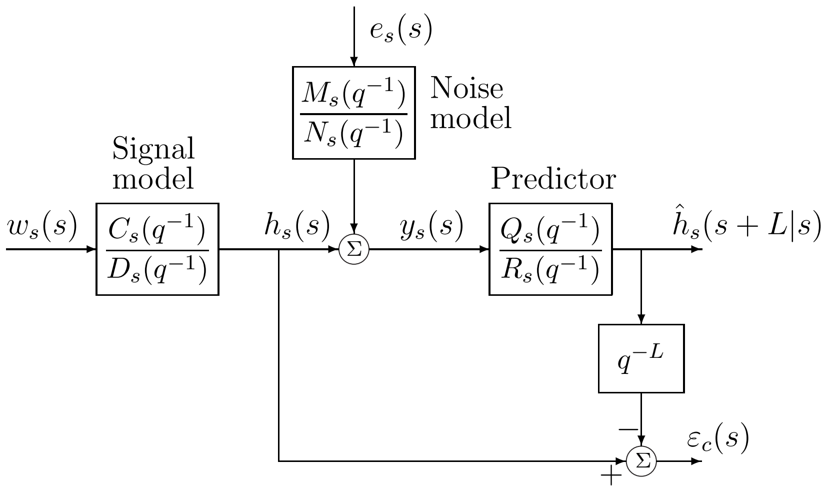

Figure 6.12: The full prediction system with signal and noise model for predictor design. The predictor and the ARMA-models may be sub-sampled. The model used for prediction is similar to the model in Figure 5.3 used for noise reduction.

As the channel does not stay stationary over long distances, it is not feasible to estimate models with a large number of parameters with any accuracy. The predictor thus has to rely on low order models of the dynamics. The delay spacing Δt plays a slightly different role here as compared to the direct predictor. The question now is how densely the channel has to be sampled, for the dynamics to be approximated by an AR or ARMA-model of moderate order. If the dynamics of the channel is perfectly band-limited, then it is sufficient to sample at the maximum Doppler frequency (then the samples are spaced by half a wavelength). The corresponding predictor would then only be able to predict the channel half a wavelength ahead, or longer. To be on the safe side, oversampling is recommended. Reasonable results are obtained with a delay spacing of a tenth of a wavelength, that is five times oversampling.

A to high oversampling ratio for the estimated models should be avoided. When AR or ARMA-models are fitted to channel data that are highly oversampled, the poles of the models cluster around $z = 1$ in the complex domain. Then the fine structure of the Doppler spectrum can not be exploited, which leads to poor prediction performance for longer ranges (beyond a tenth of a wavelength). In these cases sub-sampled AR and ARMA-models should be used. The poles will then be slightly drawn back from $z = 1$ and are spread over a larger angle around the unit circle $|z| = 1$. The sub-sampling factor is preferably chosen so that the resulting delay spacing is on the order of tenth of a wavelength. Then the passband of the model will take up 20% of the spectrum.

6.6.1 Doppler spectrum estimation

An estimated ARMA-model can be seen as as a Doppler spectrum estimate,

$$|H(\mathrm{e}^{\jmath\omega})|^2 \approx \frac{\hat{C}_s(\mathrm{e}^{\jmath\omega})}{\hat{D}_s(\mathrm{e}^{\jmath\omega})} \frac{\hat{C}_s^{\,*}(\mathrm{e}^{-\jmath\omega})}{\hat{D}_s^{\,*}(\mathrm{e}^{-\jmath\omega})} \sigma_{w_s}^2, \tag{6.75}$$

where the subscript s denotes that the model is sub-sampled. As the model is estimated from noisy observations, the estimated ARMA-model will also include modeling of the noise. This result in overestimation of the power in the Doppler spectrum, especially in the frequency regions outside plus/minus the maximum Doppler frequency. The model error can deceive the Wiener-predictor to amplify frequency regions where the noise dominates. With the proper design of the noise model design, this problem can be avoided. Alternatively spectral subtraction could be applied. This is however not trivial as the ARMA-model can have deep notches at some frequencies.

6.6.2 The noise model

Before noise reduction the additive noise is close to white as seen in Section 5.2. After noise reduction the additive noise will be colored by the smoother that acts as low-pass anti-aliasing filter. The coloring of the noise is thus known and a model $M_s(q^{-1})N_s^{-1}(q^{-1})$ for the sub-sampled dynamics of the noise could be used in the design of the predictor. We will however use the noise model to robustify the predictor. The estimated signal model indicates that there is more signal energy at high frequencies than there actually is. A noise model designed to have more noise at those frequencies than there actually is, will reduce the amplification of those frequency regions by the predictor.

A suitable and simple noise model is a white noise, where the variance of the noise is a tuning parameter. The noise gain of the smoother, used as anti-aliasing filter, times the original estimation error variance, usually renders IIR-Wiener predictors with good prediction performance. The noise gain of an FIR smoother is obtained as the squared sum of the coefficients, that is $(\mathbf{w}^{[m]})^H\mathbf{w}^{[m]}$ where m is the smoothing lag and $\mathbf{w}^{[m]}$ is obtained as in equation (5.6).

6.6.3 Model estimation

In Section 6.2.4 the least squares algorithm for estimation of predictor coefficients for the direct FIR predictor is recapitulated. That method can of course be used to estimate the parameters of an AR-model too, if only one smoothing lag is used in the regressor (preferably $m_{\max}$). The AR-model has to be monitored to make sure that the model is stable. Poles outside the unit circle have to be reflected inside, using spectral factorization.

There are a multitude of methods for estimating ARMA-models [50]. Even though the model is sub-sampled, the whole data set should be used when estimating the model. Generally the $C_s(q^{-1})$ polynomial is hard to estimate from a limited amount of data. As the dynamics of the channel is time varying, the estimation interval will be limited to a few meters. At a carrier frequency of 2 GHz this corresponds to estimation intervals of no more than 10-30 wavelengths. Under the assumption of perfect band limited Doppler spectrum it would be sufficient with 20-60 samples in this interval. The effective number of samples available for model estimation will thus be low. The model order thus has to be kept low, if ARMA-models are at all to be used for channel prediction.

6.6.4 Use of noise reduction

When designing model based predictors, it is not trivial how to include the use of smoothed regressors with different smoothing lags. The straight forward use of smoothed regressor in the direct FIR predictor can also be applied here, where the coefficients of the $Q_s(q^{-1})$ polynomial take the place of the FIR-filter coefficients and the autoregressive part with the $R_s(q^{-1})$ polynomial is appended afterwards. When evaluated on the data sets of Chapter 3, this does however not render an improved performance for the prediction of channels, as compared to using smoothing lag $m = 0$. The Kalman filter representation of the IIR Wiener-predictor helps us to interpret this result. The Kalman filter uses the model of the dynamic of the channel to improve the estimated states. There is thus an internal noise reduction mechanism in place and the extra noise reduction provided by using a larger smoothing lag in the noise reduction offers only a minor improvement to this. It is thus sufficient to use noise-reduction with lag 0 to provide the necessary anti-aliasing filtering.

6.7 Results

The performance of the predictors discussed in this Chapter is evaluated on the measured channels described in Chapter 3. These measured channels are frequency selective and thus consist of more than one tap. Each tap is predicted individually and then the channel prediction error is evaluated.

Channel SNR

The channel SNR is the ratio between the average estimation error power and the channel tap power. For a channel where all the taps have zero mean the overall channel-to-estimation-error ratio (channel SNR) is

$$\gamma_h = \frac{\sum_{k=0}^{M-1} \sigma_{h_k}^2}{\sum_{k=0}^{M-1} \sigma_{e_k}^2}, \tag{6.76}$$

where $\sigma_{e_k}^2$ is the variance of the estimation error on each tap and $\sum_{k=0}^{M-1} \sigma_{h_k}^2$ is the average gain of a channel (also called noise gain).

When the channel estimation interval is short (small time frequency product), then all the taps in the estimated channel are corrupted by approximately the same amount of noise power $\sigma_e^2 = \sigma_{e_k}^2, \forall k$, as we saw in

Section 4.3. The channel SNR can then be expressed as the average of the individual tap SNRs,

$$\gamma_h = \frac{\sum_{k=0}^{M-1} \sigma_{h_k}^2}{M\sigma_e^2} = \frac{1}{M}\sum_{k=0}^{M-1} \gamma_{h_k}. \tag{6.77}$$

Performance measures

The complex channel prediction error is

$$\boldsymbol{\varepsilon}_c(t) = \mathbf{h}(t) - \hat{\mathbf{h}}(t|t-L). \tag{6.78}$$

The performance of the channel predictor is measured as the overall NMSE

$$\tilde{V} = \frac{\mathrm{E}\left\{\sum_{k=0}^{M-1} |\varepsilon_{c_k}(t)|^2\right\}}{\mathrm{E}\left\{\sum_{k=0}^{M-1} |h_k(t)|^2\right\}}. \tag{6.79}$$

In simulations or evaluation on measured data, the expectations are approximated by averages over time. For a finite validation set consisting of channel samples $1 \dots N$, the NMSE is thus approximated by the ratio between the sample averages

$$\tilde{V} \approx \frac{\frac{1}{N}\sum_{t=1}^{N}\sum_{k=0}^{M-1} |\varepsilon_{c_k}(t)|^2}{\frac{1}{N}\sum_{t=1}^{N}\sum_{k=0}^{M-1} |h_k(t)|^2} = \frac{\sum_{t=1}^{N} \boldsymbol{\varepsilon}_c^H(t)\boldsymbol{\varepsilon}_c(t)}{\sum_{t=1}^{N} \mathbf{h}^H(t)\mathbf{h}(t)}. \tag{6.80}$$

In the measurements only noisy observations of the channels are available. As the true channels are unknown, smoothed estimates of the channels are used instead to obtain an estimate of the vector of channel tap prediction errors as outlined in (6.42),

$$\boldsymbol{\varepsilon}_c(t) \approx \tilde{\mathbf{h}}(t|t+m) - \hat{\mathbf{h}}(t|t-L), \tag{6.81}$$

where m is a smoothing lag that renders close to optimal noise reduction performance.

The Doppler spread

The normalized Doppler spread for a tap with band-limited dynamics, defined as

$$B_D = \frac{1}{f_D}\sqrt{\frac{\int_{-f_D}^{f_D}(f-f_m)^2|H_k(f)|^2 df}{\int_{-f_D}^{f_D}|H_k(f)|^2 df}} \tag{6.82}$$

is a measure for the variation of the tap. Here $|H_k(f)|^2$ is the Doppler spectrum for tap k and f_m is the average Doppler frequency defined as

$$f_m = \frac{\int_{-f_D}^{f_D} f|H_k(f)|^2 df}{\int_{-f_D}^{f_D} |H_k(f)|^2 df}. \qquad (6.83)$$

The upper limit for the normalized Doppler spread is one and for Jakes's spectrum $B_D = 1/\sqrt{2}$ [37]. A small normalized Doppler spread means that the energy in the spectrum is highly centered. A large spread means that there can be energy all over the spectrum. The performance of the prediction is seen to depend on this measure. The estimated normalized Doppler spread is generally overestimated for the noisiest taps.

6.7.1 Simulation

The expected performance for prediction of a flat Rayleigh fading channel with Jakes spectrum will indicate the performance on true measured channels.

Robust delay spacing

In Section 6.4 a robust delay spacing for prediction of a tap described by the Jakes model corrupted by white noise was presented. If we include noise reduction in a first step, both the optimal and the robust choices for the delay spacing are altered.

The Wiener smoothers, used in noise reduction, depend on the OSR of the channel and also on the SNR. The performance of the predictor using smoothed regressors is here evaluated for an OSR of 50 and 25 (the Doppler frequencies normalized by the channel sampling frequency are then $f_D = 0.01$ and $f_D = 0.02$ respectively). FIR-Wiener smoothers with 128 coefficients and smoothing lags [0 2 5 10 20] are used in the noise reduction. Using the results for the covariance of filtered regressors in (6.38) and (6.41), the performance for the optimal FIR-predictor can be calculated using (6.13). The NMSE is calculated for SNRs in the range 0-50 dB and for prediction ranges between 0.1 and 1 wavelength, varying the delay spacing.

Our finding is that the performance is rather insensitive to the choice of delay spacing. Just as in Section 6.4 a robust choice of the delay spacing, that is a function of the number of coefficients of the predictor and of the prediction range, can be obtained. Tables 6.2 and 6.3 list the robust choices

Robust delay spacing for OSR=50						
Range	Number of coefficients M					
L [λ]	2	4	8	16	32	64
0.1	0.01	0.01	0.05	0.20	0.20	0.20
0.2	0.01	0.01	0.05	0.20	0.19	0.19
0.3	0.01	0.01	0.05	0.20	0.19	0.19
0.4	0.01	0.01	0.06	0.20	0.19	0.19
0.5	0.01	0.01	0.06	0.20	0.19	0.18
0.6	0.43	0.01	0.06	0.19	0.19	0.18
0.7	0.19	0.01	0.06	0.19	0.19	0.18
0.8	0.01	0.01	0.06	0.19	0.19	0.18

Table 6.2: Robust choice for delay spacing (measured in wavelengths) in the FIR predictor using smoothed regressors for a tap described by the Jakes model. The channel is sampled with 100 samples per wavelength and a Wiener-smoother with 128 coefficients and smoothing lags [0 2 5 10 20] are used for noise reduction. One sample corresponds to a traveled distance of 0.01 wavelength.

Robust delay spacing for OSR=25						
Range	Number of coefficients M					
L [λ]	2	4	8	16	32	64
0.1	0.02	0.02	0.10	0.22	0.22	0.22
0.2	0.02	0.02	0.10	0.22	0.20	0.20
0.3	0.02	0.02	0.10	0.22	0.20	0.20
0.4	0.02	0.02	0.10	0.20	0.20	0.20
0.5	0.02	0.02	0.10	0.20	0.20	0.20
0.6	0.42	0.02	0.10	0.20	0.20	0.20
0.7	0.18	0.02	0.10	0.20	0.20	0.20
0.8	0.02	0.02	0.10	0.20	0.20	0.20

Table 6.3: Similar to Table 6.2 but for an OSR of 25. The channel is sampled with 50 samples per wavelength (one channel sample corresponds to 0.02 wavelength).

for delay spacing of the FIR-predictors for channels with 50 and 25 times oversampling when a bank of FIR smoother are applied in the first step. The SNR is assumed to be in the range 10-50 dB. In the following performance evaluations, these robust choices of the delay spacing will be used for the FIR predictor. Delay spacings for OSRs not covered by the tables are obtained by interpolation or extrapolation of the values in the tables.

Prediction performance

Figure 6.13 shows how predictable a tap described by the Jakes model is with and without smoothed regressors. The OSR is 50 and the predictor has eight coefficients and a robust choice of delay spacing, given by Table 6.2. The noise reduction is obtained using a bank of FIR smoother with 128 coefficients and the regressor is formed as in Section 6.2.3. By using smoothed regressors the FIR predictor can obtain the same performance as the FIR predictor using no noise reduction at 7-9 dB lower SNR.

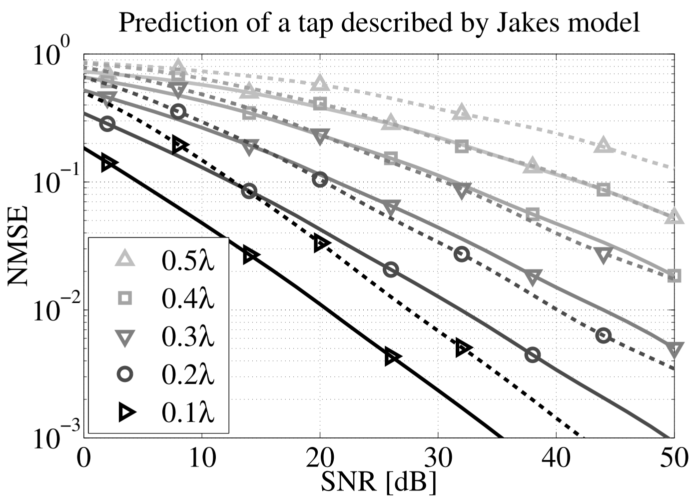

Figure 6.13: The prediction NMSE as a function of SNR for different prediction ranges. The tap is described by the Jakes model. The FIR predictor has 8 coefficients and the performance is shown for smoothed regressor (solid lines) and no noise reduction (dashed line).

An increase of the number of coefficients from 8 to 32 in the predictor result in only a minor decrease in the NMSE, as seen in Figure 6.14. The major gain is attained by the use of noise reduction. The prediction range can be extended by a tenth of a wavelength at the same performance as for not using the noise reduction.

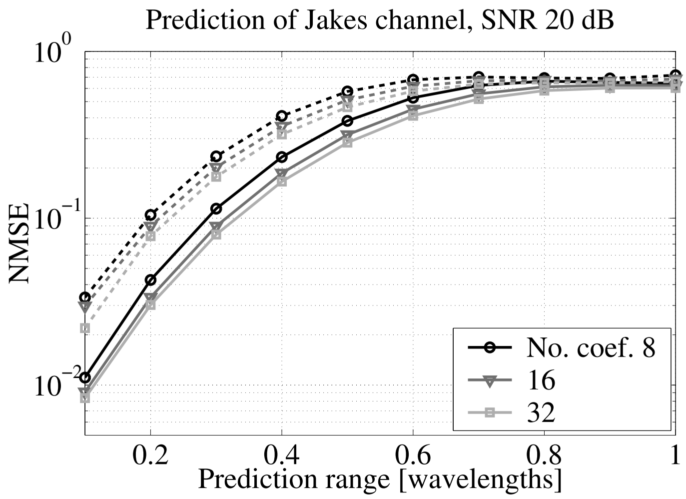

Figure 6.14: The prediction NMSE at 20 dB for a tap described by the Jakes model as a function of prediction range. Solid and dashed lines corresponds to smoothed and noisy regressors respectively.

6.7.2 Channel prediction of measured channels

The measurements in the data base described in Chapter 3 are used to evaluate the performance of the channel prediction methods described in this Chapter.

Noise reduction

A bank of FIR-Wiener smoothers for a Jakes channel, as in Section 6.2.3, has been applied to all taps. The SNR and Doppler frequency for the taps are estimated as in Section 5.2. A ten percent safety margin is added to the estimated Doppler frequency to avoid a too small bandwidth of the smoother. The bank of smoothers has smoothing lags [0 20 5 10 20], and each smoother has 128 coefficients.

As only noisy observations of the channel taps are available, the true tap is taken as the smoothed tap obtained with smoothing lag 20. This is a good approximation of the true taps for SNRs above 0 dB.

Selection of data

A total of 41 measurement locations, out of 80, are used in the evaluation. The selection of the measurement locations was based on three criteria:

- Only measurements where the estimated average channel-to-estimation-error power ratio (channel SNR) is above 10 dB are used.
- Measurements where the speed was below 30 km/h are not used as then the total measurement would cover a distance of less than 8,5 wavelength, which is too short to obtain reasonable prediction performance statistics of the channel.
- The algorithms of this Chapter are designed for channels with stationary statistics. If abrupt changes were present within the data windows of 0.14 s, it would be appropriate to use different predictor parameters before and after the change. To avoid such situations in the evaluation, a few channels where the average power changed by more than 3 dB from the first quarter of the data to the last quarter, were ruled out.

Predictor design

As most of the prediction performance for a Jakes channel using smoothed regressors is obtained using only eight coefficients, the FIR predictor for the the measured channels, using smoothed regressors, has eight coefficients. The delay spacing is taken from the Tables 6.2 and 6.3. The sample covariance estimates of the correlation matrix for the regressor $\hat{\mathbf{R}}_\varphi$ from (6.46) and the cross-correlation vector for regressor and signal $\hat{\mathbf{r}}_{x\varphi}$ from (6.47) are used in (6.12) to estimate the predictor coefficients. This simply corresponds to LS-estimation of the predictor coefficients as in (6.45). Independent of prediction range, 900 samples are used in the target vector for the training. The validation interval was on the order of 400 down to 100 samples, depending on prediction range and Doppler frequency.

The model based predictors have a delay spacing of a tenth of a wavelength. The AR-models are estimated using LS. All poles that fall outside the unit circle are reflected inside, using spectral factorization. The ARMA-models are obtained using the prediction error method (PEM). The variance of the measurement noise, used as a design variable in the Kalman filter, is set to the variance of the estimation error, filtered by a FIR smoother using lag zero. Out of the 1430 samples 900 samples are used for identification of the models. The remaining 530 samples are used for validation. The model estimation interval corresponds to at least 5,4 wavelengths of traveled distance. The model based Kalman predictors only utilize the signals from the FIR smoother with lag zero.

Prediction performance on the taps

The tap prediction is highly dependent on the tap SNR and on the normalized Doppler spread as seen in Figure 6.15, where the prediction NMSE for all the individual taps of the 41 measurements are plotted. The predictor is the FIR predictor with eight coefficients using smoothed regressors, as described above. The dependence of the NMSE on the tap SNR and on the normalized Doppler spread are shown in the left and right columns respectively. Plotted together with the scatter plot for the NMSE dependence on the SNR is the corresponding theoretical curve for the prediction NMSE for a tap described by the Jakes model. The OSR of the tap is 50, corresponds to a vehicle speed of about 50 km/h in these measurements, using the same noise reduction and predictor as for the measurements are used.

A tap described by the Jakes model has a $B_D \approx 0.71$. In the measured channels, taps with both lower and higher normalized Doppler spreads than 0.71 are present. Generally the taps with a lower B_D have a lower prediction NMSE.

For a prediction range of 0.1 wavelength, the tap prediction NMSE is seen to strongly depend on the SNR. It actually follows the theoretical curve for a tap described by the Jakes model quite well. The spread in performance increase when the prediction range increase. On the other hand, the taps with small Doppler spread remains well predictable even at a prediction range of half a wavelength. This group of taps is also seen in the SNR plots as a smaller cluster with better NMSE below the large cloud of taps. Thus, for short prediction ranges the tap SNR is the most important factor for the performance. For longer prediction ranges, the dynamics of the taps play a more important role.

Prediction performance on the channel

The median prediction NMSE (6.80) over the 41 measured channels is seen in Figure 6.16. The difference in median performance between the applied prediction methods is quite small,whereas the spread in predictability of the channels among the different measurements is quite dramatic. The Kalman predictor based on the estimated ARMA-model is actually a little bit worse than the predictor based on an AR-model. This is solely due to the sig-

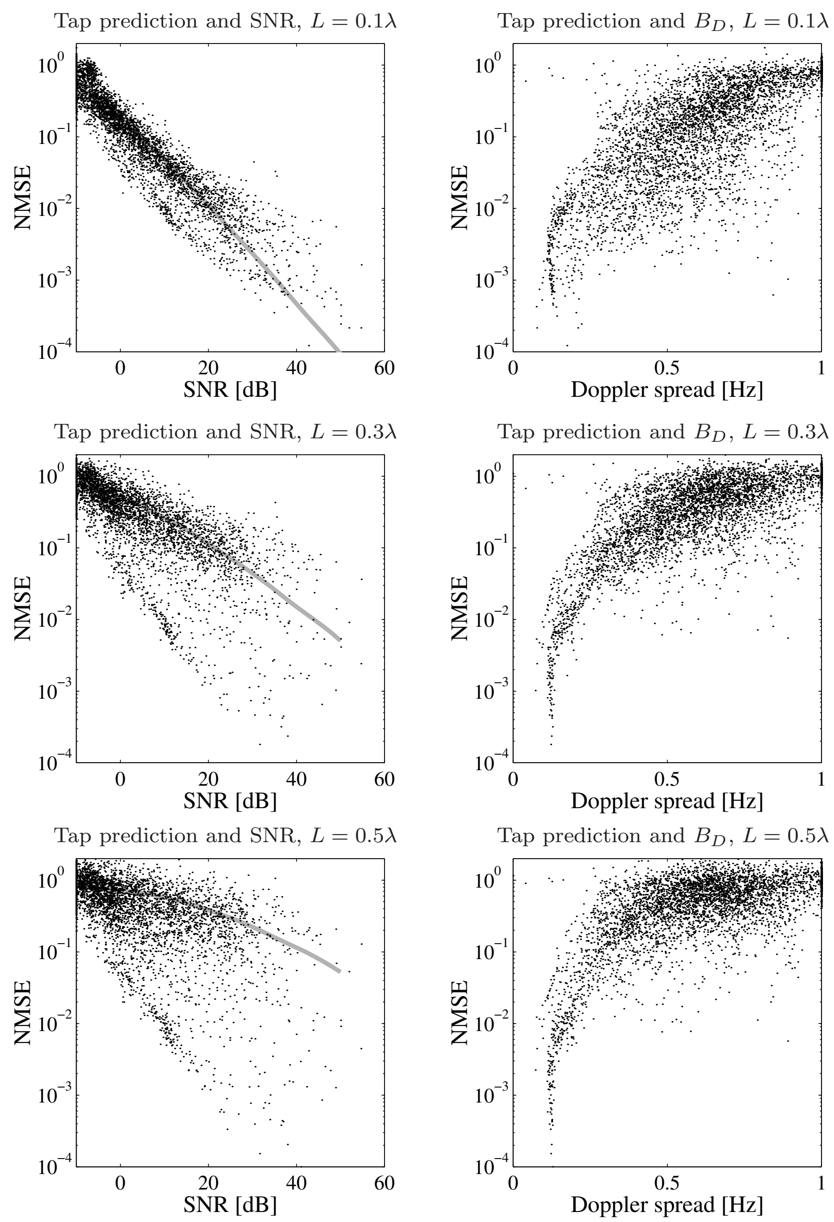

Figure 6.15: The tap prediction NMSE as a function of SNR and normalized Doppler spread for the individual taps for the prediction ranges 0.1, 0.3 and 0.5 wavelengths. The gray lines are the theoretical NMSE for prediction of a tap described by the Jakes model, for an OSR of 50 as in Figure 6.13.

nificant model estimation error obtained on such short estimation intervals.

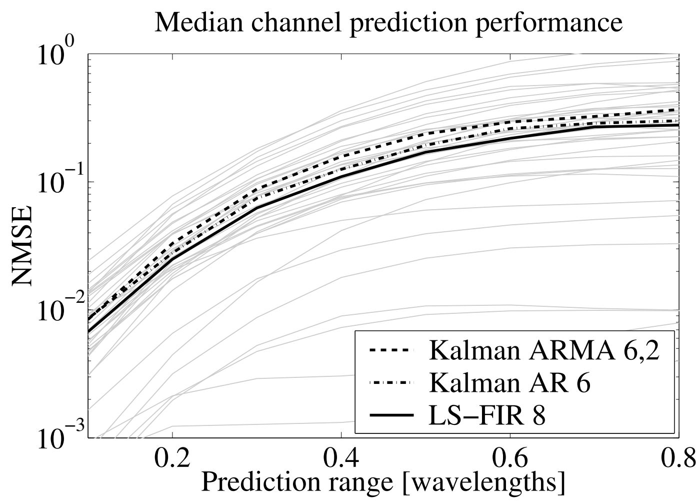

Figure 6.16: The median channel prediction NMSE for indirect model-based Kalman predictors based on ARMA 6,2 (dashed line) and AR 6 (dash-dotted line) models for the dynamics. These predictors only use noise reduction with smoothing lag zero. The direct LS-estimated FIR predictor with eight coefficient using smoothed regressors with as large lags as possible (solid line) is slightly better than the others. The gray lines are the NMSE for the 41 individual channels, using the FIR predictor.

The scatter plots in Figures 6.17, 6.18 and 6.19 show the prediction NMSE as a function of channel SNR, for the 41 different measured channels. The FIR predictor with eight coefficients is used. The attained prediction NMSE strongly depends on the channel SNR. For longer prediction ranges the dependence on the channel SNR decreases, as then the dynamics for the individual channels play a larger role. This results in a larger spread in prediction NMSE, as seen in Figure 6.19.

6.8 Conclusion

- Noise reduction is crucial for the performance of the channel prediction.
  - A bank of Wiener-smoothers with different lags can be used to produce regressors where all the regressor variables have the at

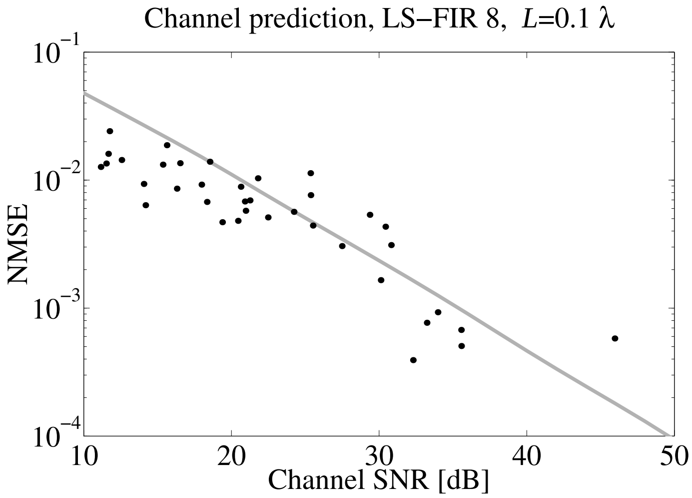

Figure 6.17: A scatter plot of the channel prediction NMSE at $L = 0.1\lambda$ as a function of channel SNR when using an FIR predictor with 8 coefficients. Smoothed regressors are used. The gray line is the theoretical NMSE for prediction of one single tap described by the Jakes model for an OSR of 50, as in Figure 6.13

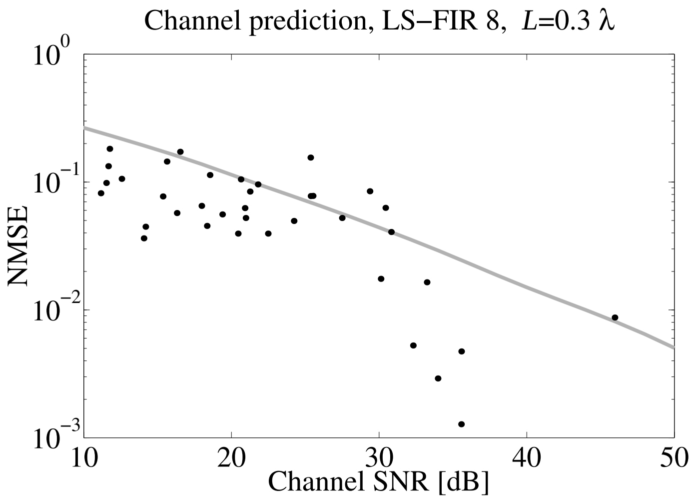

Figure 6.18: scatter plot of the channel prediction NMSE just as in Figure 6.17 but for a prediction range of 0.3λ.

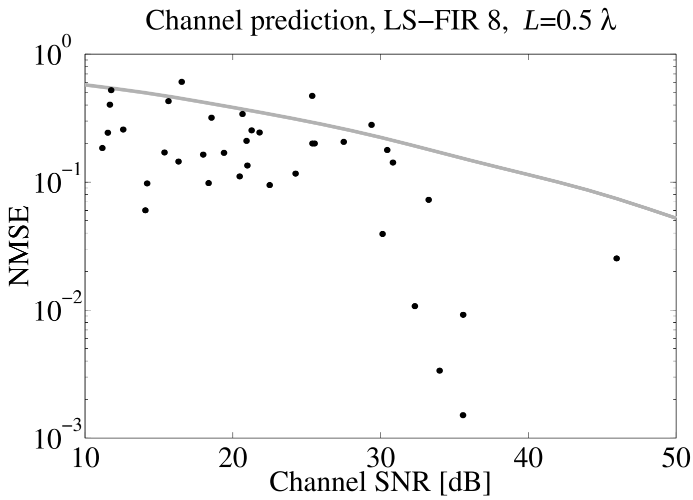

Figure 6.19: Just as Figure 6.17 but for a prediction range of half a wavelength. The spread in performance is much larger for the longer prediction range.

the time best available noise reduction.

- The Wiener-smoothers based on the Jakes model act as anti-aliasing filters, facilitating for sub-sampled predictors.
- Noise reduction reduces the number of predictor coefficients that are necessary for attaining a given performance.

- Iterative prediction is only optimal in the case of no estimation error on the channel.
- Instead of iterative prediction, model based Kalman or Wiener predictors should be used, taking the tap estimation error into account.
- The tap-prediction performance of an FIR-predictor can be highly improved by use of smoothed regressors.
- The delay spacing of the predictor strongly influences the performance of the prediction.
- Robust choices of the delay spacing have been obtained for both smoothed and noisy regressors.
- Performance evaluation on measured channels:

- There is a large spread in prediction performance among the measured channels.
- The channel prediction performance is highly dependent on the channel SNR.
- For longer prediction ranges the spread in performance is larger, as then the dynamics of the individual taps play a major role in addition to the channel SNR.
- The performance for the direct FIR predictor using smoothed regressors and the Wiener predictor using AR- and ARMA-models is quite similar.

Chapter 7

Power Prediction Based on Linear Regression

7.1 Introduction

In this Chapter we will cover how to go from linear prediction of the complex valued tap to quadratic prediction of the power. Direct use of the absolute square of the complex valued prediction is shown to render a biased power predictor (Section 7.2.1). The unbiased quadratic predictor, that is optimal in the MSE sense, is derived in Section 7.2.2 and it is shown that the same prediction coefficients that are optimal for the linear prediction of the complex valued tap are optimal also for the unbiased quadratic predictor for the power, under the assumption of Rayleigh fading [61]. The joint probability density function of the true and predicted power, a bivariate gamma distribution, is derived in Section 7.4 for this predictor. The distributions will be used in Chapter 8 in the design of a link adaptation system.

In Section 7.5 the performance of the unbiased quadratic predictor is compared to a linear power predictor, that uses old and present observations of the observed power instead of the observed complex valued time series. The linear power predictor is expressed in terms of the covariance sequences for the complex valued time series and noise. The correlation function for the power is obtained as the absolute square of the correlation function for the tap. It thus decays much faster than the correlation function for the complex tap, which result in a prediction performance that is worse than for the unbiased quadratic predictor.

In Section 7.6 the estimate of the local average of the power and its use as a power prediction is analyzed. To obtain reliable estimates of the average power the estimation interval need to be several tenths of wavelengths.

The proposed power predictors are evaluated on measured mobile radio channels. The results, presented in Section 7.7, are compared to the theoretical results for prediction of the power of a tap described by the Jakes model. The attained power prediction NMSE is found to be highly dependent on the channel estimation error.

7.2 Power Prediction using Complex Regressors

It is the power of the signal $h(t)$ in (6.1) that is of main interest to our investigation, that is, the absolute square of the complex value,

$$p(t) = |h(t)|^2. \tag{7.1}$$

The average error and the MSE of FIR power predictors that use the regression vector (6.5), containing delayed noisy measurements, are studied in this section. The signal $h(t)$ and the noise $e(t)$ are assumed to have zero mean and circular Gaussian distributions.

7.2.1 The absolute square: A biased quadratic predictor

We first examine the use of $|\hat{h}(t+L|t)|^2$, with $\hat{h}(t+L|t)$ given by (6.4), as a power predictor,

$$\hat{p}_b(t+L|t) = |\hat{h}(t+L|t)|^2 = \boldsymbol{\theta}^H \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\boldsymbol{\theta}. \tag{7.2}$$

This predictor utilizes quadratic functions of the regressor variables and is thus a nonlinear predictor. The power prediction error is then given by

$$\varepsilon_{p_b}(t) = |h(t)|^2 - |\hat{h}(t|t-L)|^2. \tag{7.3}$$

This predictor for the power is used for channel prediction in e.g. [22] and [55].

The bias

The average error will be

$$\begin{aligned} \mathrm{E}\{\varepsilon_{p_b}(t)\} &= \mathrm{E}\{|h(t)|^2 - |\hat{h}(t|t-L)|^2\} \\ &= \sigma_h^2 - \boldsymbol{\theta}^H \mathrm{E}\{\boldsymbol{\varphi}(t-L)\boldsymbol{\varphi}^H(t-L)\}\boldsymbol{\theta} \\ &= \sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}. \end{aligned} \tag{7.4}$$

If the optimal coefficients for the complex predictor (6.12) are used in the expression (7.4) for the average error, then we obtain

$$\mathrm{E}\{\varepsilon_{p_b}(t)\} = \sigma_h^2 - \mathbf{r}_{h\varphi}\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi} = \sigma_{\varepsilon_c}^2. \tag{7.5}$$

The absolute square of the complex prediction is thus a *biased power predictor*. When using the optimal prediction coefficients for the complex predictor, this bias becomes equal to the variance of the complex prediction error. The power of $h(t)$ is thus underestimated, which could be seen already in (6.18).

The MSE for the biased power predictor

When using the FIR predictor (6.4) the MSE for the power prediction error in (7.3) becomes

$$\begin{aligned}
&\mathrm{E}\{|\varepsilon_{p_b}(t)|^2\} = \mathrm{E}\{(|h(t)|^2 - |\hat{h}(t|t-L)|^2)^2\}\\
&= \mathrm{E}\{|h(t)|^4 - 2|h(t)|^2|\hat{h}(t|t-L)|^2 + |\hat{h}(t|t-L)|^4\}\\
&= \mathrm{E}\{|h(t)|^4\} - 2\boldsymbol{\theta}\mathrm{E}\{|h(t)|^2\boldsymbol{\varphi}(t-L)\boldsymbol{\varphi}^H(t-L)\}\boldsymbol{\theta} + \mathrm{E}\{|\boldsymbol{\theta}^H\boldsymbol{\varphi}(t-L)\boldsymbol{\varphi}^H(t-L)\boldsymbol{\theta}|^2\}.
\end{aligned} \tag{7.6}$$

In Appendix 7.B it is shown that this MSE can be expressed in terms of covariances, as

$$\mathrm{E}\{|\varepsilon_{p_b}(t)|^2\} = 2\sigma_h^4 - 2\boldsymbol{\theta}^H\mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H\boldsymbol{\theta} - 2\sigma_h^2\boldsymbol{\theta}^H\mathbf{R}_\varphi\boldsymbol{\theta} + 2|\boldsymbol{\theta}^H\mathbf{R}_\varphi\boldsymbol{\theta}|^2. \tag{7.7}$$

This is the criterion that should be minimized to obtain the coefficient vector $\boldsymbol{\theta}$ that provides the best power predictor of the type (7.2).

Suboptimal MSE for the biased predictor

If the prediction coefficients that by (6.12) are optimal for the complex prediction, $\boldsymbol{\theta}_c = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$, are used and the absolute square of the complex prediction (7.2) is used as the power prediction, then the MSE, according to (7.7), will be

$$\begin{aligned}
\mathrm{E}\{|\varepsilon_{p_b}(t)|^2\} &= 2\sigma_h^4 - 2|\mathbf{r}_{h\varphi}^H\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}|^2 - 2\sigma_h^2\mathbf{r}_{h\varphi}^H\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi} + 2|\mathbf{r}_{h\varphi}^H\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}|^2\\
&= 2\sigma_h^2(\sigma_h^2 - \mathbf{r}_{h\varphi}^H\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}) \qquad (7.8)\\
&= 2\sigma_h^2\sigma_{\varepsilon_c}^2. \qquad (7.9)
\end{aligned}$$

The MSE in the power prediction will thus be proportional to the variance of the complex prediction error.

7.2.2 Unbiased quadratic power prediction

To avoid the bias introduced when using (7.2) we propose an unbiased power predictor which utilizes quadratic functions of the regressor variables

$$\boxed{\hat{p}(t+L|t) = \boldsymbol{\theta}^H \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\boldsymbol{\theta} + \sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}.} \tag{7.10}$$

This corresponds to the absolute square of the complex prediction as in (7.2), but with compensation for the bias,

$$\hat{p}(t+L|t) = |\hat{h}(t+L|t)|^2 + \sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta} = |\hat{h}(t+L|t)|^2 + \sigma_h^2 - \sigma_{\hat{h}}^2, \tag{7.11}$$

where we have used (6.17). To get an unbiased power estimate, the variance of the complex prediction is subtracted and the variance of the true signal is added instead. The prediction error is then given as

$$\varepsilon_p(t) = p(t) - \hat{p}(t|t-L), \tag{7.12}$$

and it will have zero mean.

In order to obtain the variance of the power prediction error, we use the relationship between autocorrelation and variance

$$\begin{aligned} \mathrm{E}\{|v - m_v|^2\} &= \mathrm{E}\{|v|^2\} - \mathrm{E}\{v^*\}m_v - \mathrm{E}\{v\}m_v^* + |m_v|^2 \\ &= \mathrm{E}\{|v|^2\} - |m_v|^2 \end{aligned} \tag{7.13}$$

where v is a random variable and $\mathrm{E}\{v\} = m_v$ is its mean. As the predictor in (7.10) is the same as in (7.2) but with the bias removed, the variance for the power prediction error, when using the unbiased power predictor, is given by the difference between (7.7) and the square of the bias (7.4),

$$\begin{aligned} &\sigma_{\varepsilon_p}^2 = \mathrm{E}\{|\varepsilon_p(t)|^2\} \\ &= 2\sigma_h^4 - 2\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H \boldsymbol{\theta} - 2\sigma_h^2 \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta} + 2|\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}|^2 - |\sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}|^2 \\ &= \sigma_h^4 - 2\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H \boldsymbol{\theta} + |\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}|^2. \end{aligned} \tag{7.14}$$

As this predictor has no bias, the variance and the MSE coincide. We summarize the discussion above in the following result.

Theorem 7.1 *Let a predictor of a power signal $p(t) = |h(t)|^2$, with $h(t) \in$ $\mathbf{CN}(0, \sigma_h^2)$, have the structure*

$$\hat{p}(t+L|t) = \boldsymbol{\theta}^H \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\boldsymbol{\theta} + \sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}$$

where $\boldsymbol{\varphi}(t)$ *is a column vector containing regressors with covariance matrix* $\mathbf{R}_\varphi$*. A coefficient vector* $\boldsymbol{\theta}$ *that minimizes the corresponding MSE*

$$\sigma_{\varepsilon_p}^2 = \sigma_h^2 - 2\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H \boldsymbol{\theta} + |\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}|^2$$

is then given by

$$\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$$

with $\mathbf{r}_{h\varphi} = \mathrm{E}\{h(t)\boldsymbol{\varphi}(t-L)\}$.

Proof 7.1 *To find the* $\boldsymbol{\theta}$ *that minimizes* $\sigma_{\varepsilon_p}^2$*, we take the partial derivative of the MSE with respect to* $\boldsymbol{\theta}$ *and set it equal to zero*

$$\frac{\partial \sigma_{\varepsilon_p}^2}{\partial \boldsymbol{\theta}} = -2\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H + 2\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}\boldsymbol{\theta}^H \mathbf{R}_\varphi = 0$$

which results in the equation

$$\boldsymbol{\theta}^H (\mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H - \mathbf{R}_\varphi \boldsymbol{\theta}\boldsymbol{\theta}^H \mathbf{R}_\varphi) = 0$$

Here $\boldsymbol{\theta} = 0$ *is the trivial solution. The other solutions are given by*

$$\boldsymbol{\theta}\boldsymbol{\theta}^H = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1}.$$

One obvious solution is $\boldsymbol{\theta} = \boldsymbol{\theta}_p$*, with*

$$\boldsymbol{\theta}_p = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}. \tag{7.15}$$

■

Note that $\boldsymbol{\theta}_p$ by (7.15) is identical to the coefficient vector $\boldsymbol{\theta}_c$ (6.12) that minimizes the MSE (6.8) when predicting the complex signal $h(t)$. The solution (7.15) is however not unique. Actually all $\boldsymbol{\theta}_p$ that satisfy $\boldsymbol{\theta}_p = \mathrm{e}^{i\vartheta}\boldsymbol{\theta}_c$, with $\vartheta \in [0\ 2\pi[$, are solutions. This is because the phase information becomes irrelevant when the absolute square is taken.

The MSE-optimal quadratic unbiased predictor

Using the optimal coefficients for a complex Gaussian signal, the power predictor in (7.10) can by (6.12) and (6.13) be written as

$$\begin{aligned} \hat{p}(t+L|t) &= |\hat{h}(t+L|t)|^2 + \sigma_h^2 - \mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi} && (7.16) \\ &= |\hat{h}(t+L|t)|^2 + \sigma_{\varepsilon_c}^2. && (7.17) \end{aligned}$$

The optimal unbiased quadratic power predictor (7.10), is thus the sum of the absolute square of the optimal linear prediction of the complex signal $h(t)$ and the variance of the complex prediction error. Its mean value is readily obtained from (6.17) and (6.18) as being equal to the true average power,

$$\mathrm{E}\{\hat{p}(t+L|t)\} = \sigma_{\hat{h}}^2 + \sigma_{\varepsilon_c}^2 = \sigma_h^2. \tag{7.18}$$

Note that when using this predictor, the predicted power will never be less than $\sigma_{\varepsilon_c}^2$, that is the variance of the complex prediction error. This means that it will not predict zero power even when the complex prediction indicates a deep fading dip ($\hat{h}(t+L|t) = 0$).

The minimum MSE and correlation for the unbiased predictor

The optimal unbiased quadratic power predictor results in a minimum power MSE, obtained by inserting $\boldsymbol{\theta}_p$, from (7.15), into (7.14),

$$\begin{aligned}\min_{\boldsymbol{\theta}} \sigma_{\varepsilon_p}^2 &= \sigma_h^4 - 2\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi} \mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi} + |\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{R}_\varphi \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}|^2 \\ &= \sigma_h^4 - |\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}|^2. \end{aligned} \tag{7.19}$$

This can also be expressed in terms of the variance (6.13) of the complex prediction error using the optimal predictor,

$$\min_{\boldsymbol{\theta}} \sigma_{\varepsilon_p}^2 = \sigma_h^4 - |\sigma_h^2 - \sigma_{\varepsilon_c}^2|^2 = \sigma_{\varepsilon_c}^2 (2\sigma_h^2 - \sigma_{\varepsilon_c}^2). \tag{7.20}$$

The variance of the power prediction error is thus bounded as $0 \leq \sigma_{\varepsilon_p}^2 \leq \sigma_h^4$ and it increases monotonicly with $\sigma_{\varepsilon_c}^2$, when the optimal coefficients are used.

In Appendix 7.A the following expressions, useful for characterizing the properties of the power predictor (7.17), are derived. Using (6.18) and (7.18) the variance for the unbiased power prediction is

$$\sigma_{\hat{p}}^2 = \mathrm{E}\{(\hat{p}(t+L|t) - \sigma_h^2)^2\} = \sigma_{\hat{h}}^4 = |\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}|^2 = |\sigma_h^2 - \sigma_{\varepsilon_c}^2|^2, \tag{7.21}$$

see (7.106) and (7.108). The variance for the predicted power is thus lower than for the true power. The cross-correlation between the predicted power and the power prediction error is

$$\rho_{\hat{p}\varepsilon_p} = \mathrm{E}\{\hat{p}(t|t-L)\varepsilon_p(t)\} = 0. \tag{7.22}$$

Hence, using the optimal coefficients in the prediction, there is no correlation between the predicted power and the prediction error that could be further utilized by adding a linear function of $\varepsilon_p(t)$ to the predictor.

7.2.3 Comparison of performance

The results above can be used to compare the performance of the unbiased power predictor and the biased version. In both cases the coefficients that are optimal for the complex prediction is used. As the NMSE for the complex prediction error is bounded by $0 \leq \sigma_{\varepsilon_c}^2/\sigma_h^2 \leq 1$, when optimal coefficients are used, we can evaluate the NMSE for the power prediction for all possible NMSE's for the linear complex prediction error using (7.9) and (7.20). The power prediction NMSE for the biased predictor is thus

$$V_{biased} = \frac{2\sigma_h^2\sigma_{\varepsilon_c}^2}{2\sigma_h^4} = \frac{\sigma_{\varepsilon_c}^2}{\sigma_h^2}, \tag{7.23}$$

which happens to equal the NMSE for the linear complex prediction.

The NMSE for the unbiased quadratic predictor can also be expressed in terms of the NMSE for the linear complex prediction,

$$V_{unbiased} = \frac{\sigma_{\varepsilon_c}^2(2\sigma_h^2 - \sigma_{\varepsilon_c}^2)}{2\sigma_h^4} = \frac{\sigma_{\varepsilon_c}^2}{\sigma_h^2}\left(1 - \frac{\sigma_{\varepsilon_c}^2}{2\sigma_h^2}\right). \tag{7.24}$$

In Figure 7.1, the performance at different complex NMSE are shown. We see that it is mainly when the complex time signal is hard to predict that the benefit of using the unbiased predictor is fully exploited. The very simplest power predictor, just using the average power $\bar{p} = \sigma_h^2$, has an MSE that is equal to half the variance of the power (which is $\sigma_p^2 = 2\sigma_h^4$),

$$\mathrm{E}\{(p(t) - \bar{p})^2\} = \mathrm{E}\{p^2(t)\} - \bar{p}^2 = 2\sigma_h^4 - \sigma_h^4 = \sigma_h^4 = \frac{\sigma_p^2}{2}. \tag{7.25}$$

The biased predictor deploying $\boldsymbol{\theta}_c$ will actually be worse than prediction using the average, for cases where $\sigma_{\varepsilon_c}^2 > 0.5$.

The difference in the performance can be understood like this: Even when the complex prediction totally fails and $\boldsymbol{\theta} = 0$ are the best coefficients for the prediction of the complex signal, the optimal unbiased quadratic power predictor sets the prediction to the average power, whereas the biased predictor predicts the power to be zero. On the other hand, when the complex prediction improves, the bias will decrease and the gain obtained by using the unbiased predictor will be reduced.

7.2.4 Prediction based on observed time-series

In this subsection we briefly discuss the case where the statistics of the signal $h(t)$ is unknown and that only the observations $y(t)$ are available.

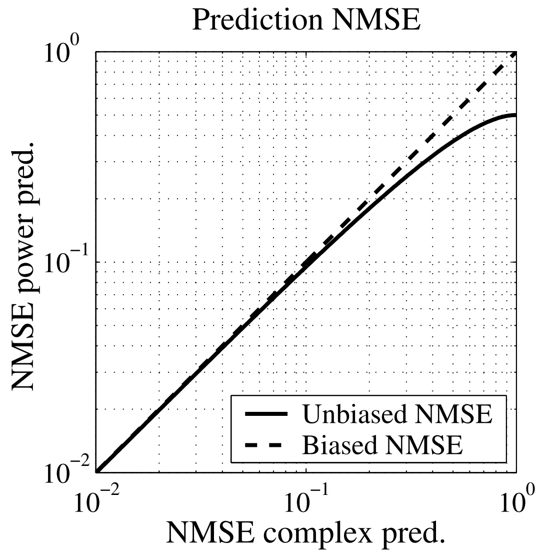

Figure 7.1: The NMSE $\sigma^2_{\varepsilon_p}/2\sigma^4_h$ for the power prediction as a function of the NMSE for the complex prediction, $\sigma^2_{\varepsilon_c}/\sigma^2_h$. Optimal linear prediction of the complex time series $h(t)$ is assumed. Equation (7.23) and (7.24) are used to evaluate the NMSE for the biased and unbiased power predictor respectively.

The statistics of $y(t)$ can be estimated from the time-series and is assumed known. The complex predictor can then be designed to minimize the MSE for the predicted complex observation

$$\mathrm{E}\{|\varepsilon_y(t)|^2\} = \mathrm{E}\{|y(t) - \boldsymbol{\varphi}^H(t-L)\boldsymbol{\theta}|^2\}, \tag{7.26}$$

where the regressor vector $\boldsymbol{\varphi}(t-L)$ is given by (6.5). Following the results in Section 6.2, Wiener optimization of this predictor results in the optimal coefficients

$$\boldsymbol{\theta}_y = \mathbf{R}_\varphi^{-1}\mathbf{r}_{y\varphi} = \mathbf{R}_\varphi^{-1}(\mathbf{r}_{h\varphi} + \mathbf{r}_{e\varphi}) \tag{7.27}$$

where $\mathbf{r}_{y\varphi}$ is the cross-covariance between the observed complex time signal and the regressor and $\mathbf{r}_{e\varphi}$ is the cross-covariance between noise and regressor.

As in Section 7.2.2, for the true power of $h(t+L)$, we obtain the optimal quadratic power predictor for the observed power of $y(t+L)$ as

$$\hat{p}_y(t+L|t) = |\boldsymbol{\varphi}^H(t)\mathbf{R}_\varphi^{-1}\mathbf{r}_{y\varphi}|^2 + \sigma^2_{\varepsilon_y}. \tag{7.28}$$

This is *not* the optimal quadratic prediction of the true power $p(t)$ of $h(t)$, as the prediction coefficients $\boldsymbol{\theta}_y$ are used instead of the unknown coefficients

$\boldsymbol{\theta}_p$ that are obtained in (7.15). If the variance of the observation noise is available, at least an unbiased predictor of the power of $h(t+L)$ can be obtained as

$$\hat{p}(t+L|t) = \hat{p}_y(t+L|t) - r_e. \tag{7.29}$$

If the auto-correlation of the noise is zero for lags larger than L, then $\mathbf{r}_{h\varphi} = \mathbf{r}_{y\varphi}$, and consequently (7.29) will actually be the MSE-optimal quadratic predictor of the power of $h(t+L)$.

7.3 Unbiased Quadratic Power Prediction of a Frequency Selective Channel

A frequency selective channel can be modeled as a time varying FIR-filter $\mathbf{h}(t)$, where each tap $h_k(t)$, can be seen as a flat Rayleigh fading channel. The total power is thus the sum of the power contribution from each tap,

$$p(t) = \sum_{k=1}^{M} p_k(t) = \sum_{k=1}^{M} |h_k(t)|^2. \tag{7.30}$$

The linear regression

$$\hat{h}_k(t+L|t) = \boldsymbol{\varphi}_k^H(t)\boldsymbol{\theta}_k \ , \tag{7.31}$$

represents an L step ahead FIR-prediction of the channel tap $h_k(t)$. Here, $\boldsymbol{\theta}_k$ is the vector of predictor coefficients for tap k. The regression vector $\boldsymbol{\varphi}_k(t)$ consists of possibly smoothed observations of the tap $h_k(t)$, up to time t, as in (6.31).

The unbiased power predictor (7.10) for tap k is then obtained as

$$\hat{p}_k(t+L|t) = |\hat{h}_k(t+L|t)|^2 + \sigma_{h_k}^2 - \boldsymbol{\theta}_k^H \mathbf{R}_{\varphi_k} \boldsymbol{\theta}_k \ , \tag{7.32}$$

where $\sigma_{h_k}^2 = \mathrm{E}\{|h_k(t)|^2\}$ and $\hat{h}_k(t+L|t)$ is given by (7.31).

As each tap of a frequency selective channel can be regarded as a flat Rayleigh fading channel, that is $h_k(t) \sim \mathbf{CN}(0, \sigma_{h_k}^2)$, the optimal choice of $\boldsymbol{\theta}_k$ is then obtained as in (7.15),

$$\boldsymbol{\theta}_k = \mathbf{R}_{\varphi_k}^{-1} \mathbf{r}_{h_k\varphi_k} \ , \tag{7.33}$$

where $\mathbf{R}_{\varphi_k}$ is the covariance matrix for the regressor to tap k and $\mathbf{r}_{h_k\varphi_k} = \mathrm{E}\{h_k(t)\boldsymbol{\varphi}_k(t)\}$.

An unbiased predictor for the total power is finally obtained by summing the contributions from each significant tap

$$\hat{p}(t+L|t) = \sum_{k=1}^{n} \hat{p}_k(t+L|t) = \sum_{k=1}^{n} \left(|\hat{h}_k(t+L|t)|^2 + \sigma_{h_k}^2 - \boldsymbol{\theta}_k^H \mathbf{R}_{\varphi_k} \boldsymbol{\theta}_k \right). \tag{7.34}$$

If all channel taps are unpredictable, then the channel power will be predicted by its average $\sum_k \sigma_{h_k}^2$. In receivers that use only a subset of the taps, the available power is the sum over this subset. The predicted available power is then the sum over the predictions for this subset, whereas the other taps act as interference.

In Section 7.7 this method will be used for power prediction on measured frequency selective channels. Only the taps that significantly contribute to the total power has to be predicted. The power contribution from the other taps can be seen as a minor disturbance.

7.4 Distributions for Power, Prediction and Error

To be able to analyze the performance of the predictor it is not sufficient to know the mean and variance of the prediction error. To fully describe the properties of the predictor we need the probability density functions for the true and the predicted power. It is especially relevant to study the joint probability density function, which can be used to analyze the performance of a system using the predictor. In Chapter 8 the derived pdf's are used to to obtain the optimal boundaries for changing the modulation in an adaptive modulation system.

7.4.1 Distributions of the power and the predicted power

Consider the channel model (6.1). The signal $x(t)$, the noise $e(t)$ and the observation $y(t)$, are assumed to be complex circular Gaussian random variables, thus **CN** distributed, with zero mean. The same holds for the prediction, $\hat{h}(t+L|t)$, as it is composed by linear filtering of the Gaussian sequence $y(t)$. (In general, linear filtering will make the distribution of the filter output, approximate a Gaussian distribution better than the distribution of the input does.) The complex prediction error, $\varepsilon_c(t)$, cf (6.7) will thus also be **CN** distributed, as it only consist of sums of other **CN** distributed variables.

The pdf for power

Let z be a complex circular Gaussian random variable with zero mean and variance σ_z^2, that is $z \in \mathbf{CN}(0, \sigma_z^2)$. Then, $p_z = |z|^2$ will have a $\chi^2(2)$ distribution [62]. The pdf for a $\chi^2(2)$ distribution is exponential in p_z,

$$f_{p_z}(p_z) = \frac{1}{\sigma_z^2} \mathrm{e}^{-p_z/\sigma_z^2} \mathrm{U}(p_z), \tag{7.35}$$

where $\mathrm{U}(\cdot)$ is Heaviside's step function. The mean of p_z is the variance of the complex variable, $m_{p_z} = \sigma_z^2$, and the variance of p_z is $\sigma_{p_z}^2 = \sigma_z^4$.

As the absolute square of a $\mathbf{CN}$ distributed random variable is $\chi^2(2)$ distributed, the true power $p(t)$ and the observed power $p_y(t)$, will both be $\chi^2(2)$ distributed.

The pdf for the predicted power

The biased quadratic predictor (7.2) is $\chi^2(2)$ distributed as well, since $\hat{h}(t+L|t)$ will be $\mathbf{CN}$ distributed. The distribution of the unbiased quadratic predictor (7.11)

$$\hat{p}(t+L|t) = |\hat{h}(t+L|t)|^2 + \sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}, \tag{7.36}$$

will therefore be a shifted $\chi^2(2)$ distribution. Using that the variance $\sigma_{\hat{h}}^2$ of the complex prediction is $\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}$, according to (6.17), this distribution can be expressed in terms of the distribution of

$$|\hat{h}(t+L|t)|^2 = \hat{p}(t+L|t) - \sigma_h^2 + \sigma_{\hat{h}}^2 \tag{7.37}$$

as

$$f_{\hat{p}}(\hat{p}(t+L|t)) = \frac{1}{\sigma_{\hat{h}}^2} \mathrm{e}^{-(\hat{p}(t+L|t)-\sigma_h^2+\sigma_{\hat{h}}^2)/\sigma_{\hat{h}}^2} \mathrm{U}\Big(\hat{p}(t+L|t)-\sigma_h^2+\sigma_{\hat{h}}^2\Big). \tag{7.38}$$

More conveniently, the pdf can also be expressed using the variances of the signal and the complex prediction error, using $\sigma_{\hat{h}}^2 = \sigma_h^2 - \sigma_{\varepsilon_c}^2$ from (6.18)

$$f_{\hat{p}}(\hat{p}(t+L|t)) = \frac{\mathrm{U}\big(\hat{p}(t+L|t)-\sigma_{\varepsilon_c}^2\big)}{\sigma_h^2 - \sigma_{\varepsilon_c}^2} \exp\left(-\frac{\hat{p}(t+L|t) - \sigma_{\varepsilon_c}^2}{\sigma_h^2 - \sigma_{\varepsilon_c}^2}\right). \tag{7.39}$$

7.4.2 The joint probability density for the true and the predicted power

The joint probability density function (jpdf) for the true and predicted power can be used to obtain results for bit error rates for certain communication systems that use linear smoothing or prediction of the channel [7]. To obtain the jpdf, we start with two correlated **CN** distributed stochastic variables h and $\hat{h}$ (in the following the time indices are dropped for notational convenience). As both $|h|^2$ and $|\hat{h}|^2$ are $\chi^2(2)$ distributed, their joint probability density is a bivariate χ^2 distribution (also called bivariate Γ distribution). From Appendix 7.C we have that the jpdf between two $\chi^2(2)$ distributed stochastic variables, formed by the absolute squares of two zero mean **CN** distributed variables h and $\hat{h}$, is

$$f_{|h|^2|\hat{h}|^2}(|h|^2, |\hat{h}|^2) = \frac{1}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2} \exp\left(\frac{\sigma_h^2\sigma_{\hat{h}}^2}{|r_{h\hat{h}}|^2 - \sigma_h^2\sigma_{\hat{h}}^2} \left(\frac{|h|^2}{\sigma_h^2} + \frac{|\hat{h}|^2}{\sigma_{\hat{h}}^2} \right) \right) \mathrm{I}_0\left(\frac{2|r_{h\hat{h}}||h||\hat{h}|}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2} \right), \quad (7.40)$$

where σ_h^2 and $\sigma_{\hat{h}}^2$ are the variances for h and $\hat{h}$ respectively, and where $r_{h\hat{h}}$ is the cross-covariance between h and $\hat{h}$. Here $\mathrm{I}_0(\cdot)$ is the modified Bessel function of the first kind of order zero (it is obtained by inserting an imaginary argument into the ordinary Bessel function of the first kind of order zero, and it does not decay with increasing argument).

Assume now that h and $\hat{h}$ in (7.40) relate to the complex signal in (6.1) and its prediction by (6.4). Using linear prediction, we know from (6.14) that $r_{h\hat{h}} = \boldsymbol{\theta}^H \mathbf{r}_{h\varphi}$ and from (6.17) that $\sigma_{\hat{h}}^2 = \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}$. The jpdf for the true power $p = |h|^2$ and the predicted power using the biased quadratic predictor $\hat{p}_b = |\hat{h}|^2$, is then given by (7.40). As in (7.36), the biased power prediction can be expressed in terms of the unbiased power predictor as

$$\hat{p}_b = |\hat{h}|^2 = \hat{p} - \sigma_h^2 + \sigma_{\hat{h}}^2. \quad (7.41)$$

For the unbiased power predictor $\hat{p}$, using (6.14), (6.17) and performing the variable substitution (7.41) above in (7.40) the jpdf thus is

$$f_{p\hat{p}}(p,\hat{p}) = \frac{\mathrm{U}(p)\mathrm{U}(\hat{p}-\sigma_h^2+\sigma_{\hat{h}}^2)}{\sigma_h^2\sigma_{\hat{h}}^2-|r_{h\hat{h}}|^2}\times$$
$$\exp\!\left(\frac{\sigma_h^2\sigma_{\hat{h}}^2}{|r_{h\hat{h}}|^2-\sigma_h^2\sigma_{\hat{h}}^2}\left(\frac{p}{\sigma_h^2}+\frac{\hat{p}-\sigma_h^2+\sigma_{\hat{h}}^2}{\sigma_{\hat{h}}^2}\right)\right)\mathrm{I}_0\!\left(\frac{2|r_{h\hat{h}}|\sqrt{p(\hat{p}-\sigma_h^2+\sigma_{\hat{h}}^2)}}{\sigma_h^2\sigma_{\hat{h}}^2-|r_{h\hat{h}}|^2}\right), \tag{7.42}$$

with $r_{h\hat{h}} = \boldsymbol{\theta}^H\mathbf{r}_{h\varphi}$ and $\sigma_{\hat{h}}^2 = \boldsymbol{\theta}^H\mathbf{R}_\varphi\boldsymbol{\theta}$. Note that since $|h|^2$ and $|\hat{h}|^2$ in (7.40) are always positive, the step-functions are included in (7.42) to ensure that the arguments remain positive.

If the predictor coefficients $\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$, that minimizes the MSE for the power (7.15), are used in the unbiased quadratic power predictor, we can perform the substitution

$$\sigma_h^2 - \sigma_{\varepsilon_c}^2 = \sigma_{\hat{h}}^2 = r_{h\hat{h}} \tag{7.43}$$

from (6.18) and (6.16). Then, the jpdf for p and $\hat{p}$ in (7.42) can be expressed as[1]

$$f_{p\hat{p}}(p,\hat{p}) = \frac{\mathrm{U}(p)\mathrm{U}(\hat{p}-\sigma_{\varepsilon_c}^2)}{\sigma_{\varepsilon_c}^2(\sigma_h^2-\sigma_{\varepsilon_c}^2)}\exp\!\left(-\frac{p}{\sigma_{\varepsilon_c}^2}-\frac{(\hat{p}-\sigma_{\varepsilon_c}^2)\sigma_h^2}{(\sigma_h^2-\sigma_{\varepsilon_c}^2)\sigma_{\varepsilon_c}^2}\right)\mathrm{I}_0\!\left(\frac{2}{\sigma_{\varepsilon_c}^2}\sqrt{p(\hat{p}-\sigma_{\varepsilon_c}^2)}\right). \tag{7.44}$$

The jpdf for the true and the predicted power, using the optimal unbiased quadratic predictor, can thus be expressed in terms of the variance σ_h^2 of the complex time series $h(t)$ and the variance $\sigma_{\varepsilon_c}^2$ of the complex prediction error $\varepsilon_c(t) = h(t) - \hat{h}(t|t-L)$.

For prediction of a the power of a tap in a mobile radio channel it is convenient to express the jpdf in terms of the prediction error variance, or of the received SNR in a communication system. Statistics expressed in this way will be directly applicable in the evaluation of communication systems that utilize power prediction. This is exemplified in the study of optimized link adaption in Chapter 8.

[1] The corresponding jpdf for the true power and the biased quadratic power prediction is obtained by the variable substitution $\hat{p} = \hat{p}_b + \sigma_{\varepsilon_c}^2$.

7.4.3 Conditional probability density functions

Conditional probability density function for the true power

To be able to evaluate the performance of a system utilizing predicted powers we need to know the statistical behavior of the true power given the prediction. The conditional probability density function (cpdf) for the power is thus of great interest. It can be obtained as

$$f_{p\hat{p}}(p|\hat{p}) = \frac{f_{p\hat{p}}(p,\hat{p})}{f_{\hat{p}}(\hat{p})}, \tag{7.45}$$

where the division only is performed in the region where $f_{\hat{p}}(\hat{p}) \neq 0$. The cpdf for p and $\hat{p}$ can then be obtained by inserting (7.39) and (7.44) in (7.45), yielding

$$f_{p\hat{p}}(p|\hat{p}) = \frac{\mathrm{U}(p)\mathrm{U}(\hat{p}-\sigma^2_{\varepsilon_c})}{\sigma^2_{\varepsilon_c}} \exp\left(-\frac{p+\hat{p}-\sigma^2_{\varepsilon_c}}{\sigma^2_{\varepsilon_c}}\right) \mathrm{I}_0\left(\frac{2}{\sigma^2_{\varepsilon_c}}\sqrt{p(\hat{p}-\sigma^2_{\varepsilon_c})}\right). \tag{7.46}$$

In Chapter 8 this expression is used in the design and performance evaluation of a system with link adaption.

The conditional prediction error density

The conditional pdf for the power prediction error ε_p, using the unbiased power predictor (7.10) with the optimal parameters (7.15) can be obtained from (7.46), using the transformation $p = \hat{p} + \varepsilon_p$,

$$f_{\varepsilon_p}(\varepsilon_p|\hat{p}) =$$
$$\frac{\mathrm{U}(\hat{p}+\varepsilon_p)\mathrm{U}(\hat{p}-\sigma^2_{\varepsilon_c})}{\sigma^2_{\varepsilon_c}} \exp\left(-\frac{\varepsilon_p+2\hat{p}-\sigma^2_{\varepsilon_c}}{\sigma^2_{\varepsilon_c}}\right) \mathrm{I}_0\left(\frac{2}{\sigma^2_{\varepsilon_c}}\sqrt{(\hat{p}-\sigma^2_{\varepsilon_c})(\hat{p}+\varepsilon_p)}\right). \tag{7.47}$$

The predictor error has zero mean for all $\hat{p}$, as shown in (7.136) Appendix 7.C.2, but the variance of the prediction error $\sigma^2_{\varepsilon_p}$, obtained as in (7.144) Appendix 7.C.2, depends on the prediction

$$\sigma^2_{\varepsilon_p}(\hat{p}) = \sigma^2_{\varepsilon_c}(2\hat{p} - \sigma^2_{\varepsilon_c}). \tag{7.48}$$

The relative standard deviation $\sigma_{\varepsilon_p}(\hat{p})/\hat{p}$ is seen in Figure 7.2. The relative uncertainty is thus largest for the smallest predicted powers.

The predicted power is in the range $\sigma^2_{\varepsilon_c} \leq \hat{p} < \infty$ and the prediction error is in the range $-\hat{p} \leq \varepsilon_p < \infty$. For $\hat{p} = \sigma^2_{\varepsilon_c}$ (obtained when $\hat{h}(t+L|t) = 0$), the

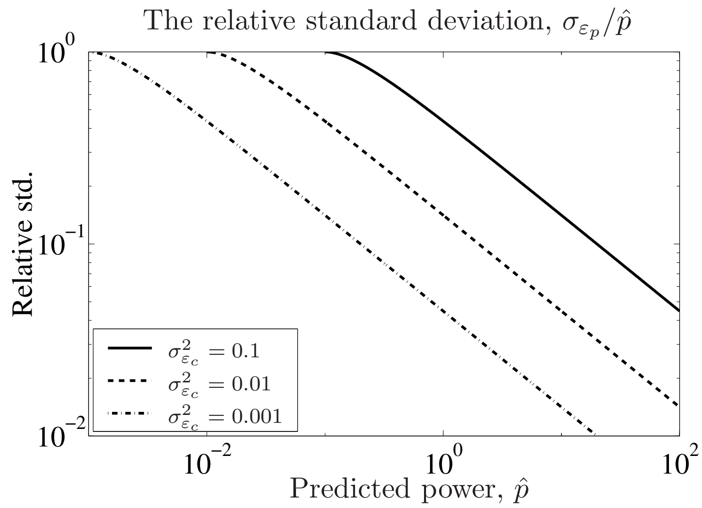

Figure 7.2: The relative standard deviation for the predicted power $\sigma_{\varepsilon_p}/\hat{p}$ as a function of $\hat{p}$ for different variances for the complex channel prediction error.

power prediction error has a $\chi^2(2)$ distribution shifted to the left. When the prediction $\hat{p}$ increases the distribution for the error becomes more Gaussian and its variance increases according to (7.48).

Example 7.1

Let the variance of the signal be $\sigma_h^2 = 1$, and let the NMSE and the variance of the complex prediction error be $\sigma^2_{\varepsilon_c} = 0.1$. Then the NMSE for the power prediction is 0.095 by (7.24). The pdf for the prediction error *conditioned* on the predicted power according to equation (7.47) is depicted in Figure 7.3. As seen in Figure 7.3, and calculated in Appendix 7.C.2, the power prediction error has a $\chi^2(2)$ like pdf for small power predictions $\hat{p}$, but as $\hat{p}$ increases, the pdf for the error becomes more Gaussian. Thus, when the the power is predicted to be high, the error is symmetric around the predicted value which is not the case for low predicted powers. The pdf is much wider for high predicted powers than for low. The variance for the power prediction error is thus seen to increase with the predicted power, which is in accordance with (7.48).

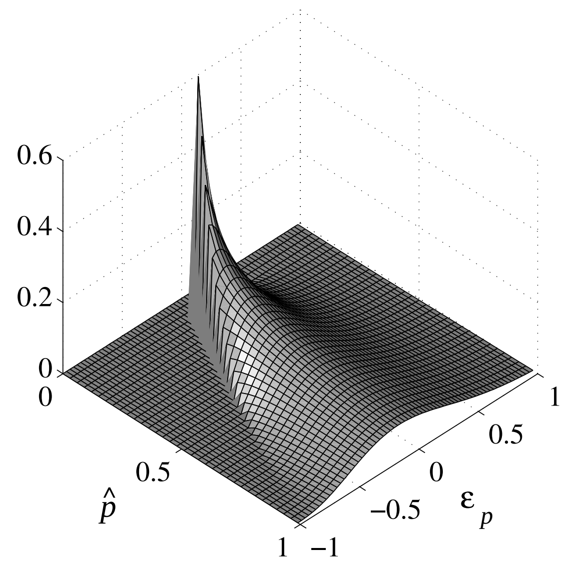

Figure 7.3: The conditional pdf (7.47) for the prediction error conditioned on the predicted power, with $\sigma_h^2 = 1$ and $\sigma_{\varepsilon_c}^2 = 0.1$.

7.4.4 The prediction error probability density

In the following, we consider the optimal unbiased quadratic power predictor (7.17) of Section 7.2.2. To derive the probability density function of the prediction error for the power, we express the error in terms of the complex linear prediction and its corresponding prediction error,

$$\begin{aligned}\varepsilon_p(t) &= p(t) - \hat{p}(t|t-L) = |h(t)|^2 - |\hat{h}(t|t-L)|^2 - \sigma_{\varepsilon_c}^2 \\ &= |\hat{h}(t|t-L) + \varepsilon_c(t)|^2 - |\hat{h}(t|t-L)|^2 - \sigma_{\varepsilon_c}^2 \\ &= |\varepsilon_c(t)|^2 + \hat{h}(t|t-L)\varepsilon_c^*(t) + \hat{h}^*(t|t-L)\varepsilon_c(t) - \sigma_{\varepsilon_c}^2 \quad (7.49)\end{aligned}$$

Using the optimal predictor coefficients $\boldsymbol{\theta}_c$, the prediction $\hat{h}(t|t-L)$ and the prediction error $\varepsilon_c(t)$ will, according to equation (7.22), be uncorrelated. This highly simplifies the derivation of the distributions.

The pdf for the error

The pdf for the power prediction error using optimal linear prediction, as derived in Appendix 7.C.3, can be expressed in terms of the variance of the complex prediction error and the variance of the complex time series as

$$f_{\varepsilon_p}(\varepsilon_p) = \begin{cases} \frac{1}{\rho_+ + \rho_-} \mathrm{e}^{-(\varepsilon_p + \sigma^2_{\varepsilon_c})/\rho_+} & \text{for } \varepsilon_p \geq -\sigma^2_{\varepsilon_c} \\ \frac{1}{\rho_+ + \rho_-} \mathrm{e}^{(\varepsilon_p + \sigma^2_{\varepsilon_c})/\rho_-} & \text{otherwise} \end{cases} \tag{7.50}$$

with

$$\rho_\pm = \frac{1}{2}\left(\sqrt{\sigma^2_{\varepsilon_c}(4\sigma^2_h - 3\sigma^2_{\varepsilon_c})} \pm \sigma^2_{\varepsilon_c}\right) = \frac{\sigma^2_h}{2}\left(\sqrt{\frac{\sigma^2_{\varepsilon_c}}{\sigma^2_h}\left(4 - 3\frac{\sigma^2_{\varepsilon_c}}{\sigma^2_h}\right)} \pm \frac{\sigma^2_{\varepsilon_c}}{\sigma^2_h}\right). \tag{7.51}$$

This is a skew Laplacian distribution which is heavy to the front. In the pdf the exponential fall-off is faster for errors that are more negative than minus the variance of the complex prediction error, than for errors that are larger. Due to the use of the unbiased power predictor $\mathrm{E}\{\varepsilon_p\} = 0$. The smaller the ratio between the variance of the complex prediction error and the variance of the channel is, the more symmetric is the distribution for the power prediction error.

Example 7.2

Let the variance of the signal be $\sigma^2_h = 1$, and consider the pdf for the variances 0.1 and 0.5 for the complex prediction error. The NMSE for the unbiased power prediction is then 0.095 and 0.375 respectively, by (7.24). Figure 7.2 depicts the pdf for the power prediction error according to equation (7.50) and (7.51). The power prediction error is the result of the subtraction between two stochastic variables with asymmetrical one sided distributions. The result will thus be a skew distribution. The distribution becomes more skew when the variance $\sigma^2_{\varepsilon_c}$, of the complex prediction error increases, as also seen directly in (7.51). When the variance for the complex prediction error $\sigma^2_{\varepsilon_c}$ approaches σ^2_h then $\rho_+ \to \sigma^2_h$ and $\rho_- \to 0$. The peak in the distribution, that is the most likely outcome, is not at zero but at $-\sigma^2_{\varepsilon_c}$. The most likely value for the power prediction error is thus $\varepsilon_p = -\sigma^2_{\varepsilon_c}$, but on average it will be zero. This is an effect of the bias compensation.

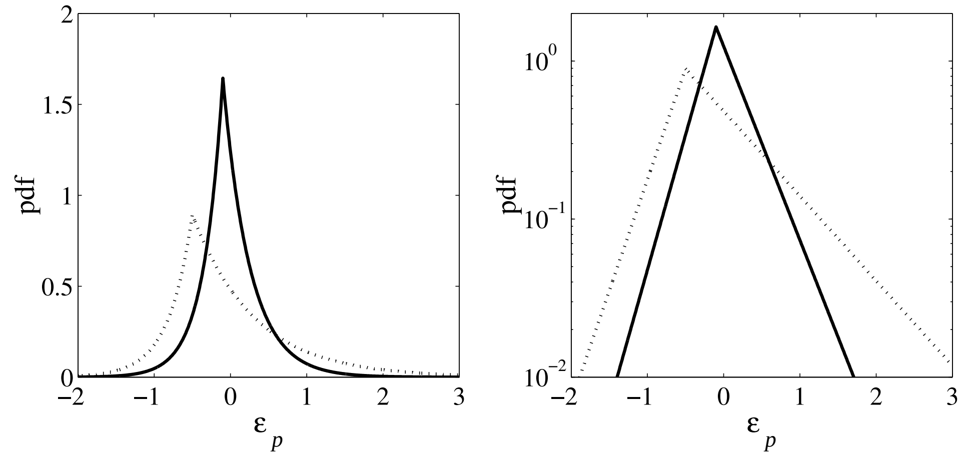

Figure 7.4: The pdf for the power prediction error when $\sigma^2_{\varepsilon_c} = 0.1$, solid line, and $\sigma^2_{\varepsilon_c} = 0.5$, dotted line. (Linear scale to the left and logarithmic scale to the right.)

7.5 Linear Power Prediction using Observed Power in the Regressor

A linear power predictor using observed power in the regressor can be derived in a similar manner as the linear predictor for the complex time series $h(t)$. In this section we will show that the performance of the this power predictor can be expressed using the same covariances as for the complex predictor.

7.5.1 Linear power prediction

For the signal $h(t)$ by (6.1), the true power is $p(t) = |h(t)|^2$, whereas the observed power is

$$p_y(t) = |y(t)|^2 = |h(t) + e(t)|^2 = |h(t)|^2 + h(t)e^*(t) + x^*(t)e(t) + |e(t)|^2 \ . \tag{7.52}$$

The noise on the power is thus both multiplicative and additive. Just as for the complex signal, a linear predictor for the power time series can be expressed as (cf. (7.104))

$$\hat{p}(t + L|t) = \boldsymbol{\phi}^T(t)\boldsymbol{\vartheta} + \bar{p} \ , \tag{7.53}$$

where $\boldsymbol{\vartheta}$ is the vector of predictor coefficients and $\bar{p}$ is the average power of $p(t)$. It is beneficial for the linear predictor to have a regressor with zero mean and then to add the mean afterwards, as in (7.53). The regressor thus consists of the power observations with their average subtracted,

$$\boldsymbol{\phi}(t) = [p_y(t) - \bar{p}_y \;\; p_y(t-\Delta t) - \bar{p}_y \ldots p_y(t-(M-1)\Delta t) - \bar{p}_y]^T \; . \tag{7.54}$$

The average power of $h(t)$ is

$$\bar{p} = \mathrm{E}\{p(t)\} = \mathrm{E}\{|h(t)|^2\} = \sigma_h^2 \; , \tag{7.55}$$

which is the variance of the complex time series $h(t)$. For the observations $y(t)$, the average power is

$$\begin{aligned} \bar{p}_y &= \mathrm{E}\{p_y(t)\} = \mathrm{E}\{|h(t)+e(t)|^2\} \\ &= \mathrm{E}\{|h(t)|^2\} + \mathrm{E}\{h(t)e^*(t)\} + \mathrm{E}\{x^*(t)e(t)\} + \mathrm{E}\{|e(t)|^2\} \\ &= \sigma_h^2 + \sigma_e^2 \; , \end{aligned} \tag{7.56}$$

which is the sum of the variances for signal and noise respectively.

The prediction error resulting from the use of the predictor (7.53) is

$$\epsilon_p(t) = p(t) - \hat{p}(t|t-L) = p(t) - \bar{p} - \boldsymbol{\phi}^T(t-L)\boldsymbol{\vartheta} \; . \tag{7.57}$$

The first two terms on the last line $p(t) - \bar{p} = |h(t)|^2 - \sigma_h^2$ represent the time series for the power with the mean subtracted.

The MSE for the prediction is

$$\begin{aligned} \mathrm{E}|\epsilon_p(t)|^2 &= \mathrm{E}\{|p(t) - \bar{p} - \boldsymbol{\phi}^T(t-L)\boldsymbol{\vartheta}|^2\} \\ &= \mathrm{E}\{|p(t) - \sigma_h^2|^2\} - \mathrm{E}\{\boldsymbol{\vartheta}^T\boldsymbol{\phi}(t-L)(p(t)-\sigma_h^2)\} \\ &\quad - \mathrm{E}\{(p(t)-\sigma_h^2)\boldsymbol{\phi}^T(t-L)\boldsymbol{\vartheta}\} + \mathrm{E}\{\boldsymbol{\vartheta}^T\boldsymbol{\phi}(t-L)\boldsymbol{\phi}^T(t-L)\boldsymbol{\vartheta}\} \\ &= \sigma_h^4 - \boldsymbol{\vartheta}^T\mathbf{r}_{p\phi} - \mathbf{r}_{p\phi}^T\boldsymbol{\vartheta} + \boldsymbol{\vartheta}^T\mathbf{R}_\phi\boldsymbol{\vartheta} \; , \end{aligned} \tag{7.58}$$

where $\mathbf{r}_{p\phi}$ is the cross-covariance between $p(t)$ and the regressor $\boldsymbol{\phi}(t-L)$ and where $\mathbf{R}_\phi$ is the covariance matrix of the regressor. In (7.58), we exploited that, since $h(t) \in \mathbf{CN}(0, \sigma_h^2)$ and is circular, the variance of the power can be expressed as

$$\begin{aligned} \sigma_p^2 &= \mathrm{E}\{|p(t)-\bar{p}|^2\} = \mathrm{E}\{|h(t)|^2 - \sigma_h^2|^2\} \\ &= \mathrm{E}\{|h(t)|^4\} - 2\mathrm{E}\{|h(t)|^2\}\sigma_h^2 + \sigma_h^4 \\ &= 2\sigma_h^4 - 2\sigma_h^4 + \sigma_h^4 = \sigma_h^4 \; . \end{aligned} \tag{7.59}$$

According to the Wiener-Hopf equations, the coefficients of a linear predictor (7.53) that minimize the MSE in (7.58) are given by

$$\boldsymbol{\vartheta} = \mathbf{R}_\phi^{-1}\mathbf{r}_{p\phi} \; . \tag{7.60}$$

In Appendix 7.A.2, equation (7.116), it is shown that the cross-covariance $\mathbf{r}_{p\phi}$ is

$$\begin{aligned}\mathbf{r}_{p\phi} &= \mathrm{E}\{(p(t)-\bar{p})\boldsymbol{\phi}(t-L)\} \\ &= \left[\ |r_h(L)|^2 \ \ |r_h(L-\Delta t)|^2 \dots |r_h(L-(M-1)\Delta t)|^2\ \right]^T \\ &= \mathbf{r}_{h\varphi} \odot \mathbf{r}^*_{h\varphi}\ , \end{aligned} \tag{7.61}$$

where $\odot$ denotes the Hadamard product (element wise multiplication). The cross-covariance between the power and the regressor of observed powers can thus be expressed in terms of the corresponding cross-covariance between the complex time series $h(t)$ and the regressor vector $\boldsymbol{\varphi}(t-L)$.

The elements of the covariance matrix $\mathbf{R}_\phi$ are obtained from the covariance for the power observations, which according to equation (7.120) in Appendix 7.A is

$$r_{p_y}(\tau) = \mathrm{E}\{(p_y(t)-\bar{p}_y)(p_y(t-\tau)-\bar{p}_y)\} = |r_h(\tau)+r_e(\tau)|^2\ . \tag{7.62}$$

The covariance matrix for the power regressor $\mathbf{R}_\phi$ can thus be expressed in terms of the covariance matrix for the complex regressor,

$$\mathbf{R}_\phi = \begin{bmatrix} \sigma^2_{p_y} & r_{p_y}(-\Delta t) & & r_{p_y}((1-M)\Delta t) \\ r_{p_y}(\Delta t) & \sigma^2_{p_y} & & \\ \vdots & & \ddots & \\ r_{p_y}((M-1)\Delta t) & & & \sigma^2_{p_y} \end{bmatrix} = \mathbf{R}_\varphi \odot \mathbf{R}^*_\varphi\ . \tag{7.63}$$

According to the results above we can express the optimal linear FIR-predictor using the observed power in the regressor as

$$\hat{p}(t+L|t) = \boldsymbol{\phi}^T(t)\boldsymbol{\vartheta}_p + \sigma^2_h\ , \tag{7.64}$$

where

$$\boldsymbol{\vartheta}_p = \mathbf{R}_\phi^{-1}\mathbf{r}_{p\phi} = (\mathbf{R}_\varphi \odot \mathbf{R}^*_\varphi)^{-1}(\mathbf{r}_{h\varphi} \odot \mathbf{r}^*_{h\varphi}) \tag{7.65}$$

and the minimum MSE is obtained as for the complex predictor in (6.13)

$$\begin{aligned}\sigma^2_{\epsilon_p} &= \sigma^2_p - \mathbf{r}^H_{p\phi}\mathbf{R}^{-1}_\phi \mathbf{r}_{p\phi} \\ &= \sigma^4_h - (\mathbf{r}_{h\varphi} \odot \mathbf{r}^*_{h\varphi})^H(\mathbf{R}_\varphi \odot \mathbf{R}^*_\varphi)^{-1}(\mathbf{r}_{h\varphi} \odot \mathbf{r}^*_{h\varphi})\ . \end{aligned} \tag{7.66}$$

The NMSE for the predicted power is thus, for $\mathrm{E}\{p^2\} = \mathrm{E}\{|h|^4\} = 2\sigma^4_h$ since $h(t) \in \mathbf{CN}(0, \sigma^2_h)$,

$$V_{linear} = \frac{\sigma^2_{\epsilon_p}}{2\sigma^4_h} = \frac{1}{2} - \frac{(\mathbf{r}_{h\varphi} \odot \mathbf{r}^*_{h\varphi})^H(\mathbf{R}_\varphi \odot \mathbf{R}^*_\varphi)^{-1}(\mathbf{r}_{h\varphi} \odot \mathbf{r}^*_{h\varphi})}{2\sigma^4_h}. \tag{7.67}$$

The performance for the linear power predictor, using observed powers in the regressor, can thus be expressed in terms of the covariance and cross-covariance for the underlying complex signal.

7.5.2 Performance comparison

The results for the linear power predictor of Section 7.5.1 should be compared to the results obtained using complex observations in the regressor (7.19). When only one coefficient is used in the predictors, that is $M = 1$, the two predictors actually are the same.

As the auto-covariance for the power is related to the auto-covariance for the complex time series as $r_p(\tau) = |r_h(\tau)|^2$, for the complex Gaussian $h(t)$, we see that the correlation decays faster with an increasing lag τ for the power than for the complex time series. Consequently, it is harder to predict the power than the complex time series.

Example 7.3 Power prediction of an AR2-process

In this example we will compare the performance of the linear power predictor (7.64), the quadratic power predictor (7.17) and the biased quadratic power predictor (7.2), when the complex signal is generated by a second order AR-process driven by white complex Gaussian noise. The delay spacing Δt is set to one. The AR2-process has two poles on the distance $\rho < 1$ from the origin at the angles $\pm\vartheta$. The difference equation for the signal is

$$\begin{aligned} h(t) &= 2\rho\cos\vartheta\, h(t-1) - \rho^2 h(t-2) + w(t) \\ y(t) &= h(t) + e(t) \end{aligned} \tag{7.68}$$

where $w(t)$ is the innovation process with variance σ_w^2 and the measurement noise $e(t)$ is white with variance σ_e^2. Let the variance σ_w^2, of the driving noise be normalized as $\sigma_w^2 = (1-\rho^2)(1-2\rho^2\cos 2\vartheta + \rho^4)/(1+\rho^2)$ to make $\sigma_h^2 = 1$. The auto-covariance for the signal $h(t)$ is then

$$r_h(\tau) = \frac{(1-\rho^2)\cos\vartheta\sin(\vartheta|\tau|) + (1+\rho^2)\cos(\vartheta\tau)\sin\vartheta}{(1+\rho^2)\sin\vartheta}\rho^{|\tau|}. \tag{7.69}$$

As the measurement noise is white, the cross-correlation $\mathbf{r}_{h\varphi}$, between $h(t)$ and the regressor $\boldsymbol{\varphi}(t-L)$, with complex observations, only depends on the auto-correlation for $h(t)$,

$$\mathbf{r}_{h\varphi} = [r_h(L)\; r_h(L+1)]^T. \tag{7.70}$$

The covariance matrix for the regressor is

$$\mathbf{R}_\varphi = \mathbf{R}_h + \sigma_e^2 \mathbf{I}. \tag{7.71}$$

Let $\rho = 0.9748$ and $\vartheta = 0.4417$, with $\sigma_w^2 = 0.0178$ to make $\sigma_h^2 = 1$, in this example. Then the auto-covariance (7.69) is close to $\mathrm{J}_0(2\pi f_D \tau)$, with $f_D = 0.1$, for the first four lags. The time series $h(t)$ is thus an approximation of a Rayleigh fading tap with Jakes spectrum.

The predictors have only two coefficients ($M = 2$). In the absence of measurement noise, $\sigma_e^2 = 0$, this is sufficient to obtain the best possible prediction of the complex time series. We will in the following look at the one-step predictor ($L = 1$).

In Figure 7.5 the NMSE for the three power predictors are shown as a function of the SNR (σ_e^2 is varied). The NMSE for the predictions are evaluated using the equations (7.23), (7.24) and (7.67). For low SNR

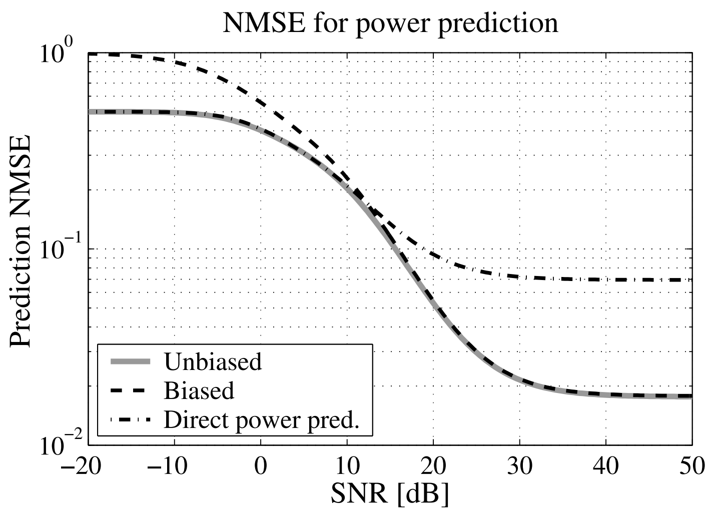

Figure 7.5: A plot of the NMSE for one step power prediction of an AR2-process as a function of SNR for the unbiased quadratic predictor (7.17), (solid line), the biased quadratic predictor (7.2), (dashed) and the linear power predictor (7.64), (dash-dotted).

the unbiased quadratic predictor and the linear power predictor have the same performance with an NMSE equal to one half (the same as if only the average is known). This is because both these predictors converge towards the average predictor $\hat{p}(t+1|t) = \sigma_h^2$, when the SNR is decreased. In contrast, the biased power predictor has an NMSE of one. For high SNR we approach the noise free situation where the NMSE for the linear power predictor (7.64) is 0.0695 according to (7.67). The unbiased quadratic power predictor (7.17) and the biased quadratic power predictor (7.2) then almost coincide, as the complex one-step prediction error will then only consist of the small innovation $w(t)$, with NMSEs of 0.0176 and 0.0178 respectively (evaluated using (7.24) and (7.23) with the complex variance given by $\sigma_{\varepsilon_c}^2 = \sigma_w^2$).

The improvement when using the unbiased quadratic predictor as compared to the other two is clearly seen in Figure 7.6, where the relative gain in dB is shown. The performance of the unbiased quadratic predictor is

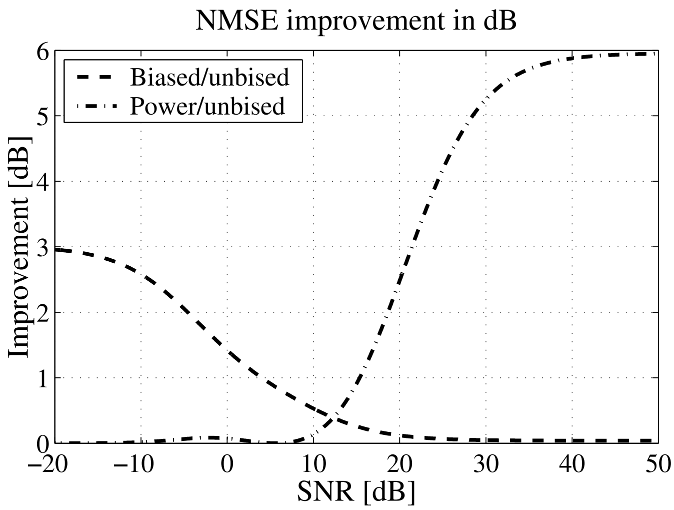

Figure 7.6: The improvement in NMSE, measured in dB, achieved by using the unbiased quadratic predictor as compared to the biased (dashed) or the linear power predictor (dash-dotted).

similar to that of the linear power predictor up to an SNR of 10dB, as both predictors then mainly rely on using the average power as the prediction. At an SNR of 40dB the difference is already close to 6dB. The opposite holds when comparing with the biased quadratic predictor, which has a performance that is 3dB worse than the unbiased quadratic predictor for low SNR, where the biased predictor sets the predicted power to zero and the unbiased quadratic predictor sets it to the average power. For high

SNR they have almost equal performance, as then the complex prediction error is low and the bias correspondingly small. At high SNR, it is always beneficial to predict the power based on the complex time series as compared to using the power directly.

7.6 The Average Power

In Section 7.5 it was assumed that the average power is known. Generally this is not the case and the average power has to be estimated. In the following the statistical properties of a local time average of the power will be examined for the Jakes channel (Section 7.6.1). The properties of this estimate as a power predictor is investigated in Section 7.6.2. The analysis is performed on a continuously observed power with no noise. If the channel is sampled at a high rate, then the averaging of power samples will give similar results to those obtained with the continuous average power estimator. The effect of an additive noise is not treated here but it can be shown to cause a bias in the average power estimate.

7.6.1 Estimation of the average power

The estimation of the average power is based on the observed power. These are sampled and noisy observations. In the following we will see that power estimation based on local averages over a few wavelengths of traveled distance are quite inaccurate even when the continuous power, without noise, can be observed.

A local time-average of the power is obtained by summing the observations over a time interval T and divide by the length of the interval. If the number of terms in the sum is large and the sampling rate high, the sum can be approximated by the integral of the observations as

$$\bar{p}_T = \frac{1}{T}\int_0^T |h(t)|^2 dt. \tag{7.72}$$

The deviation from the the true average power is

$$\varepsilon_{\bar{p}_T} = \bar{p} - \bar{p}_T \tag{7.73}$$

The expectation of the estimation error is zero as the expectation of the local average is the average power ($\mathrm{E}\{\bar{p}_T\} = \bar{p}$). The variance of the estimation error is

$$\sigma^2_{\varepsilon_{\bar{p}_T}} = \mathrm{E}\{\varepsilon^2_{\bar{p}_T}\} = \mathrm{E}\{\bar{p}^2_T\} - \bar{p}^2, \tag{7.74}$$

which depends on $\mathrm{E}\{\bar{p}^2_T\}$. Apply (7.72) to form

$$\begin{aligned}\mathrm{E}\{\bar{p}^2_T\} &= \mathrm{E}\left\{\frac{1}{T^2}\iint_0^T |h(t)|^2|h(\tau)|^2 dtd\tau\right\} \\ &= \frac{1}{T^2}\iint_0^T (\sigma_h^4 + |r_h(t-\tau)|^2)dtd\tau,\end{aligned} \tag{7.75}$$

where we have used (7.100) in Appendix 7.A to evaluate the expectation of the product of the absolute squares of two stochastic variables with zero mean and circular complex Gaussian distributions. As $\sigma_h^4 = \bar{p}^2$ does not depend on t or τ, that term can be put outside the integral. The variance of the error can thus be expressed as

$$\mathrm{E}\{\varepsilon^2_{\bar{p}_T}\} = \frac{1}{T^2}\iint_0^T |r_h(t-\tau)|^2 dtd\tau = \frac{2}{T^2}\int_0^T\int_0^\tau |r_h(t)|^2 dtd\tau. \tag{7.76}$$

The variance of the estimation error for the average power is thus given by a double integral over the squared covariance of the complex signal. For Jakes model the covariance is given by $r_h(\tau) = \mathrm{J}_0(2\pi f_D \tau)$ for which (7.76) is plotted in Figure 7.7. For a Jakes channel the estimation interval for average power estimation has to be at least five wavelengths to obtain an estimation error variance that is below 0.1. To obtain a variance below 0.01 the interval is must be above 80 wavelengths. It is thus not sufficient to use short estimation intervals of a few wavelengths, to obtain good estimates of the average power.

7.6.2 Using the average power as prediction

The local average (7.72) can be used as a power predictor. Define the running power average as

$$\bar{p}_T(t) = \frac{1}{T}\int_{t-T}^t p(\tau)d\tau. \tag{7.77}$$

If this power estimate is used to predict the power a time L ahead, then the prediction error is

$$\varepsilon_{\bar{p}_T}(t) = p(t) - \bar{p}_T(t-L), \tag{7.78}$$

which has zero average. It is thus an unbiased predictor (in the case of no noise). The MSE of this prediction, which is equal to the variance of the

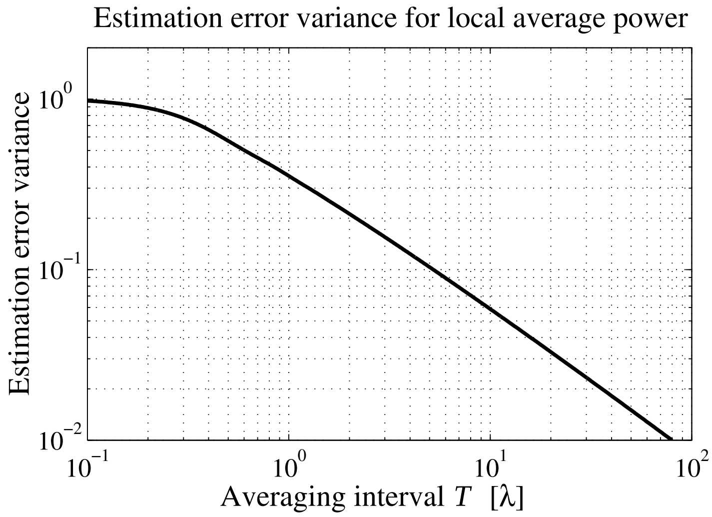

Figure 7.7: Variance of the estimation error for the local estimated average power for a Jakes channel with $\bar{p} = 1$.

prediction error, is

$$\begin{aligned} \mathrm{E}\{\varepsilon_{\bar{p}_T}^2\} &= \mathrm{E}\left\{(p(t) - \bar{p}_T(t))^2\right\} \\ &= \mathrm{E}\{p^2(t)\} - 2\mathrm{E}\{p(t)\bar{p}_T(t-L)\} + \mathrm{E}\{\bar{p}_T^2(t-L)\}. \end{aligned} \tag{7.79}$$

This expression is treated term by term. As $p(t) = |h(t)|^2$ and $h(t) \in \mathbf{CN}(0, \sigma_h^2)$, we have that the first term is $2\sigma_h^4 = 2\bar{p}^2$. The second term

$$\mathrm{E}\{p(t)\bar{p}_T(t-L)\} = \frac{1}{T}\int_{t-T-L}^{t-L} \underbrace{\mathrm{E}\{p(t)p(\tau)\}}_{\bar{p}^2 + |r_h(t-\tau)|^2} d\tau = \bar{p}^2 + \frac{1}{T}\int_L^{T+L} |r_h(\rho)|^2 d\rho, \tag{7.80}$$

where we have applied (7.101) and made the variable substitution $\rho = t - \tau$ in the last integral. If L is large, then the corresponding correlation will be small and thus the integral above will also be small. The last term of (7.79) is given by (7.75). The total MSE of the predictor in (7.79) can thus, using (7.79), (7.80) and (7.75), be expressed as

$$\begin{aligned} \mathrm{E}\{\varepsilon_{\bar{p}_T}^2\} &= 2\bar{p}^2 - 2\bar{p}^2 - \frac{2}{T}\int_L^{T+L} |r_h(\tau)|^2 d\tau + \bar{p}^2 + \frac{1}{T^2}\iint_0^T |r_h(t-\tau)|^2 dt d\tau \\ &= \bar{p}^2 + \frac{1}{T^2}\iint_0^T |r_h(t-\tau)|^2 dt d\tau - \frac{2}{T}\int_L^{T+L} |r_h(\tau)|^2 d\tau. \end{aligned} \tag{7.81}$$

As the second order moment of the power is $\mathrm{E}\{p^2(t)\} = 2\bar{p}^2$ the NMSE for this power predictor is obtained by dividing (7.81) by $2\bar{p}^2$. Figure 7.8 shows the NMSE for the Jakes model as a function of the length of the window T and different prediction ranges L. Here both window length and range are expressed as corresponding traveled distances as measured in wavelengths. The results in Figure 7.8 are for an ideal case with continuous

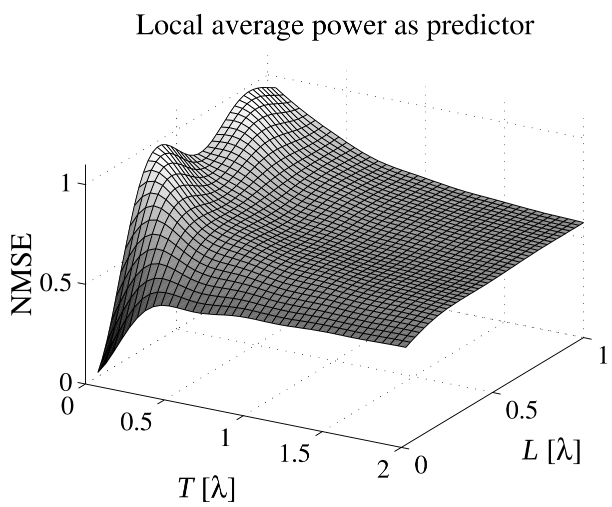

Figure 7.8: NMSE of the power prediction using the estimated local average power as predictor.

observations and no noise. For longer estimation intervals T, the prediction NMSE approaches that of the true average predictor, that is 0.5 independent of prediction range. For a short prediction range $L < 0.1$, and a short estimation interval $T < 0.2$, the NMSE of the local average as predictor actually is better than 0.5. This is due to that the estimation interval is sufficiently correlated to the power L steps ahead. The short estimation interval makes the predictor similar to using the last observed power sample as predictor.

Even under these favorable conditions an average over observations for a quite long distances is needed to obtain good estimates of the average power.

7.6.3 The last sample as predictor

When a estimation interval of just one sample is used in the average predictor, we obtain the last sample predictor. It relies on an assumption that the power does not change that much over the prediction range, so the last observations of the power is a good indication of what it will be later.

The prediction error for the last sample predictor, in a noiseless case, is

$$\varepsilon_p(t) = p(t) - p(t-L). \tag{7.82}$$

It is a zero mean predictor and the variance is

$$\begin{aligned}\sigma^2_{\varepsilon_p} &= \mathrm{E}\{\varepsilon^2_p(t)\} = \mathrm{E}\{p^2(t) - 2p(t)p(t-L) + p^2(t-L)\} \\ &= 2\sigma^4_h - 2(\sigma^4_h + |r_h(L)|^2) + 2\sigma^4_h = 2\sigma^4_h - 2|r_h(L)|^2. \end{aligned} \tag{7.83}$$

The NMSE for the prediction error using the last sample as prediction, is thus

$$V = \frac{\sigma^2_{\varepsilon_p}}{\sigma^2_p} = 1 - \frac{|r_h(L)|^2}{\sigma^4_h}. \tag{7.84}$$

The NMSE for the last sample predictor for a Jakes channel, is seen in Figure 7.9. The prediction range has to be below 0.07 wavelengths to obtain an NMSE under 0.1. For ranges beyond 0.18 wavelengths it is even worse than just using the average power as prediction. The last sample predictor is thus only an alternative for very short prediction ranges, even under such favorable conditions as no noise.

7.6.4 Frequency selective channels

A frequency selective channel can be modeled as a time varying FIR-filter where each tap can be seen as flat Rayleigh fading channel. The total power is thus the sum of the power contribution from each tap,

$$p(t) = \sum_{k=1}^{M} p_k(t) = \sum_{k=1}^{M} |h_k(t)|^2. \tag{7.85}$$

In the following section we will examine two types of channels, one where all the taps has equal average power and one where the taps are ordered after an exponentially decaying average power, see Figure 7.10.

The total average power is just the sum of the average power for each tap

$$\bar{p} = \mathrm{E}\{p(t)\} = \sum_{k=1}^{M} \underbrace{\mathrm{E}\{|h_k(t)|^2\}}_{\bar{p}_k = \sigma^2_{h_k}} = \sum_{k=1}^{M} \bar{p}_k, \tag{7.86}$$

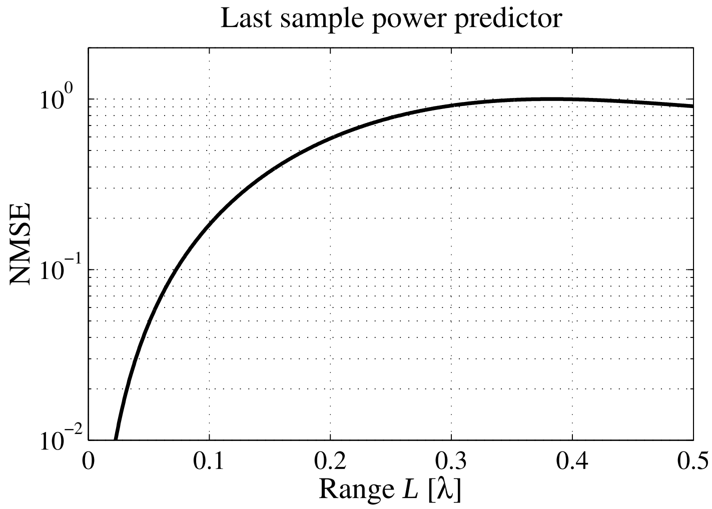

Figure 7.9: The NMSE of the power prediction of a Jakes channel using the last observed sample as predictor.

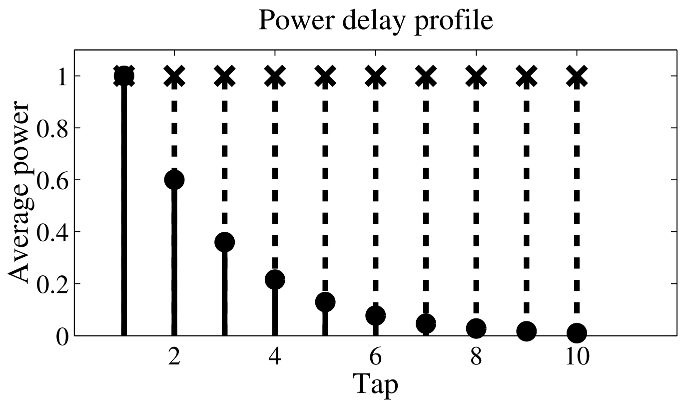

Figure 7.10: The power delay profile for two frequency selective channels: One with ten on average equally strong taps and the other with exponential decaying average powers of the taps.

whereas the second order moment of the power is expressed as

$$\mathrm{E}\{p^2(t)\} = \sum_{k=1}^{M}\sum_{l=1}^{M}\underbrace{\mathrm{E}\{|h_k(t)|^2|h_l(t)|^2\}}_{\bar{p}_k\bar{p}_l+|r_{h_kh_l}|^2} = \bar{p}^2 + \sum_{k=1}^{M}\sum_{l=1}^{M}|r_{h_kh_l}|^2, \tag{7.87}$$

where we have applied (7.101) and where $r_{h_k h_l}$ is the cross-covariance between tap k and l. The variance for the power is obtained as

$$\mathrm{E}\{(p(t)-\bar{p})^2\} = \mathrm{E}\{p^2(t)\} - \bar{p}^2 = \sum_{k=1}^{M}\sum_{l=1}^{M} |r_{h_k h_l}|^2. \tag{7.88}$$

It thus depends on the variance and cross-covariance of the complex taps.

Uncorrelated taps

In the following we will assume that the taps of the impulse response are uncorrelated, that is

$$r_{h_k h_l} = \delta_{kl} r_{h_k} = \delta_{kl}\bar{p}_k. \tag{7.89}$$

This assumption simplifies the analysis significantly, but it is not valid when pulse-shaping of the transmitted symbols are used. Then de-correlation of the taps can be performed in a first step to obtain uncorrelated taps.

With uncorrelated taps the second order moment for the power is

$$\mathrm{E}\{p^2(t)\} = \bar{p}^2 + \sum_{k=1}^{M} \bar{p}_k^2, \tag{7.90}$$

and the variance for the power is obtained as

$$\mathrm{E}\{(p(t)-\bar{p})^2\} = \mathrm{E}\{p^2(t)\} - \bar{p}^2 = \sum_{k=1}^{M} \bar{p}_k^2. \tag{7.91}$$

In the two examples below we can see what the power contribution from a number of uncorrelated taps means for the power prediction performance when using the average power as predictor.

Prediction using the average power

The NMSE for using the true average power $\bar{p}$, as a power predictor can in the case of uncorrelated taps be obtained by using (7.86) in (7.90) and (7.91) as

$$V = \frac{\mathrm{E}\{(p(t)-\bar{p})^2\}}{\mathrm{E}\{p^2(t)\}} = \frac{\sum_k \bar{p}_k^2}{\left(\sum_k \bar{p}_k\right)^2 + \sum_k \bar{p}_k^2} \tag{7.92}$$

Example 7.4 Taps with equal power

This channel has M uncorrelated taps, all of which has the same average power, $\bar{p}_k = 1$. Insert this into (7.92) and the NMSE of the total average as a predictor is obtained as

$$V = \frac{\sum_k 1^2}{\left(\sum_k 1\right)^2 + \sum_k 1^2} = \frac{M}{M^2 + M} = \frac{1}{M+1}. \tag{7.93}$$

The NMSE decreases as the number of taps M increases. For $M = 1$ we obtain $V = 0.5$, as expected.

Here one of the obvious advantages with broadband radio transmission is seen. The higher the bandwidth, the larger the number of significant taps. With a large number of contributing taps, the variation of the total power will decrease as it is unlikely that all the taps fade simultaneously. When the variation of the power level decreases, the average power will work better as a predictor.

Example 7.5 Taps with exponentially decaying average power

In this channel the taps are uncorrelated and have an exponentially decaying average power $\bar{p}_k = ba^k$ where $0 < a < 1$ and $0 \leq b$. The total average power is

$$\bar{p} = \sum_{k=1}^{M} \bar{p}_k = \sum_{k=1}^{M} ba^k = ba\frac{1 - a^M}{1 - a}. \tag{7.94}$$

The sum of the squared powers is

$$\sum_{k=1}^{M} \bar{p}_k^2 = \sum_{k=1}^{M} b^2 a^{2k} = b^2 a^2 \frac{1 - a^{2M}}{1 - a^2}. \tag{7.95}$$

By using (7.94) and (7.95) the NMSE (7.92), for a channel with exponential decaying PDP, can be expressed as

$$V = \frac{b^2 a^2 \frac{1-a^{2M}}{1-a^2}}{\left(ba\frac{1-a^M}{1-a}\right)^2 + b^2 a^2 \frac{1-a^{2M}}{1-a^2}} = \frac{(1 + a^M)(1 - a)}{2(1 - a^{M+1})}. \tag{7.96}$$

Note that $M = 1$ result in $V = 0.5$. If we let the number of taps go to infinity, then the exponential terms will tend to zero and (7.96) becomes

$$\lim_{M\to\infty} V = \frac{1-a}{2}. \tag{7.97}$$

Thus, the NMSE for the average power predictor converges. If there is a large number of strong taps, that is, a is close to one, then there is a lot of diversity and the power of the channel will vary less. In that case the average power is a good predictor. If a is close to zero, then only a few taps will contribute significantly to the power. In that case the variation of the power level will be significant and the average predictor will render a poor performance with an NMSE close to 0.5.

In both these examples we see that if there are many taps that carry power in the impulse response, then the relative variability of the total power will be reduced. This improves the performance of the average power as a predictor for frequency selective channels.

7.7 Results

7.7.1 Simulations

Just as in the evaluation of the performance for complex channel prediction in Section 6.7.1 the Jakes model is studied. The expected performance for power prediction of a flat Rayleigh fading channel with Jakes spectrum will indicate the performance on the prediction of the absolute square of taps of the measured channels.

Prediction performance with no noise reduction

The delay spacing that is optimal for complex prediction of a tap is also optimal for prediction of the absolute square of a tap.[2] When the optimal delay spacing is obtained as in Figure 6.7, the optimal power prediction NMSE for the predictor can be calculated for each SNR for the unbiased quadratic predictor (7.16). In Figure 7.11 and 7.12 the optimal NMSE for the quadratic unbiased predictor of the power using complex observations in

[2]This follows from equation (7.20) where the variance for the power prediction error is expressed in terms of the variance of the complex prediction error. In the range $0 \leq \sigma^2_{\varepsilon_c} \leq 1$ this function obtain the minimum value for the smallest possible $\sigma^2_{\varepsilon_c}$.

the regressors and linear prediction using observed power in the regressor, of a Rayleigh fading tap with Jakes spectrum contaminated by white noise is depicted. As seen in Figure 7.12 the direct linear power predictor is outperformed by the unbiased quadratic predictor, when the prediction range is a quarter of a wavelength. This is expected as the correlation function for the power decays much faster than for the complex tap.

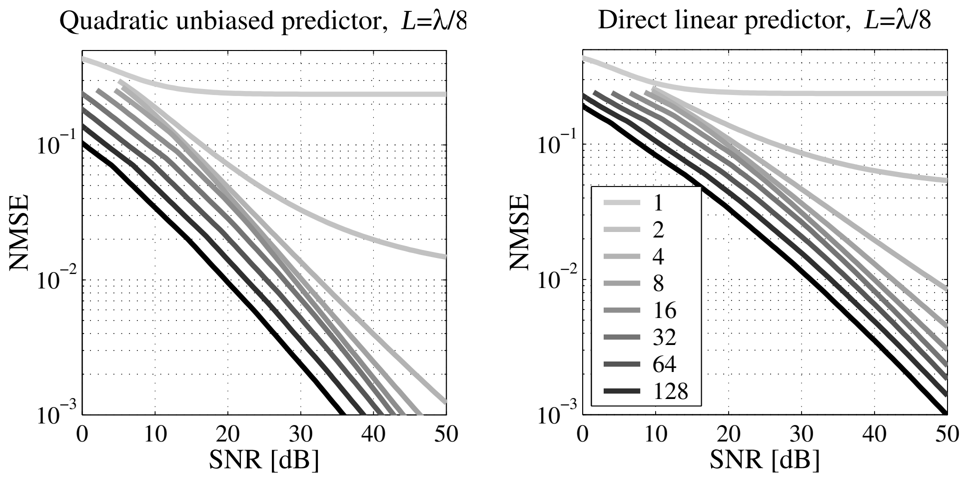

Figure 7.11: The optimal NMSE for power prediction one eighth of a wavelength ahead, using the unbiased quadratic (left) and the linear power predictor (right). Performance improve as the number of coefficients increase.

The achievable performance depends on the number of coefficients M, that is the complexity. If there is a target NMSE, then the curves in Figures 7.11 and 7.12 can be used to find the necessary complexity at a given SNR. For the prediction NMSE of 10^{-1} and 10^{-2}, the Table 7.1 list how much lower SNR that is tolerated when the number of coefficients are doubled, e.g. for a prediction range of $\lambda/4$ a doubling of the number of coefficients from 64 to 128 will render an NMSE lower than 10^{-2} even if the SNR decreases by 2.1 dB. These results are valid for taps described by the Jakes model that are corrupted by white noise. No noise reduction is applied.

Prediction performance with noise reduction

Figure 7.13 shows how predictable the absolute square of a tap described by the Jakes model is, with and without smoothed regressors. The OSR is 50 and the unbiased quadratic predictor (7.16) is based on an FIR predictor for

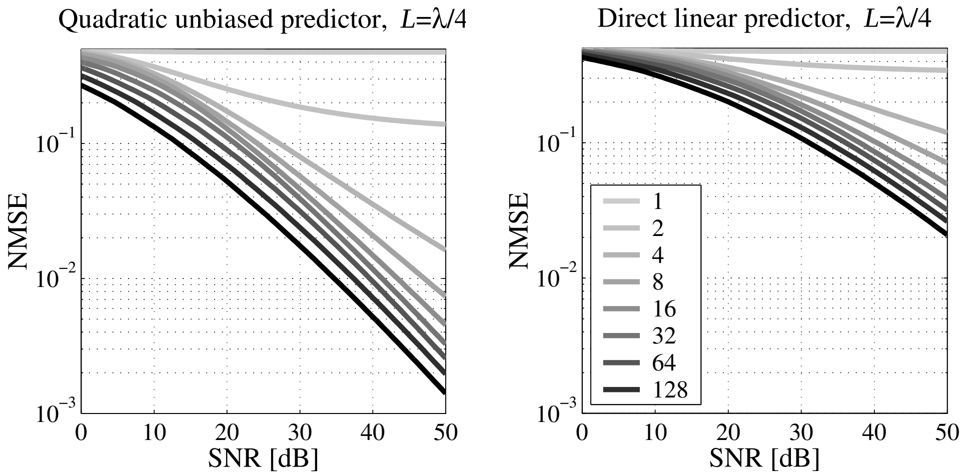

Figure 7.12: The optimal NMSE for power prediction one quarter of a wavelength ahead, using the unbiased quadratic (left) and the linear power predictor (right).

Prediction range = 0.125λ							
Doubling M		2-4	4-8	8-16	16-32	32-64	64-128
SNR gain [dB]	10^{-1}	3.0	0.6	1.6	2.5	2.7	2.9
at NMSE	10^{-2}		2.6	1.2	1.3	2.3	2.7
Prediction range = 0.25λ							
Doubling M		2-4	4-8	8-16	16-32	32-64	64-128
SNR gain [dB]	10^{-1}		4.4	2.2	1.4	2.0	2.5
at NMSE	10^{-2}				2.5	1.7	2.1

Table 7.1: The improved tolerance towards noise (error on the channel estimates) at a given NMSE when the prediction ranges are one eighth and a quarter of a wavelength. The empty cells occur when the prescribed NMSE is not achieved for an SNR below 50dB.

the complex taps, as described in Section 6.7.1. The complex predictor has eight coefficients and a robust choice of delay spacing, given by Table 6.2. The noise reduction is obtained using a bank of FIR smoother with 128 coefficients and the regressor is formed as in Section 6.2.3. By using smoothed regressors the unbiased quadratic predictor, with eight coefficients, can attain the same performance as the unbiased quadratic predictor using no noise reduction at 7-9 dB lower SNR.

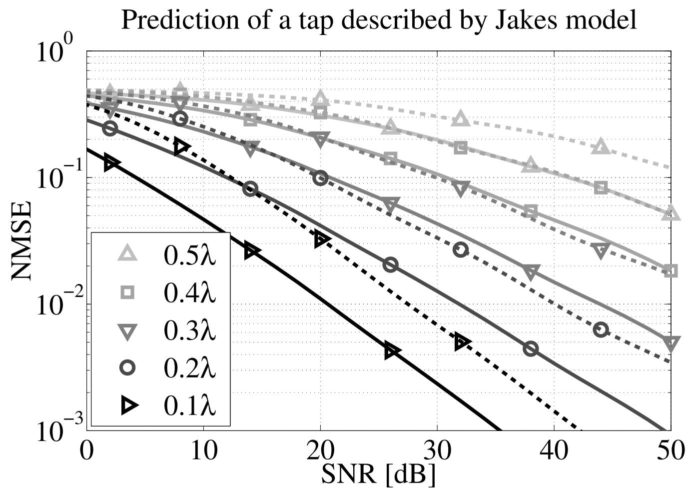

Figure 7.13: The power prediction NMSE as a function of SNR for different prediction ranges. The tap is described by the Jakes model. The unbiased quadratic predictor is based on a complex FIR predictor with 8 coefficients and the performance is shown for smoothed regressor (solid lines) and no noise reduction (dashed line).

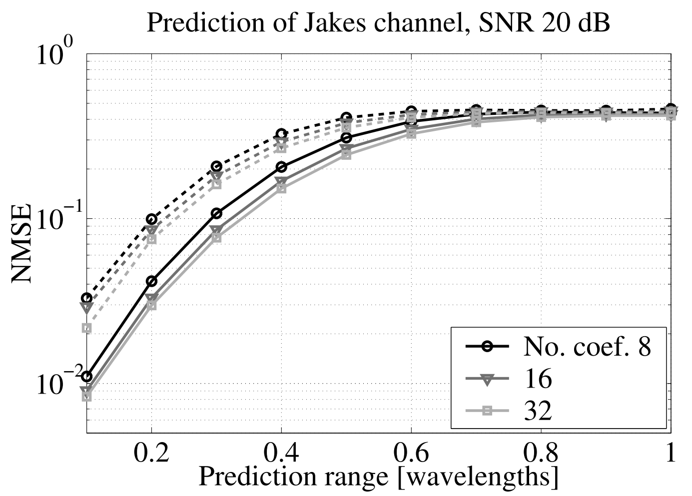

Figure 7.14: The power prediction NMSE at 20 dB for a tap described by the Jakes model as a function of prediction range. Solid and dashed lines corresponds to smoothed and noisy regressors respectively.

7.7.2 Power prediction on measured channels

To establish the performance of the predictors treated in this Chapter, the predictors are tested on measured channel impulse responses, described in Chapter 3. The quadratic power predictors are based on the complex tap predictors described in Section 6.7.2. The noise reduction, delay spacing and choice of measurements used for the performance evaluation are also described in Section 6.7.2. Once the predictions of the absolute square of the individual taps are obtained the prediction of the total power in the channel is obtained as the sum of the predicted tap powers (7.34).

Quadratic power prediction

A power prediction can be formed from the prediction of the complex tap as the absolute square of the complex prediction with an added compensation for the bias, as seen in (7.16). When the model based predictors are used the bias is not given by equation (7.4) or (7.5). The bias then has to be estimated directly from the data in a separate step.

When the LS-estimated FIR predictor for the tap (6.48) is used the unbiased power predictor can be obtained just using the estimated sample covariances, $\hat{\mathbf{R}}_{\varphi_k}$ from (6.46) and $\hat{\mathbf{r}}_{h_k\varphi_k}$ from (6.47). The prediction of the power of the tap is then obtained as in (7.16) by substituting the true covariances by the estimated. This is an unbiased power predictor under the assumption of complex Gaussian taps. To further improve the performance, smoothed taps are used as regressors.

For the model based predictors the AR models are estimated using least squares and has a delay spacing of 0.1 wavelengths. The model estimation is performed on smoothed taps, using a lag of 20 samples. The prediction is performed on smoothed taps with lag 0. The power bias has to be estimated in a separate step. From Section 7.6 we know that quite long estimation intervals are needed to obtain reliable estimates for the average power. The same should hold for estimation of the bias. The bias compensation thus might actually reduce the prediction performance, especially when the variance for the complex prediction error is small.

Independent of the prediction range, 900 samples are used in the target vector for the estimation of the predictor coefficients. The validation interval was on the order of 400 down to 100 samples, depending on prediction range and the Doppler frequency.

Average and last sample prediction

The two simplest power predictors, the average power and the last sample predictor, described in Section 7.6, are also included in the evaluation. The average power is estimated using the same training interval as for the more advanced predictors and it is evaluated on the same validation interval. The performance of the average power as predictor says something about the variability of the power. The better the average power is a as predictor the less does the channel power vary.

7.7.3 Performance on the measured taps

The power prediction NMSE for all the taps in the measured channels are evaluated using the unbiased quadratic power predictor based on the LS-estimated FIR predictor for the complex taps, with eight coefficients and smoothed regressors. The scatter plots in Figure 7.15 show the NMSE as a function of the tap-to-estimation error SNR. The solid curve is the median NMSE for the predicted taps within an SNR-bin of width 6 dB. The dashed line is the corresponding theoretical NMSE for a tap described by the Jakes model, using eight coefficients and the same noise reduction (the OSR is 50). The spread in performance increases with the prediction range. From Figure 6.15 we know that taps with small Doppler spread show a higher predictability. This is also the case for power prediction. The results shown in Figure 7.15 for the prediction ranges 0.1 and 0.3 wavelength are summarized for all the prediction ranges 0.1-0.5 wavelength in Figure 7.16. The power prediction median NMSE that is achieved on the measured taps is in the vicinity of that for a Jakes tap. For high SNR the obtained NMSE is worse than the theoretical, for short prediction ranges. For longer ranges the achieved SNR is actually lower than for a Jakes tap.

Another view of the same data is presented in Figure 7.17. The taps with SNR in the range 32-38 dB have a median NMSE about 0.01 at prediction range of 0.2 wavelengths. At the same range taps with SNR in the range 8-14 dB has a median NMSE close to ten times higher. At a given prediction NMSE performance constraint, taps with high SNR can be predicted much longer as compared to taps with low SNR. To obtain channel estimates with as low estimation error as possible is thus crucial for the success of the prediction scheme.

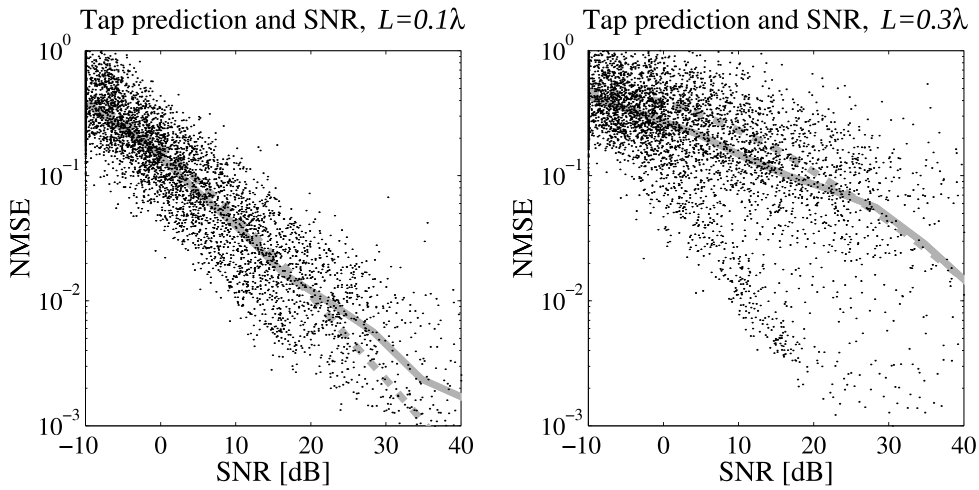

Figure 7.15: Scatter plots of the tap power NMSE for all the measured taps as a function of SNR for the unbiased quadratic power predictor based on the complex tap FIR predictor with 8 coefficients described in Section 6.7. The solid gray line is the median NMSE in a SNR bin of width 6 dB and the dashed gray line is the theoretical power prediction NMSE for a tap described by the Jakes model, for an OSR of 50. The prediction range is 0.1 wavelength (to the left) and 0.3 wavelengths (to the right).

7.7.4 Performance on the measured channels

The power predictions of the individual taps at each measurement location are summed to obtain the power prediction of the corresponding channels. In Figure 7.18 the power prediction using the absolute square of the model based Kalman predictor for the complex taps, with added estimated bias compensation is compared to the unbiased quadratic predictor based on the FIR predictor for the complex taps. The median value of the power prediction NMSE for the 41 measurement locations, at each range, is used as performance measure. The performance of the power prediction based on model based prediction is similar to that for the power predictor based on FIR prediction of the complex channel. Even though the FIR predictor deploys smoothed regressors and is designed to minimize the prediction error at each range, it renders roughly the same performance as power prediction based on a complex Kalman predictor with an AR8 model for the dynamics. The predictor based on an AR4 model has a slightly worse performance, whereas the AR12 model does not render a corresponding improvement.

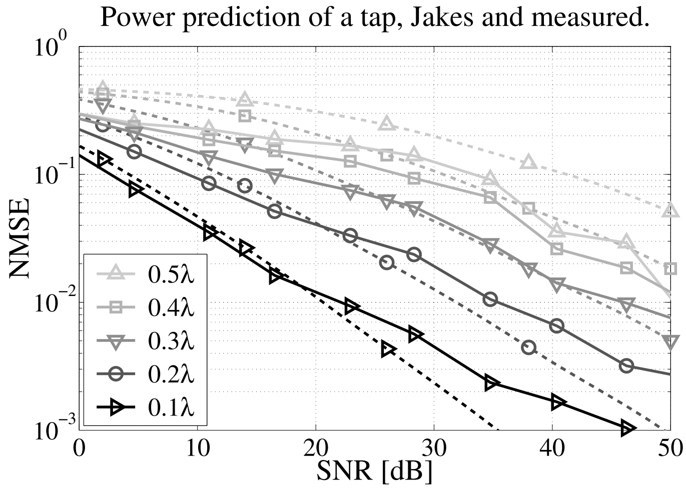

Figure 7.16: The power prediction NMSE as a function of channel estimation error SNR for different prediction ranges. The unbiased quadratic power predictor (solid lines) is based on an LS-estimated FIR predictor using smoothed regressors, with 8 coefficients, for the complex tap. The median value for the NMSE for taps with SNRs that fall within a bin of width 6 dB is plotted. For comparison the corresponding theoretical curves of Figure 7.13 for power prediction of a tap described by the Jakes model, using the same noise reduction and delay spacing, is included (dashed lines).

This can be due to estimation errors on the AR12 model. For these short estimation intervals, about 10 wavelengths long, larger models should thus be avoided.

Which type of predictor that should be used in not an obvious choice and complexity considerations have to be made. If only one prediction range is of interest, then the predictor based on the FIR predictor is probably the best choice. Whereas if all prediction ranges from 0.1 to 0.5 is of interest, then the Kalman based predictor might be preferable as the whole range of predictions are obtained at a very low extra computational cost.

The channel prediction NMSE for the unbiased quadratic predictor is compared to the biased predictor (the sum of the absolute square of the taps with no added bias compensation), the last sample predictor and the average power predictor in Figure 7.18. The unbiased power predictor provides the best performance for all prediction ranges, measured over the validation set. As the range L increases, its NMSE slowly approaches that of the average predictor from below. The biased power predictions and the last

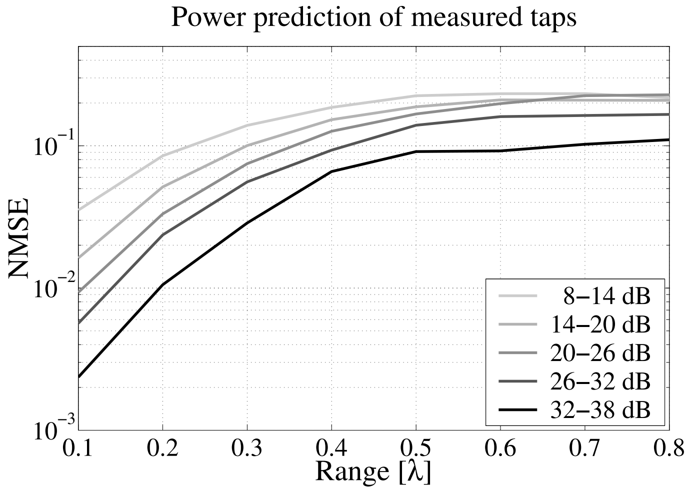

Figure 7.17: The median power prediction NMSE for an unbiased quadratic power predictor based on an LS-estimated FIR predictor using smoothed regressors, with 8 coefficients, for the complex tap. The five curves correspond to different SNRs on the complex taps. The higher the SNR is, the lower is the NMSE.

sample predictions become slightly worse than using the average power at large prediction ranges. The average power actually provides a quite good predictor with a median NMSE slightly below 0.1, as seen in Figure 7.19. This is due to the long impulse responses where a lot of taps carry a significant amount of power, as discussed in Section 7.6. To obtain better power prediction performance we have to rely on the quadratic power predictors that obtain prediction NMSE under 0.01 for the total power up to a quarter of a wavelength ahead.

The scatter plots in Figure 7.20 show the power prediction NMSE as a function of channel SNR. The unbiased quadratic predictor based on the FIR predictor with eight coefficients is used. The achieved power prediction NMSE for the channel depends on the channel SNR. The SNR is not the only factor influencing the prediction. The dynamics of the individual taps and the variability of the total power also play a role. A measure for how much the power varies is the performance of the estimated average as predictor. The scatter plot in Figure 7.21 show the NMSE for the sum of the predicted powers from the unbiased quadratic predictors as a function of the NMSE of the average power as predictor. For a short prediction range

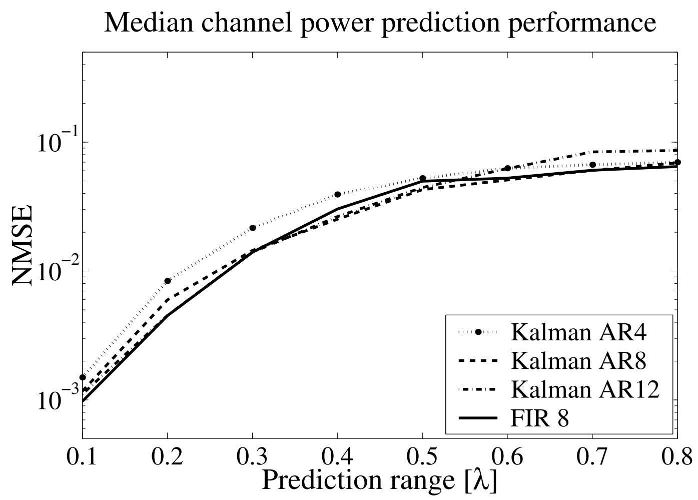

Figure 7.18: The median power prediction NMSE evaluated at 41 measurement locations. The power prediction is obtained based on complex predictions of the taps using model based or FIR predictors (with smoothed regressors).

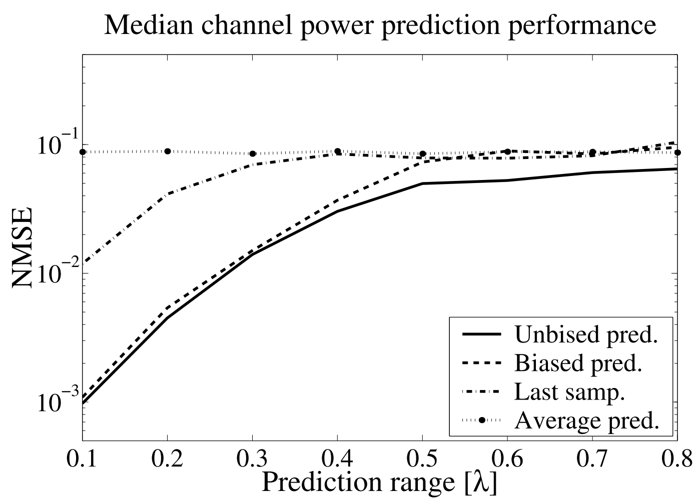

Figure 7.19: The average power prediction normalized MSE evaluated at 41 measurement locations.

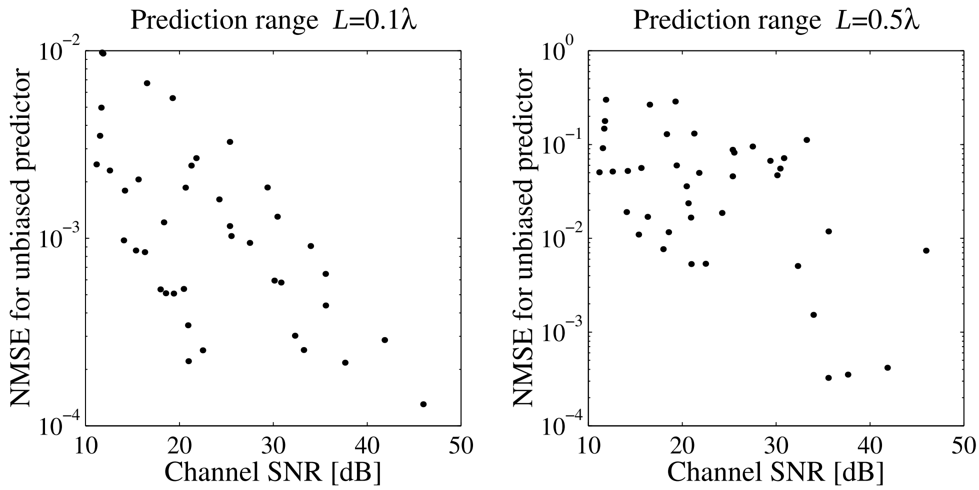

Figure 7.20: Scatter plots of the tap power NMSE for the unbiased predictor as a function of the channel SNR. A higher channel SNR result in better power prediction.

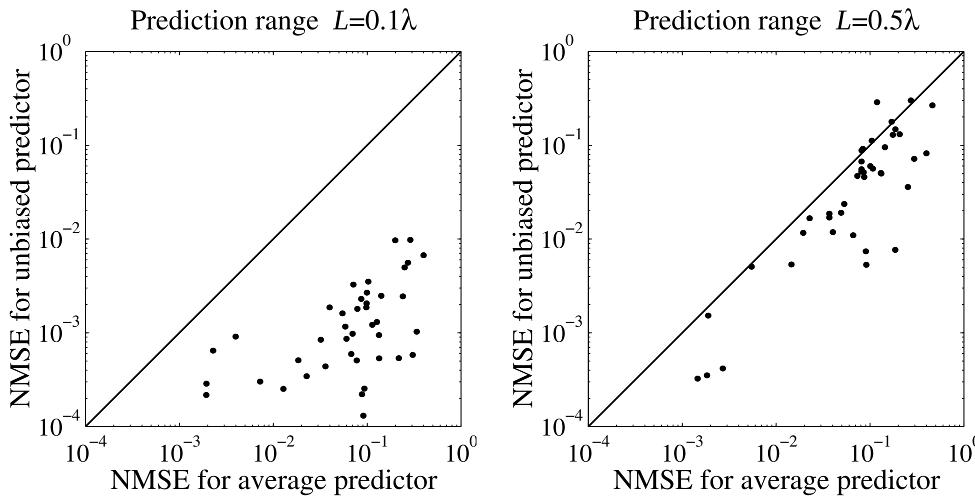

Figure 7.21: Scatter plots of the tap power NMSE for the unbiased predictor as a function of the NMSE for the average power as power predictor. The performance of the average power as predictor is directly dependent on how much the channel power actually varies. Dots below the line corresponds to channels where the unbiased power predictor outperforms the average power as predictor.

(L=0.1λ), the NMSE for the sum of the predicted powers from the unbiased quadratic predictors and the NMSE for the average power as predictor show a slight dependence. For a range of half a wavelength the prediction NMSE of the sum of the predicted powers from the quadratic unbiased predictors approaches that of the average power as predictor. This is an expected result as the complex prediction will reduce the amplification at such a long range. The bias compensation will thus play a larger role and when the amplification in the complex prediction tends to zero, the bias compensation tend to the average power. There is a clear advantage in using the sum of the predicted powers from the quadratic unbiased predictors as compared to using the average power as predictor independent of the variability of the power, especially for short prediction ranges.

7.8 Conclusion

- When a power signal originates from the squared magnitude of a complex signal corrupted by an independent additive noise, it is beneficial to use the observed complex signal to predict future values of the power as compared to using the observed power.

- A quadratic predictor of the power using the observed complex taps is designed.

  - To directly use the absolute square of a linear prediction of the complex time series renders a biased power prediction.
  - Using the optimal linear predictor, this bias will equal the variance for the complex prediction error.
  - The bias can be compensated for to form an unbiased quadratic predictor.
  - For a Rayleigh fading channel, that is complex Gaussian samples, the MMSE linear prediction of the complex time series also renders the MMSE unbiased power prediction.
  - When the complex valued time series is unpredictable the unbiased quadratic predictor renders the average power whereas the biased predictor predict zero power.

- The probability density functions for the power and the predicted power are important in the design of link adaptation systems.

  - The joint pdf for the true and predicted power is given by the bivariate gamma distribution.
  - With known auto-covariances for the complex signal and the additive noise the correlation between the true and predicted power is known.

- Just as the prediction coefficients are tuned to achieve the best possible prediction so should the delay spacing of the samples in the regressor be tuned. The delay spacing strongly influences the achieved performance.

- For a flat Rayleigh fading channel the correlation function for the power is the absolute square of the correlation function of the complex channel. The correlation function for the power thus falls of faster than for the complex channel. Power prediction based on the observed powers thus does not achieve as good performance as predictors based on the complex channel.

- The power prediction performance for an unbiased quadratic predictor on a tap described by the Jakes model can be greatly improved using noise reduction. For an OSR of 50, a bank of Wiener smoothers with 128 coefficients, the performance of a predictor using only eight coefficients achieves the same performance as a predictor not applying noise reduction at 7-9 dB lower channel-to-estimation error SNR.

- The performance of the unbiased quadratic power predictor is evaluated on measured broadband channels.

  - The median performance for power prediction of the measured taps for the unbiased quadratic predictor using smoothed regressors is similar to the performance on a tap described by the Jakes model.
  - The unbiased quadratic predictor has a better performance than the biased quadratic predictor.
  - The performance of the model based predictors, based on complex prediction of the taps with a Kalman filter and AR models, is similar to the one obtained using an unbiased quadratic predictor based on FIR prediction of the complex channel.
  - The performance is highly dependent on the channel-to-estimation error SNR. To achieve reliable channel predictions the channel estimation error has to be low.

7.A Covariances and Cross-covariances

As we usually work with the power of circular complex Gaussian random variables, we will review some basic results concerning the correlation for the power of such stochastic variables. For Gaussian random variables a, b, c and d, the expectation of the product is

$$\mathrm{E}\{abcd\} = \mathrm{E}\{ab\}\mathrm{E}\{cd\} + \mathrm{E}\{ac\}\mathrm{E}\{bd\} + \mathrm{E}\{ad\}\mathrm{E}\{bc\} + \mathrm{E}\{a\}\mathrm{E}\{b\}\mathrm{E}\{c\}\mathrm{E}\{d\}. \tag{7.98}$$

In particulare, for Gaussian random variables with zero mean we have

$$\mathrm{E}\{abcd\} = \mathrm{E}\{ab\}\mathrm{E}\{cd\} + \mathrm{E}\{ac\}\mathrm{E}\{bd\} + \mathrm{E}\{ad\}\mathrm{E}\{bc\}. \tag{7.99}$$

This is a relation that we will use often in the following.

For two zero mean, circular complex Gaussian random variables, $a \in \mathbf{CN}(0, \sigma_a^2)$ and $b \in \mathbf{CN}(0, \sigma_b^2)$ the corresponding powers are $p_a = |a|^2$ and $p_b = |b|^2$. The correlation/covariance (they are the same as a and b have zero means) between a and b is $r_{ab} = \mathrm{E}\{ab^*\}$. The covariance between p_a and p_b can be expressed using (7.99)

$$\begin{aligned} \mathrm{E}\{p_a p_b\} &= \mathrm{E}\{|a|^2|b|^2\} = \mathrm{E}\{aa^*bb^*\} \\ &= \mathrm{E}\{aa^*\}\mathrm{E}\{bb^*\} + \mathrm{E}\{ab\}\mathrm{E}\{a^*b^*\} + \mathrm{E}\{ab^*\}\mathrm{E}\{a^*b\}, \end{aligned} \tag{7.100}$$

but as a and b are circular we have that $\mathrm{E}\{ab\} = \mathrm{E}\{a^*b^*\} = 0$ and we obtain

$$\mathrm{E}\{p_a p_b\} = \mathrm{E}\{aa^*\}\mathrm{E}\{bb^*\} + \mathrm{E}\{ab^*\}\mathrm{E}\{a^*b\} = \sigma_a^2\sigma_b^2 + |r_{ab}|^2. \tag{7.101}$$

As a special case of (7.101) we have the second order moment of the power of a,

$$\mathrm{E}\{p_a^2\} = \mathrm{E}\{aa^*\}\mathrm{E}\{aa^*\} + \mathrm{E}\{aa^*\}\mathrm{E}\{a^*a\} = 2\sigma_a^4. \tag{7.102}$$

The mean of the power is readily obtained as

$$\bar{p}_a = \mathrm{E}\{p_a\} = \mathrm{E}\{aa^*\} = \sigma_a^2, \tag{7.103}$$

which is the variance of a. The variance of the power can then be obtained using (7.102) and (7.103) as

$$\mathrm{E}\{(p_a - \bar{p}_a)^2\} = \mathrm{E}\{p_a^2\} - \bar{p}_a^2 = 2\sigma_a^4 - \sigma_a^4 = \sigma_a^4, \tag{7.104}$$

which is the square of the variance of a. The covariance between p_a and p_b is obtained using (7.101) and (7.103)

$$\mathrm{E}\{(p_a - \bar{p}_a)(p_b - \bar{p}_b)\} = \mathrm{E}\{p_a p_b\} - \bar{p}_a\bar{p}_b = |r_{ab}|^2. \tag{7.105}$$

The covariance between the absolute squares of a and b is the absolute square of the covariance between a and b.

7.A.1 Power predicted with the unbiased quadratic predictor

The expected value for the predicted power is σ_h^2 when using the unbiased quadratic predictor (7.11). The variance of the predicted power is

$$\begin{aligned}\sigma_{\hat{p}}^2 &= \mathrm{E}\{(\hat{p}(t+L|t)-\sigma_h^2)^2\} = \mathrm{E}\{(|\hat{h}(t+L|t)|^2-\sigma_{\hat{h}}^2)^2\} \\ &= \mathrm{E}\{|\hat{h}(t+L|t)|^4\} - \sigma_{\hat{h}}^4 = \sigma_{\hat{h}}^4, \end{aligned} \tag{7.106}$$

which follows from (7.104). From (6.17) we then obtain

$$\sigma_{\hat{p}}^2 = \sigma_{\hat{h}}^4 = |\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}|^2 \tag{7.107}$$

and for the optimal predictor coefficients (7.15) $\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$, the variance (7.107) of the predicted power becomes

$$\sigma_{\hat{p}}^2 = |\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}|^2. \tag{7.108}$$

As both the true power $p(t)$ and the power prediction $\hat{p}(t)$ using the unbiased predictor in (7.11), has the variance for the complex time series, σ_h^2, as their averages, the corresponding cross-covariance is obtained as

$$\begin{aligned} r_{p\hat{p}} &= \mathrm{E}\{(p(t)-\sigma_h^2)(\hat{p}(t|t-L)-\sigma_h^2)\} \\ &= \mathrm{E}\{(|h(t)|^2-\sigma_h^2)(|\hat{h}(t|t-L)|^2-\sigma_{\hat{h}}^2)\} = |r_{h\hat{h}}|^2 = |\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}|^2, \end{aligned} \tag{7.109}$$

where we have used (7.105) in the second last equality and (6.14) for $r_{h\hat{h}}$ in the last. With the optimal coefficients $\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$, this is

$$r_{p\hat{p}} = |\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}|^2, \tag{7.110}$$

which is the same as the variance of the power prediction (7.108).

The cross-correlation between power and the prediction error using the unbiased predictor in (7.11), is

$$\begin{aligned} \rho_{p\varepsilon_p} &= \mathrm{E}\{p(t)\varepsilon_p(t)\} = \mathrm{E}\{p(t)(p(t)-\hat{p}(t|t-L))\} \\ &= \mathrm{E}\{|h(t)|^4\} - \mathrm{E}\{|h(t)|^2(|\hat{h}(t|t-L)|^2 + \sigma_h^2 - \sigma_{\hat{h}}^2)\} \\ &= 2\sigma_h^4 - \sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2 - \sigma_h^2(\sigma_h^2 - \sigma_{\hat{h}}^2) \\ &= \sigma_h^4 - |r_{h\hat{h}}|^2 = \sigma_h^4 - |\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}|^2 \end{aligned} \tag{7.111}$$

where we have used (7.101) and (7.102) in the second last row and (6.14) for $r_{h\hat{h}}$ in the last equality. With the optimal coefficients $\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$, this is

$$\rho_{p\varepsilon_p} = \sigma_h^4 - |\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}|^2. \tag{7.112}$$

The cross-correlation according to (7.112) will thus equal the minimal variance for the prediction error, according to (7.19).

The cross-correlation between predicted power and the prediction error is

$$\begin{aligned}\rho_{\hat{p}\varepsilon_p} &= \mathrm{E}\{\hat{p}(t|t-L)\varepsilon_p(t)\} \\ &= \mathrm{E}\{(p(t)-\varepsilon_p(t))\varepsilon_p(t)\} = \rho_{p\varepsilon_p} - \sigma^2_{\varepsilon_p} \ . \end{aligned} \tag{7.113}$$

Use the results from equation (7.111), for $\rho_{p\varepsilon_p}$, and (7.14), for the error variance, to obtain

$$\begin{aligned}\rho_{\hat{p}\varepsilon_p} &= \sigma_h^4 - |\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}|^2 - \left(\sigma_h^4 - 2|\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}|^2 + |\mathbf{r}^H_{h\varphi}\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}|^2\right) \\ &= |\boldsymbol{\theta}^H \mathbf{r}_{h\varphi}|^2 - |\mathbf{r}^H_{h\varphi}\mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}|^2. \end{aligned} \tag{7.114}$$

With $\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1}\mathbf{r}_{h\varphi}$ we obtain $\rho_{\hat{p}\varepsilon_p} = 0$. *Hence, using the optimal coefficients in the unbiased power prediction, there is no remaining correlation between the predicted power and the power prediction error at time t.*

7.A.2 Power predicted by linear regression in delayed power observations

When using the predictor (7.53), the cross-covariance $\mathbf{r}_{p\phi}$ between the true power and a regressor consisting of delayed observed power samples, is

$$\begin{aligned}[\mathbf{r}_{p\phi}]_i &= \mathrm{E}\{(p(t)-\bar{p})\boldsymbol{\phi}_i(t-L)\} \\ &= \mathrm{E}\{(|h(t)|^2 - \sigma_h^2)(|h(t-\tau_i)|^2 + |e(t-\tau_i)|^2 + x^*(t-\tau_i)e(t-\tau_i) \\ &\quad +h(t-\tau_i)e^*(t-\tau_i) - \sigma_h^2 - \sigma_e^2)\} \ . \end{aligned} \tag{7.115}$$

Here, $\tau_i = L + (1-i)\Delta t$. This expression can be expanded as

$$\begin{aligned}[\mathbf{r}_{p\phi}]_i &= \mathrm{E}\{|h(t)|^2|h(t-\tau_i)|^2\} - \mathrm{E}\{|h(t)|^2\}(\sigma_h^2+\sigma_e^2) \\ &\quad +\mathrm{E}\{|h(t)|^2|e(t-\tau_i)|^2\} - \sigma_h^2\mathrm{E}\{|e(t-\tau_i)|^2\} \\ &\quad +\sigma_h^2(\sigma_h^2+\sigma_e^2) - \sigma_h^2\mathrm{E}\{|h(t-\tau_i)|^2\} \\ &= (\sigma_h^4 + |r_h(\tau_i)|^2) - \sigma_h^2(\sigma_h^2+\sigma_e^2) + \sigma_h^2\sigma_e^2 \\ &\quad -\sigma_h^2\sigma_e^2 + \sigma_h^2(\sigma_h^2+\sigma_e^2) - \sigma_h^4 = |r_h(\tau_i)|^2, \end{aligned} \tag{7.116}$$

where we have made use of (7.101). The elements of the cross-covariance vector $\mathbf{r}_{p\phi}$ (7.61) thus consist of the absolute square of auto-correlations for $h(t)$. The elements of the cross-covariance vector between the complex signal $h(t)$, and the complex regressor $\boldsymbol{\varphi}(t-L)$, $\mathbf{r}_{h\varphi}$ (6.9) are $[\mathbf{r}_{h\varphi}]_i = r_h(\tau_i)$. The elements of $\mathbf{r}_{p\phi}$ can thus be expressed as

$$[\mathbf{r}_{p\phi}]_i = |[\mathbf{r}_{h\varphi}]_i|^2, \tag{7.117}$$

and the full vector is thus given by

$$\mathbf{r}_{p\phi} = \mathbf{r}_{h\varphi} \odot \mathbf{r}^*_{h\varphi}, \tag{7.118}$$

where $\odot$ denotes the Hadamard product (element wise multiplication).

Observed power

The covariance function for the observations of the power is

$$r_{p_y}(\tau) = \mathrm{E}\{(p_y(t) - \bar{p}_y)(p_y(t-\tau) - \bar{p}_y)\} = |r_y(\tau)|^2, \tag{7.119}$$

where we have used (7.105). When the signal $h(t)$ and the noise $e(t)$ are uncorrelated $r_y(\tau) = r_h(\tau) + r_e(\tau)$ and the covariance function becomes

$$r_{p_y}(\tau) = |r_y(\tau)|^2 = |r_h(\tau) + r_e(\tau)|^2. \tag{7.120}$$

7.B The MSE for the Biased and Unbiased Power Predictor

When using the biased power predictor (7.2) the MSE for the power prediction error in (7.3) becomes

$$\begin{aligned} \mathrm{E}\{|\varepsilon_{p_b}(t)|^2\} &= \mathrm{E}\{(|h(t)|^2 - |\hat{h}(t|t-L)|^2)^2\} \\ &= \mathrm{E}\{|h(t)|^4 - 2|h(t)|^2|\hat{h}(t|t-L)|^2 + |\hat{h}(t|t-L)|^4\}. \end{aligned} \tag{7.121}$$

Examine the three terms of (7.121) separately. As both $h(t)$ and $\hat{h}(t|t-L)$ are distributed as circular complex Gaussian with zero mean, the first and third terms are $2\sigma_h^4$ and $2\sigma_{\hat{h}}^4$ respectively. For the same reason we can use (7.101) to express the second term as

$$\mathrm{E}\{|h(t)|^2|\hat{h}(t|t-L)|^2\} = \sigma_h^2\sigma_{\hat{h}}^2 + |r_{h\hat{h}}|^2. \tag{7.122}$$

The MSE (7.121) for the biased power predictor is thus

$$\begin{aligned} \mathrm{E}\{|\varepsilon_{p_b}|^2\} &= 2\sigma_h^4 - 2(\sigma_h^2\sigma_{\hat{h}}^2 + |r_{h\hat{h}}|^2) + 2\sigma_{\hat{h}}^4 \\ &= 2\sigma_h^4 - 2\boldsymbol{\theta}^H\mathbf{r}_{h\varphi}\mathbf{r}_{h\varphi}^H\boldsymbol{\theta} - 2\sigma_h^2\boldsymbol{\theta}^H\mathbf{R}_\varphi\boldsymbol{\theta} + 2|\boldsymbol{\theta}^H\mathbf{R}_\varphi\boldsymbol{\theta}|^2, \end{aligned} \tag{7.123}$$

where we have used (6.17) for $\sigma_{\hat{h}}^2$ and (6.14) for $r_{h\hat{h}}$. This is the criterion that should be minimized to obtain the coefficient vector $\boldsymbol{\theta}$ that provides the best biased power predictor of the type (7.2).

The variance

The MSE for the unbiased power predictor (7.11) is obtained from (7.123) by subtracting the square of the bias,

$$\begin{aligned}\sigma_{\varepsilon_p}^2 = \mathrm{E}\{|\varepsilon_p|^2\} &= \mathrm{E}\{(|x|^2 - |\hat{h}|^2)^2\} - |\sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}|^2 \\ &= \sigma_h^4 - 2\boldsymbol{\theta}^H \mathbf{r}_{h\varphi} \mathbf{r}_{h\varphi}^H \boldsymbol{\theta} + |\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}|^2 \end{aligned} \quad (7.124)$$

This is also the variance for the prediction error for both the unbiased and biased power predictor.

7.B.1 Optimal coefficients for the biased predictor

Is it optimal to use the coefficient vector $\boldsymbol{\theta}_c$ in (6.12) that minimizes the prediction error (6.7) for the complex predictor, in the biased quadratic power predictor (7.2)? To investigate if this is the case, we take the partial derivative of the MSE, $\mathrm{E}\{|\varepsilon_{p_b}(t)|^2\}$ with respect to $\boldsymbol{\theta}$ in (7.123) and set it equal to zero.[3]

$$\frac{\partial \mathrm{E}\{|\varepsilon_{p_b}(t)|^2\}}{\partial \boldsymbol{\theta}} = -2\boldsymbol{\theta}^H \mathbf{r}_{h\varphi} \mathbf{r}_{h\varphi}^H - 2\sigma_h^2 \boldsymbol{\theta}^H \mathbf{R}_\varphi + 4\boldsymbol{\theta}^H \mathbf{R}_\varphi \boldsymbol{\theta}\boldsymbol{\theta}^H \mathbf{R}_\varphi = 0 \quad (7.125)$$

which results in the equation

$$\boldsymbol{\theta}^H (\mathbf{r}_{h\varphi} \mathbf{r}_{h\varphi}^H + \sigma_h^2 \mathbf{R}_\varphi - 2\mathbf{R}_\varphi \boldsymbol{\theta}\boldsymbol{\theta}^H \mathbf{R}_\varphi) = 0 \ . \quad (7.126)$$

The optimal coefficients should fulfill this equation. It is obvious that $\boldsymbol{\theta} = \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}$, which is the optimal solution with respect to the MSE criterion (6.8) for complex prediction, is *not a solution*. However, a scaled version of those coefficients

$$\boldsymbol{\theta}_{p_b} = \sqrt{\frac{\sigma_h^2 + \mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}}{2\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}}} \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi} = \sqrt{\frac{\sigma_h^2 - \sigma_{\varepsilon_c}^2/2}{\sigma_h^2 - \sigma_{\varepsilon_c}^2}} \boldsymbol{\theta}_c \quad (7.127)$$

will satisfy (7.126). Note that that the expression inside of the parenthesis in (7.126) will not be zero with this choice of $\boldsymbol{\theta}_{p_b}$. The vector of predictor coefficients $\boldsymbol{\theta}_{p_b}$, is however orthogonal to the expression inside the parenthesis and the scalar product is thus zero. Here, $\sigma_{\varepsilon_c}^2$ is the variance of the complex prediction error using the optimal coefficients, see (6.13). As $\sigma_{\varepsilon_c}^2 \leq \sigma_h^2$, the

[3] How to find the derivative with respect to a complex vector is described in e.g.Appendix B of [63].

scaling factor is larger or equal to one. This scaling thus amplifies the power prediction obtained with the optimal coefficients for the complex prediction,

$$\begin{aligned}\hat{p}(t+L|t) &= \frac{\sigma_h^2 - \sigma_{\varepsilon_c}^2/2}{\sigma_h^2 - \sigma_{\varepsilon_c}^2} \boldsymbol{\theta}_c^H \boldsymbol{\varphi}(t)\boldsymbol{\varphi}(t)^H \boldsymbol{\theta}_c \\ &= \frac{\sigma_h^2 - \sigma_{\varepsilon_c}^2/2}{\sigma_h^2 - \sigma_{\varepsilon_c}^2} |\hat{h}(t+L|t)|^2 \ . \end{aligned} \tag{7.128}$$

This solution is of minor interest as the predictor will still be biased, although the use of $\boldsymbol{\theta}_{p_b}$ reduces the bias by half, as compared to using $\boldsymbol{\theta}_c$.

If the optimal coefficients, $\boldsymbol{\theta}_{p_b}$, given by (7.127) for the biased power predictor are used in (7.123), then the MSE is obtained as

$$\begin{aligned}\mathrm{E}\{|\varepsilon_{p_b}(t)|^2\} &= \frac{3}{2}\sigma_h^4 - \sigma_h^2 \mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi} - \frac{1}{2}|\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi}|^2 \\ &= \sigma_{\varepsilon_c}^2 \left(2\sigma_h^2 - \sigma_{\varepsilon_c}^2/2\right), \end{aligned} \tag{7.129}$$

which is $\sigma_{\varepsilon_c}^4/2$ larger than the MSE (7.20) for the unbiased quadratic predictor (7.17).

7.C Derivation of the Distribution

7.C.1 The jpdf for the power

To obtain the jpdf for the true power and the power, as predicted by the unbiased quadratic predictor, we will use the bivariate χ^2 distribution.

Assume that we have two correlated **CN** distributed stochastic variables x and $\hat{h}$, both zero mean and with variances σ_h^2 and $\sigma_{\hat{h}}^2$ and correlation $r_{h\hat{h}}$. As both $|h|^2$ and $|\hat{h}|^2$ are $\chi^2(2)$ distributed, their joint probability density is a bivariate χ^2 distribution. The jpdf for the magnitudes, $|h|$ and $|\hat{h}|$, is a bivariate Rayleigh distribution [64], [7]

$$\begin{aligned} f_{|h||\hat{h}|}(|h|, |\hat{h}|) = \frac{4|h||\hat{h}|}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2} \exp\left(-\frac{\sigma_h^2\sigma_{\hat{h}}^2}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2}\left(\frac{|h|^2}{\sigma_h^2} + \frac{|\hat{h}|^2}{\sigma_{\hat{h}}^2}\right)\right) \\ \times \mathrm{I}_0\left(\frac{2|r_{h\hat{h}}||h||\hat{h}|}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2}\right). \end{aligned} \tag{7.130}$$

To obtain the corresponding bivariate χ^2 distribution we make a transformation to the powers, $|h|^2$ and $|\hat{h}|^2$. The jpdf for the transformed random variables is obtained using the Jacobian, as described in [62],

$$J(|h|, |\hat{h}|) = \begin{vmatrix} \frac{\partial |h|^2}{\partial |h|} & \frac{\partial |h|^2}{\partial |\hat{h}|} \\ \frac{\partial |\hat{h}|^2}{\partial |h|} & \frac{\partial |\hat{h}|^2}{\partial |\hat{h}|} \end{vmatrix} = \begin{vmatrix} 2|h| & 0 \\ 0 & 2|\hat{h}| \end{vmatrix} . \tag{7.131}$$

The determinant of the Jacobian is

$$|J(|h|, |\hat{h}|)| = 4|h||\hat{h}|. \tag{7.132}$$

The bivariate χ^2 distribution for $|h|^2$ and $|\hat{h}|^2$ is thus obtained as

$$\begin{aligned} f_{|h|^2|\hat{h}|^2}(|h|^2, |\hat{h}|^2) &= \frac{f_{|h||\hat{h}|}(\sqrt{|h|^2}, \sqrt{|\hat{h}|^2})}{|J(|h|, |\hat{h}|)|} \\ &= \frac{1}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2} \exp\left(-\frac{\sigma_h^2\sigma_{\hat{h}}^2}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2} \left(\frac{|h|^2}{\sigma_h^2} + \frac{|\hat{h}|^2}{\sigma_{\hat{h}}^2} \right) \right) \mathrm{I}_0\left(\frac{2|r_{h\hat{h}}||h||\hat{h}|}{\sigma_h^2\sigma_{\hat{h}}^2 - |r_{h\hat{h}}|^2} \right). \end{aligned} \tag{7.133}$$

7.C.2 The conditional pdf for the error

The cpdf for the prediction error and the the optimal quadratic power prediction is given by (7.47)

$$f_{\varepsilon_p}(\varepsilon_p|\hat{p}) = \frac{1}{\sigma_{\varepsilon_c}^2} \exp\left(-\frac{\varepsilon_p + 2\hat{p} - \sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2} \right) \mathrm{I}_0\left(\frac{2}{\sigma_{\varepsilon_c}^2} \sqrt{(\hat{p} - \sigma_{\varepsilon_c}^2)(\hat{p} + \varepsilon_p)} \right) \times \mathrm{U}(\hat{p} + \varepsilon_p)\mathrm{U}(\hat{p} - \sigma_{\varepsilon_c}^2). \tag{7.134}$$

In this section the mean, variance and the asymptotic behavior of the error, as a function of the predicted power, are derived.

The power prediction error, when using $\boldsymbol{\theta} = \boldsymbol{\theta}_p$ from (7.15) in the predictor (7.10), can by (6.7) and (7.17) be expressed in terms of the complex prediction $\hat{h}$ and its prediction error variance $\sigma_{\varepsilon_c}^2$ as (7.49)

$$\begin{aligned} \varepsilon_p &= p - \hat{p} = |h|^2 - |\hat{h}|^2 - \sigma_{\varepsilon_c}^2 = |\hat{h} + \varepsilon_c(t)|^2 - |\hat{h}|^2 - \sigma_{\varepsilon_c}^2 \\ &= |\varepsilon_c|^2 + \hat{h}\varepsilon_c^* + \hat{h}^*\varepsilon_c - \sigma_{\varepsilon_c}^2. \end{aligned} \tag{7.135}$$

Here, $\hat{h}$ and ε_c are uncorrelated, as the optimal prediction coefficients are used.

Mean

The mean of ε_p is

$$\mathrm{E}\{\varepsilon_p\} = \mathrm{E}\big\{|\varepsilon_c|^2 + \hat{h}\varepsilon_c^* + \hat{h}^*\varepsilon_c - \sigma^2_{\varepsilon_c}\big\} = \sigma^2_{\varepsilon_c} + 0 + 0 - \sigma^2_{\varepsilon_c} = 0, \qquad (7.136)$$

where we have used that the complex prediction and the complex prediction error are uncorrelated when the MSE-optimal prediction coefficients are used. *The power prediction error for the unbiased power predictor will thus have zero mean for all predicted powers.*

Variance

To obtain the variance for the power prediction error, conditioned on the predicted power, we shall study the mean and variance of a special case of the general (real-valued) quadratic form $Z = A|X|^2 + B|Y|^2 + CXY^* + C^*X^*Y$ with $A = 0$, $B = 1$ and $C = 1$,

$$Z = |Y|^2 + XY^* + X^*Y = Y_r^2 + Y_i^2 + 2X_rY_r + 2X_iY_i \ , \qquad (7.137)$$

where the sub-indices r and i denotes real and imaginary part respectively. Here, Y is a $\mathbf{CN}(0, \sigma_Y^2)$ distributed stochastic variable while X is below considered as a free deterministic variable which is uncorrelated with Y. Express X in polar coordinates, $X = r\cos\varphi + \jmath r\sin\varphi$, as we are interested in how Z depends on r. We thus have

$$Z = Y_r^2 + Y_i^2 + 2Y_r r\cos\varphi + 2Y_i r\sin\varphi \ . \qquad (7.138)$$

The mean of Z defined by (7.137) with respect to Y is

$$\bar{Z} = \mathrm{E}\{Z\} = \mathrm{E}\{|Y|^2 + XY^* + X^*Y\} = \sigma_Y^2 \ . \qquad (7.139)$$

The variance of Z is

$$\begin{aligned}
\sigma_Z^2 &= \mathrm{E}\{Z^2\} - \bar{Z}^2 = \mathrm{E}\big\{(Y_r^2 + Y_i^2 + 2Y_r r\cos\varphi + 2Y_i r\sin\varphi)^2\big\} - \bar{Z}^2 \\
&= \mathrm{E}\{Y_r^4\} + \mathrm{E}\{Y_i^4\} + 2\mathrm{E}\{Y_r^2\}\mathrm{E}\{Y_i^2\} \\
&\quad +4r^2\left(\mathrm{E}\{Y_r^2\}\cos^2\varphi + \mathrm{E}\{Y_i^2\}\sin^2\varphi\right) - \bar{Z}^2 \\
&= \frac{3}{4}\sigma_Y^4 + \frac{3}{4}\sigma_Y^4 + 2\frac{\sigma_Y^2}{2}\frac{\sigma_Y^2}{2} + 4r^2\frac{\sigma_Y^2}{2} - \sigma_Y^4 \\
&= \sigma_Y^2(\sigma_Y^2 + 2r^2) \ , \qquad (7.140)
\end{aligned}$$

where we have used that for zero mean real-valued Gaussian variables

$$\mathrm{E}\{Y_r^2\} = \mathrm{E}\{Y_i^2\} = \frac{\mathrm{E}\{Y^2\}}{2} = \frac{\sigma_Y^2}{2} \qquad (7.141)$$

$$\mathrm{E}\{Y_r^4\} = \mathrm{E}\{Y_i^4\} = 3\left(\frac{\mathrm{E}\{Y^2\}}{2}\right)^2 = \frac{3}{4}\sigma_Y^4 \ . \tag{7.142}$$

The power prediction error (7.135) using the unbiased quadratic predictor can be expressed in the form (7.137), with $Y = \varepsilon_c$, $X = \hat{h}$ and with the bias compensation $\sigma_{\varepsilon_c}^2$, subtracted. The variance for the power prediction error is given by (7.140) with $\sigma_Y^2 = \sigma_{\varepsilon_c}^2$ and $r = |\hat{h}|$,

$$\sigma_{\varepsilon_p}^2 = \sigma_{\varepsilon_c}^2(\sigma_{\varepsilon_c}^2 + 2|\hat{h}|^2). \tag{7.143}$$

The predicted power is $\hat{p} = |\hat{h}|^2 + \sigma_{\varepsilon_c}^2 = r^2 + \sigma_{\varepsilon_c}^2$, so the variance for the power prediction error conditioned on the predicted power is

$$\sigma_{\varepsilon_p}^2 = \sigma_{\varepsilon_c}^2(\sigma_{\varepsilon_c}^2 + 2(\hat{p} - \sigma_{\varepsilon_c}^2)) = \sigma_{\varepsilon_c}^2(2\hat{p} - \sigma_{\varepsilon_c}^2), \tag{7.144}$$

which is (7.48).

Asymptotic behavior

When $\hat{p} = \sigma_{\varepsilon_c}^2$ (which is the minimum), the pdf in equation (7.134) becomes

$$f_{\varepsilon_p \hat{p}}(\varepsilon_p|\hat{p}) = \frac{1}{\sigma_{\varepsilon_c}^2}\exp\left(-\frac{\varepsilon_p + \sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2}\right)\mathrm{U}(\varepsilon_p + \sigma_{\varepsilon_c}^2) \ , \tag{7.145}$$

which is the pdf for a $\chi^2(2)$ variable with mean $\sigma_{\varepsilon_c}^2$, that is, the pdf is shifted to the left by $\sigma_{\varepsilon_c}^2$. This will be the the distribution of the power prediction error when predicting into deep fading dips.

When $\hat{p} \to \infty$, we expect the distribution to become more Gaussian, as the two Gaussian terms $\hat{h}\varepsilon_c^* + \hat{h}^*\varepsilon_c$ in the sum in equation (7.135) will, for a specified and increasing $|\hat{h}|$, dominate over the first term. (Note that sums of Gaussian variables are also Gaussian.)

To obtain the asymptotic behavior for the pdf when $\hat{p} \to \infty$, the following two expressions are needed,

$$\sqrt{1 + \frac{1}{z}} = 1 + \frac{1}{2z} - \frac{1}{8z^2} + O\left(\frac{1}{z^3}\right) \ , \tag{7.146}$$

$$\mathrm{I}_0(z) = \frac{\mathrm{e}^z}{\sqrt{2\pi z}}\left[1 + O\left(\frac{1}{z}\right)\right] \tag{7.147}$$

where $z \to \infty$. If we use the approximation (7.146) on the argument in the modified Bessel function in (7.134) we obtain

$$
\begin{aligned}
&\frac{2}{\sigma_{\varepsilon_c}^2}\sqrt{(\hat{p}-\sigma_{\varepsilon_c}^2)(\hat{p}+\varepsilon_p)} = \frac{2\hat{p}}{\sigma_{\varepsilon_c}^2}\sqrt{1+\left(\frac{\varepsilon_p-\sigma_{\varepsilon_c}^2}{\hat{p}} - \frac{\varepsilon_p\sigma_{\varepsilon_c}^2}{\hat{p}^2}\right)} \\
&=\frac{2\hat{p}}{\sigma_{\varepsilon_c}^2}\left(1+\frac{1}{2}\left(\frac{\varepsilon_p-\sigma_{\varepsilon_c}^2}{\hat{p}} - \frac{\varepsilon_p\sigma_{\varepsilon_c}^2}{\hat{p}^2}\right) - \frac{1}{8}\left(\frac{\varepsilon_p-\sigma_{\varepsilon_c}^2}{\hat{p}} - \frac{\varepsilon_p\sigma_{\varepsilon_c}^2}{\hat{p}^2}\right)^2 + O\left(\frac{\varepsilon_p^3}{\hat{p}^3}\right)\right) \\
&=\frac{2\hat{p}}{\sigma_{\varepsilon_c}^2}\left(1+\frac{\varepsilon_p-\sigma_{\varepsilon_c}^2}{2\hat{p}} - \frac{\varepsilon_p\sigma_{\varepsilon_c}^2}{2\hat{p}^2} - \frac{(\varepsilon_p-\sigma_{\varepsilon_c}^2)^2}{8\hat{p}^2} + O\left(\frac{\varepsilon_p^3}{\hat{p}^3}\right)\right) \\
&=\frac{\varepsilon_p+2\hat{p}-\sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2} - \frac{(\varepsilon_p+\sigma_{\varepsilon_c}^2)^2}{4\hat{p}\sigma_{\varepsilon_c}^2} + O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)
\end{aligned}
\tag{7.148}
$$

This approximation of the argument is inserted into the modified Bessel function and then the approximation (7.147) is applied,

$$
\begin{aligned}
&\mathrm{I}_0\left(\frac{2}{\sigma_{\varepsilon_c}^2}\sqrt{(\hat{p}-\sigma_{\varepsilon_c}^2)(\hat{p}+\varepsilon_p)}\right) = \mathrm{I}_0\left(\frac{\varepsilon_p+2\hat{p}-\sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2} - \frac{(\varepsilon_p+\sigma_{\varepsilon_c}^2)^2}{4\hat{p}\sigma_{\varepsilon_c}^2} + O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)\right) \\
&= \frac{\exp\left(\frac{\varepsilon_p+2\hat{p}-\sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2} - \frac{(\varepsilon_p+\sigma_{\varepsilon_c}^2)^2}{4\hat{p}\sigma_{\varepsilon_c}^2} + O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)\right)}{\sqrt{2\pi\left(\frac{\varepsilon_p+2\hat{p}-\sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2} - \frac{(\varepsilon_p+\sigma_{\varepsilon_c}^2)^2}{4\hat{p}\sigma_{\varepsilon_c}^2} + O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)\right)}}\left(1+O\left(\hat{p}^{-1}\right)\right) \\
&= \sqrt{\frac{\sigma_{\varepsilon_c}^2}{4\pi\hat{p}}}\exp\left(\frac{\varepsilon_p+2\hat{p}-\sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2}\right)\exp\left(-\frac{\varepsilon_p^2}{4\sigma_{\varepsilon_c}^2\hat{p}}\right) \\
&\quad\times\frac{\exp\left(-\frac{2\varepsilon_p\sigma_{\varepsilon_c}^2+\sigma_{\varepsilon_c}^4}{4\hat{p}\sigma_{\varepsilon_c}^2} + O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)\right)}{\sqrt{\left(1+\frac{\varepsilon_p-\sigma_{\varepsilon_c}^2}{\hat{p}} + O\left(\frac{\varepsilon_p^2}{\hat{p}^2}\right)\right)}}\left(1+O\left(\hat{p}^{-1}\right)\right)
\end{aligned}
\tag{7.149}
$$

The last exponential and square root can be approximated further using

$$\frac{1}{\sqrt{1+\frac{1}{z}}} = 1 - \frac{1}{2z} + O\left(\frac{1}{z^2}\right), \tag{7.150}$$

$$\mathrm{e}^{1/z} = 1 + \frac{1}{z} + O\left(\frac{1}{z^2}\right) \tag{7.151}$$

where $z \to \infty$. The modified Bessel function in (7.134) can then be approximated as

$$\begin{aligned}
&\mathrm{I}_0\left(\frac{2}{\sigma_{\varepsilon_c}^2}\sqrt{(\hat{p}-\sigma_{\varepsilon_c}^2)(\hat{p}+\varepsilon_p)}\right) = \sqrt{\frac{\sigma_{\varepsilon_c}^2}{4\pi\hat{p}}}\exp\left(\frac{\varepsilon_p+2\hat{p}-\sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2}\right)\exp\left(-\frac{\varepsilon_p^2}{4\sigma_{\varepsilon_c}^2\hat{p}}\right)\\
&\quad\times\left(1-\frac{2\varepsilon_p+\sigma_{\varepsilon_c}^2}{4\hat{p}}+O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)\right)\left(1-\frac{\varepsilon_p-\sigma_{\varepsilon_c}^2}{4\hat{p}}+O\left(\frac{\varepsilon_p^2}{\hat{p}^2}\right)\right)\left(1+O(\hat{p}^{-1})\right)\\
&\quad=\sqrt{\frac{\sigma_{\varepsilon_c}^2}{4\pi\hat{p}}}\exp\left(\frac{\varepsilon_p+2\hat{p}-\sigma_{\varepsilon_c}^2}{\sigma_{\varepsilon_c}^2}\right)\exp\left(-\frac{\varepsilon_p^2}{4\sigma_{\varepsilon_c}^2\hat{p}}\right)\left(1+O\left(\frac{\varepsilon_p}{\hat{p}}\right)+O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)\right).
\end{aligned} \tag{7.152}$$

Insert this into equation (7.134) and the pdf for the power prediction error when $\hat{p}\to\infty$ is obtained as

$$f_{\varepsilon\hat{p}}(\varepsilon_p|\hat{p}) = \frac{1}{\sqrt{4\pi\sigma_{\varepsilon_c}^2\hat{p}}}\exp\left(-\frac{\varepsilon_p^2}{4\sigma_{\varepsilon_c}^2\hat{p}}\right)\left(1+O\left(\frac{\varepsilon_p}{\hat{p}}\right)+O\left(\frac{\varepsilon_p^3}{\hat{p}^2}\right)\right) \tag{7.153}$$

which is a Gaussian distribution. The variance is $2\sigma_{\varepsilon_c}^2\hat{p}$ which is in line with the results from (7.144) where the $\sigma_{\varepsilon_c}^4$ term is negligible compared to $2\sigma_{\varepsilon_c}^2\hat{p}$ when $\hat{p}\to\infty$.

7.C.3 Derivation of the pdf for the prediction error

To obtain an expression for the pdf of (7.49) we shall again take a look at the quadratic form

$$Z = |Y|^2 + XY^* + X^*Y\ , \tag{7.154}$$

where X and Y are uncorrelated zero mean **CN** distributed stochastic variables. The characteristic function of Z is given in [65] as

$$\Phi_z(\omega) = \frac{1}{(1+\jmath\omega\rho_+)(1-\jmath\omega\rho_-)} = \frac{1}{\rho_++\rho_-}\left(\frac{1}{\frac{1}{\rho_+}+\jmath\omega}+\frac{1}{\frac{1}{\rho_-}-\jmath\omega}\right), \tag{7.155}$$

where the damping coefficients ρ_+ and ρ_- are positive. The relationship between ρ_+, ρ_- and the variances will be given further on. The pdf is the inverse Fourier transform of the characteristic function,

$$f_z(z) = \mathcal{F}^{-1}\Phi_z(\omega) = \begin{cases} \frac{1}{\rho_++\rho_-}\mathrm{e}^{-z/\rho_+} & \text{if } z\ge 0\\ \frac{1}{\rho_++\rho_-}\mathrm{e}^{z/\rho_-} & \text{otherwise}\ , \end{cases} \tag{7.156}$$

where the part for positive and negative z relates to the first and second term in the parenthesis of equation (7.156), respectively. The stochastic variable Z has thus a skew Laplacian distribution. To obtain expressions for ρ_+ and ρ_- we look at the mean and variance of Z. These moments can be calculated from the pdf in equation (7.156) as

$$\bar{Z} = \int_{-\infty}^{\infty} z f_z(z) dz = \rho_+ - \rho_- \tag{7.157}$$

$$\sigma_Z^2 = \int_{-\infty}^{\infty} z^2 f_z(z) dz - \bar{Z}^2 = \rho_+^2 + \rho_-^2 \ . \tag{7.158}$$

The mean and variance can also be obtained from equation (7.154) as

$$\bar{Z} = \mathrm{E}\{|Y|^2 + XY^* + X^*Y\} = \sigma_Y^2 \tag{7.159}$$

$$\sigma_Z^2 = \mathrm{E}\{||Y|^2 + XY^* + X^*Y|^2\} - \bar{Z}^2 = \sigma_Y^4 + 2\sigma_X^2\sigma_Y^2, \tag{7.160}$$

since X and Y are uncorrelated. A system of equations is obtained from (7.157) to (7.160), where we can solve for ρ_+ and ρ_-,

$$\rho_+ = \frac{1}{2}\left(\sqrt{\sigma_Y^2(4\sigma_X^2 + \sigma_Y^2)} + \sigma_Y^2\right) \tag{7.161}$$

$$\rho_- = \frac{1}{2}\left(\sqrt{\sigma_Y^2(4\sigma_X^2 + \sigma_Y^2)} - \sigma_Y^2\right). \tag{7.162}$$

The pdf for Z can thus be expressed in terms of the variances for X and Y by insertion of (7.161) and (7.162) into (7.156).

The pdf for the prediction error

The prediction error of the power is obtained as (7.49) or (7.135),

$$\varepsilon_p = |\varepsilon_c|^2 + \hat{h}\varepsilon_c^* + \hat{h}^*\varepsilon_c - \sigma_{\varepsilon_c}^2. \tag{7.163}$$

This expression is similar to (7.154) but there is an extra constant term in (7.163) which has no counterpart in (7.154). This additive constant will cause a shift of the whole pdf in (7.156) by a factor of $\sigma_{\varepsilon_c}^2$. The pdf for the power prediction error using optimal linear prediction is thus obtained as

$$f_{\varepsilon_p}(\varepsilon_p) = \begin{cases} \frac{1}{\rho_+ + \rho_-} \mathrm{e}^{-(\varepsilon_p + \sigma_{\varepsilon_c}^2)/\rho_+} & \text{for } \varepsilon_p \geq -\sigma_{\varepsilon_c}^2 \\ \frac{1}{\rho_+ + \rho_-} \mathrm{e}^{(\varepsilon_p + \sigma_{\varepsilon_c}^2)/\rho_-} & \text{otherwise} \end{cases} \tag{7.164}$$

with $\rho_\pm$ obtained as in (7.161) and (7.162) with $\sigma_Y^2 = \sigma_{\varepsilon_c}^2$ and $\sigma_X^2 = \sigma_{\hat{h}}^2$,

$$\rho_\pm = \frac{1}{2}\left(\sqrt{\sigma_{\varepsilon_c}^2(4\sigma_{\hat{h}}^2 + \sigma_{\varepsilon_c}^2)} \pm \sigma_{\varepsilon_c}^2\right) = \frac{1}{2}\left(\sqrt{\sigma_{\varepsilon_c}^2(4\sigma_h^2 - 3\sigma_{\varepsilon_c}^2)} \pm \sigma_{\varepsilon_c}^2\right) \quad (7.165)$$

$$= \frac{1}{2}\left(\sqrt{(\sigma_h^2 - \mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi})(\sigma_h^2 + 3\mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi})} \pm (\sigma_h^2 - \mathbf{r}_{h\varphi}^H \mathbf{R}_\varphi^{-1} \mathbf{r}_{h\varphi})\right), \quad (7.166)$$

where we in the second equality made use of (6.18) for $\sigma_{\hat{h}}^2$ and in the last equality the expression (6.13) for $\sigma_{\varepsilon_c}^2$ was used.

Chapter 8

Application to Link Adaptation

8.1 Introduction

Adaptive modulation, or link adaptation, is a powerful technique for improving the spectral efficiency in wireless transmission over fading channels. The modulation parameters, such as signal constellation size, transmitted power level and data rate are here adjusted according to the channel conditions. The adaptation can also take requirements of different traffic classes and services such as required bit error rates, into account. In the case of *fast* link adaptation considered here, we strive to adapt to the small scale fading. The receiver estimates the received power and sends feedback information via a return channel to the transmitter, which adjusts its modulation parameters. Due to the unavoidable delays involved in power estimation, feedback transmission and modulation adjustment, the adaptation needs to be based on *predicted* estimates of the power of the fading communication channel.

The design of adaptive modulation systems is extensively studied in the literature (e.g. see [8], [4], [7]). However, a common assumption is perfect knowledge of the channel conditions at the transmitter as well as error free channel estimates at the receiver. In real systems, these assumptions are not valid. This leads to performance degradation such as decrease in the throughput and increase in the delay as well as failure in providing the required service such as the expected bit error rate. Therefore, a solution based on a more realistic assumption is of great importance and interest.

In this chapter, we intend to design an optimum adaptive modulation system for a given channel prediction error variance with a corresponding statistical model, which maximizes the spectral efficiency while satisfying a certain BER requirement. The proposed system utilizes the unbiased quadratic predictor (7.16) derived in Chapter 7 to predict the channel quality at the receiver. The statistical model developed in Chapter 7 reasonably well describes the properties of prediction errors and gives a good estimate of the mean, the variance and the probability density function of the errors. This statistical model is taken into account for design and performance analysis of the adaptive modulation schemes.

This chapter is organized as follows. Section 8.2 describes the system model and the notations which are used throughout this study. The channel prediction is explained in Section 8.3. The BER formula as a function of the predicted instantaneous SNR is evaluated in Section 8.4 and the optimal rate and power adaptation are derived in Section 8.5. Analytical results are presented in Section 8.6 and finally, some conclusions are drawn in Section 8.7.

8.2 System Model

In the adaptive modulation, M-QAM modulation with different constellation sizes are provided at the transmitter. For each transmission of a block of data, the modulation scheme and transmit power are assumed to be selected, to maximize the spectral efficiency under the instantaneous BER and the average power constraints, based on the instantaneous predicted SNR. The channel is modeled as a flat Rayleigh fading channel. At the receiver, demodulation is performed using estimates of the channel, which are here assumed to be error free.

The discrete-time model of the system is depicted in Figure 8.1. All the signals are sampled at the symbol rate, where the index n represents the signal sample at time nt_s where t_s is the symbol period. The channel is estimated at the symbol rate. The complex valued channel gain $h(n)$ has a uniformly distributed phase within $[0, 2\pi)$ and Rayleigh distributed amplitude with pdf given by

$$f_{|h|}(|h(n)|) = \frac{|h(n)|}{\sigma_h^2/2} \exp\left(-\frac{|h(n)|^2}{\sigma_h^2}\right) \tag{8.1}$$

where σ_h^2 is the average power of the fading process. The noise $v(n)$ is a sample of the complex AWGN with zero mean and time-invariant variance σ_v^2. Channel estimation at the receiver renders $y(n)$, a noisy estimate of $h(n)$,

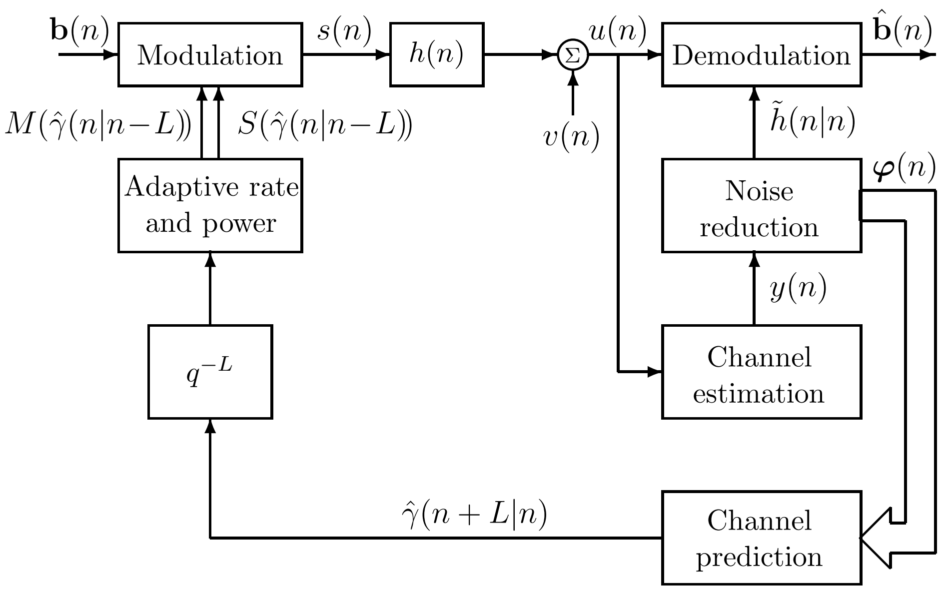

Figure 8.1: Discrete model of the system. M is the constellation size and S is the transmit power.

which is used by an FIR predictor at the receiver to predict the instantaneous received SNR denoted by $\hat{\gamma}(n|n-L)$. The noise reduction step is optional but helps to increase the prediction performance. An error free feed-back channel is assumed. Furthermore, the feed-back transmission delay is presumed to be taken up by the predictor, since the prediction is performed beyond the corresponding delay.

Based on $\hat{\gamma}(n|n-L)$, the modulation scheme with constellation size $M(\hat{\gamma}(n|n-L))$ (out of N constellations available at the transmitter) which transmits $k(\hat{\gamma}(n|n-L)) = \log_2 M(\hat{\gamma}(n|n-L))$ bits per symbol, and the transmit power $S(\hat{\gamma}(n|n-L))$ are selected. A number of blocks of $k(\hat{\gamma}(n|n-L))$ data bits denoted by $\mathbf{b}(n)$, are Gray encoded and mapped to $s(n)$ which is a symbol in the signal constellation and is transmitted over the flat Rayleigh fading channel. The received sample, $u(n)$, is used to estimate the channel gain $\tilde{h}(n|n)$ which in turn is used to demodulate $u(n)$ to detect the transmitted bits denoted by $\hat{\mathbf{b}}(n)$. Since the estimation error is believed to have a minor effect on the performance compared to the prediction error, perfect channel estimation is assumed in the demodulation, i.e. $\tilde{h}(n|n) = h(n)$. The SNR prediction $\hat{\gamma}(n|n-L)$ is however assumed to be based on noisy channel estimates and will have a prediction error, with variance σ_ε^2.

In this study, the following notations similar to those of [8] are used. Let $\bar{S}$ denote the average transmit signal power and

$$\bar{\gamma} = \frac{\bar{S}}{\sigma_v^2}\sigma_h^2 \tag{8.2}$$

denote the average received SNR. For a constant transmit power $\bar{S}$, the instantaneous received SNR is

$$\gamma(n) = \bar{\gamma}\frac{p(n)}{\sigma_h^2} \tag{8.3}$$

where $p(n) = |h(n)|^2$ is the instantaneous channel power gain. Accordingly, the instantaneous predicted received SNR is

$$\hat{\gamma}(n|n-L) = \bar{\gamma}\frac{\hat{p}(n|n-L)}{\sigma_h^2} \tag{8.4}$$

where $\hat{p}(n|n-L)$ is the predicted instantaneous channel power gain.

For a varying transmit power $S(\hat{\gamma}(n|n-L))$, the instantaneous received SNR is given by

$$\gamma(n,S) = \frac{S(\hat{\gamma}(n|n-L))}{\sigma_v^2}p(n) = \gamma(n)\frac{S(\hat{\gamma}(n|n-L))}{\bar{S}}. \tag{8.5}$$

Also, the rate region boundaries denoted by $\{\hat{\gamma}_i\}_{i=0}^{N-1}$ and $\hat{\gamma}(n) = \infty$ are defined as the ranges of $\hat{\gamma}(n)$ values over which the different constellations are used by the transmitter.

When the predicted instantaneous SNR belongs to a given rate region, i.e. $\hat{\gamma}(n) \in [\hat{\gamma}_i, \hat{\gamma}_{i+1})$, the corresponding constellation of size $M(\hat{\gamma}(n)) = M_i$ with $k(\hat{\gamma}(n)) = k_i$ bits per symbol is transmitted. Finally, there is no transmission if $\hat{\gamma}(n) < \hat{\gamma}_0$ meaning that $\hat{\gamma}_0$ is the cutoff SNR.

8.3 Prediction and Distributions for SNR

The channel gain in a Rayleigh fading mobile radio channel takes values from a complex time series which can be modeled as a correlated complex Gaussian stochastic process. The absolute square, i.e. the power, of the time series is predicted based on linear regression of the observations of the complex time series. The unbiased quadratic predictor that is optimal in the Mean Square Error (MSE) sense is derived in Chapter 7 where it is shown that the same prediction coefficients that are optimal also for the linear prediction of the complex tap, are optimal for the quadratic unbiased predictor.

The complex channel gain is observed in noise as

$$y(n) = h(n) + e(n) \tag{8.6}$$

where $e(n)$ is assumed to be a white, zero mean, complex Gaussian random variable which is independent from $h(n)$.

SNR prediction

From the past and present observations, the power of the signal at time $n+L$, i.e. $p(n+L) = |h(n+L)|^2$, is predicted by an unbiased FIR predictor. In the vector formulation, the unbiased predicted power based on the past noisy observations is given by (7.10) derived in Section 7.2.2

$$\hat{p}(n+L|n) = \boldsymbol{\theta}^H \boldsymbol{\varphi}(n)\boldsymbol{\varphi}^H(n)\boldsymbol{\theta} + \sigma_h^2 - \boldsymbol{\theta}^H \mathbf{R}_{\varphi}\boldsymbol{\theta} \tag{8.7}$$

where the regressor $\boldsymbol{\varphi}(n)$ is obtained as in (6.5) or using smoothed observations as in (6.31). The predicted SNR is obtained by insertion of (8.7) into (8.4).

Distribution functions

In order to find the optimum adaptive modulation system which maximizes the spectral efficiency under a certain BER constraint the pdf of the instantaneous SNR is required. Since $h(n)$ is a complex Gaussian random variable, the true power $p(n) = |h(n)|^2$ is $\chi^2(2)$ distributed, whereas the predicted power has a shifted $\chi^2(2)$ distributed due to the bias compensation. In Appendix 8.A it is shown that the pdf of γ conditioned on $\hat{\gamma}$ is given by (8.41)

$$f_\gamma(\gamma|\hat{\gamma}) = \frac{\mathrm{U}(\gamma)\mathrm{U}(\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c})}{\bar{\gamma}\nu_{h\varepsilon_c}} \exp(-\frac{\gamma + \hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c}}{\bar{\gamma}\nu_{h\varepsilon_c}})\mathrm{I}_0\left(\frac{2}{\bar{\gamma}\nu_{h\varepsilon_c}}\sqrt{\gamma(\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c})}\right), \tag{8.8}$$

where $\nu_{h\varepsilon_c}$ is the ratio between the variance for the complex channel prediction error and the variance for the channel, defined as

$$\nu_{h\varepsilon_c} = \frac{\sigma_{\varepsilon_c}^2}{\sigma_h^2}. \tag{8.9}$$

The pdf of $\hat{\gamma}$ is given by (8.39) in Appendix 8.A as

$$f_{\hat{\gamma}}(\hat{\gamma}) = \frac{\mathrm{U}(\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c})}{\bar{\gamma}(1 - \nu_{h\varepsilon_c})} \exp\left(-\frac{\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c}}{\bar{\gamma}(1 - \nu_{h\varepsilon_c})}\right). \tag{8.10}$$

In (8.8) and (8.10) the index n is dropped expressions since $h(n)$ and $\hat{h}(n)$ are both assumed to be stationary random processes. The power prediction has a lower limit given by the variance of the complex channel prediction error. The predicted SNR will thus have a lower limit $\hat{\gamma} \geq \bar{\gamma}\nu_{h\varepsilon_c}$.

8.4 M-QAM BER Performance

The transmitter adjusts the constellation size and the transmit power based on the predicted SNR $\hat{\gamma}$. Evaluation of the optimal power and constellation size (or rate) which maximize the spectral efficiency and satisfy the BER requirement, requires an analytical expression for the BER formula as a function of $\hat{\gamma}$. Assuming a square M-QAM constellation with Gray encoded bits, constellation size $M(\hat{\gamma})$, and transmit power $S(\hat{\gamma})$, the instantaneous BER as a function of γ and $\hat{\gamma}$ on an AWGN channel, is approximated by [65]

$$\mathrm{BER}(\gamma, \hat{\gamma}) \approx \frac{2}{\log_2 M(\hat{\gamma})}\left(1 - \frac{1}{\sqrt{M(\hat{\gamma})}}\right)\mathrm{erfc}\left(\sqrt{1.5\frac{\gamma\frac{S(\hat{\gamma})}{\bar{S}}}{M(\hat{\gamma})-1}}\right) \quad (8.11)$$

which is tight for high SNRs. In [8], it is shown that (8.11) can be further approximated as

$$\mathrm{BER}(\gamma, \hat{\gamma}) \approx 0.2 \exp\left(\frac{-1.6\gamma}{M(\hat{\gamma})-1}\frac{S(\hat{\gamma})}{\bar{S}}\right) \quad (8.12)$$

which is tight within 1 dB for $M(\hat{\gamma}) \geq 4$ and BER$\leq 10^{-3}$. By averaging (8.12) over the whole range of the instantaneous true SNR γ, the instantaneous BER as a function of the instantaneous predicted SNR is obtained as

$$\mathrm{BER}(\hat{\gamma}) = \int_0^{\infty} \mathrm{BER}(\gamma, \hat{\gamma}) f_{\gamma}(\gamma|\hat{\gamma})d\gamma \quad (8.13)$$

$$\approx \frac{a(\hat{\gamma})}{b(\hat{\gamma})}\exp\left(\frac{c(\hat{\gamma})^2}{4b(\hat{\gamma})}\right), \quad \hat{\gamma} \geq \bar{\gamma}\nu_{h\varepsilon_c} \quad (8.14)$$

where

$$a(\hat{\gamma}) = \frac{0.2}{\bar{\gamma}\nu_{h\varepsilon_c}}\exp\left(-\frac{\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c}}{\bar{\gamma}\nu_{h\varepsilon_c}}\right), \quad (8.15)$$

$$b(\hat{\gamma}) = \frac{1}{\bar{\gamma}\nu_{h\varepsilon_c}} + \frac{1.6}{M(\hat{\gamma})-1}\frac{S(\hat{\gamma})}{\bar{S}}, \quad (8.16)$$

and

$$c(\hat{\gamma}) = \frac{2}{\bar{\gamma}\nu_{h\varepsilon_c}}\sqrt{\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c}}. \tag{8.17}$$

In Figure 8.2, the instantaneous BER is illustrated for the constant transmit power $S(\hat{\gamma}) = \bar{S}$ and different constellation sizes where both (8.11) and (8.12) are used in (8.14). It is shown that the results based on the approximation given by (8.12) is tight enough.

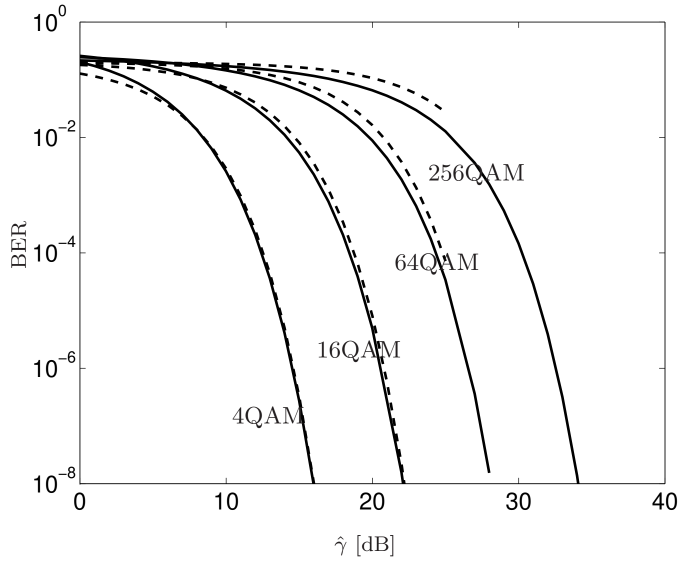

Figure 8.2: BER versus instantaneous predicted SNR of M-QAM schemes for $\bar{\gamma} = 30$ dB and $\sigma^2_{\varepsilon_p} = 0.001$ (that is a prediction NMSE of 0.0005). The solid lines and dashed lines correspond to (8.14) based on (8.11) and (8.12), respectively.

8.5 Optimal Rate and Power Adaptation

The spectral efficiency of a modulation scheme is given by the average data rate R per unit bandwidth B, (R/B). When a modulation with constellation size $M(\hat{\gamma})$ is chosen, the instantaneous data rate is $k(\hat{\gamma})/t_s$ (bps), where k is the number of data bits per symbol. Assuming the Nyquist data pulses $(B = 1/t_s)$, the spectral efficiency is given by

$$\eta_B = \frac{R}{B} = \sum_{i=0}^{N-1} k_i \int_{\hat{\gamma}_i}^{\hat{\gamma}_{i+1}} f_{\hat{\gamma}}(\hat{\gamma})d\hat{\gamma} \quad \text{bps/Hz} \tag{8.18}$$

We would like to maximize the spectral efficiency subject to the average transmit power constraint

$$\int_0^\infty S(\hat{\gamma}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \leq \bar{S} \tag{8.19}$$

and the instantaneous BER constraint given by

$$\mathrm{BER}(\hat{\gamma}) = P_b. \tag{8.20}$$

Under a BER constraint, $S(\hat{\gamma})$ can be evaluated in terms of $\hat{\gamma}$ and $M(\hat{\gamma})$ as follows: Let us define

$$x = \frac{1}{\bar{\gamma}\nu_{h\varepsilon_c} b(\hat{\gamma})}, \quad 0 \leq x \leq 1 \tag{8.21}$$

and

$$y = \frac{\hat{\gamma}}{\bar{\gamma}\nu_{h\varepsilon_c}}, \quad y \geq 1. \tag{8.22}$$

By using (8.32) and (8.22) in (8.14), equation (8.20) can be expressed as

$$0.2xe^{(1-y)(1-x)} = P_b. \tag{8.23}$$

Taking the natural logarithm of (8.23) and then using the Taylor approximation $\ln x \approx x - 1$ about $x = 1$, we obtain

$$x \approx 1 + \frac{1}{y} \ln\left(\frac{P_b}{0.2}\right) \tag{8.24}$$

and consequently

$$S(\hat{\gamma}) \approx g(\hat{\gamma}, M(\hat{\gamma})) \tag{8.25}$$

$$= \frac{-\left(M(\hat{\gamma}) - 1\right)}{1.6} \frac{\bar{S} \ln\left(\frac{P_b}{0.2}\right)}{\hat{\gamma} + \bar{\gamma}\nu_{h\varepsilon_c} \ln\left(\frac{P_b}{0.2}\right)} \mathrm{U}\left(\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c}\right). \tag{8.26}$$

As explained in Section 8.2, a signal constellation and consequently a transmission rate is assigned to each predicted SNR region boundary. Therefore, for a given number of signal constellations, the optimum predicted SNR region boundary for each constellation corresponds to the optimum transmission rate, accordingly. Moreover, the transmit power can be obtained from the instantaneous predicted SNR based on (8.25). Therefore, the optimization problem can be simplified to finding only the optimal region boundaries. The Lagrangian function is given by

$$J(\hat{\gamma}_0, \hat{\gamma}_1, \cdots, \hat{\gamma}_{N-1}) = \sum_{i=0}^{N-1} k_i \int_{\hat{\gamma}_i}^{\hat{\gamma}_{i+1}} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} + \lambda \Big(\sum_{i=0}^{N-1} \int_{\hat{\gamma}_i}^{\hat{\gamma}_{i+1}} g(\hat{\gamma}, M_i) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} - \bar{S} \Big) \quad (8.27)$$

where the optimal region boundaries are found by solving

$$\frac{\partial J}{\partial \hat{\gamma}_i} = 0, \quad 0 \leq i \leq N-1 \quad (8.28)$$

which results in

$$g(\hat{\gamma}_i, M_{i-1}) - g(\hat{\gamma}_i, M_i) = -\frac{k_{i-1} - k_i}{\lambda} \quad (8.29)$$

where $k_{-1} = 0$ and $M_{-1} = 1$ are assumed. Taking into account (8.25) and (8.29), the optimum region boundaries are obtained by

$$\hat{\gamma}_i = \ln \left(\frac{P_b}{0.2} \right) \left(\frac{\bar{S}}{1.6} \frac{M_{i-1} - M_i}{k_{i-1} - k_i} \lambda - \bar{\gamma} \nu_{h\varepsilon_c} \right) \quad (8.30)$$

where the Lagrangian multiplier $\lambda \neq 0$ is numerically evaluated based on the power constraint given as

$$\bar{S} \geq \sum_{i=0}^{N-1} \int_{\hat{\gamma}_i}^{\hat{\gamma}_{i+1}} g(\hat{\gamma}, M_i) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} = \rho \exp \left(\frac{\nu_{h\varepsilon_c}}{1 - \nu_{h\varepsilon_c}} \left(1 + \ln \left(\frac{P_b}{0.2} \right) \right) \right) \sum_{i=0}^{N-2} (M_i - 1) \left(\mathrm{Ei}(\rho \lambda P_i) - \mathrm{Ei}(\rho \lambda P_{i+1}) \right) + (M_{N-1} - 1) \mathrm{Ei}(\rho \lambda P_{N-1}) \quad (8.31)$$

where

$$\rho = \ln \left(\frac{P_b}{0.2} \right) \frac{\bar{S}}{1.6} \frac{1}{\bar{\gamma}(1 - \nu_{h\varepsilon_c})}, \quad P_i = \frac{M_{i-1} - M_i}{k_{i-1} - k_i} \quad (8.32)$$

and $\mathrm{Ei}(\cdot)$ is the Exponential Integral given by $\int (e^x/x) dx = \mathrm{Ei}(x)$. When the optimal region boundaries are determined, the maximum spectral efficiency is evaluated through (8.10) and (8.18) as

$$\eta_B = \sum_{i=0}^{N-1} k_i \Big[\exp \Big(-\frac{\max(\bar{\gamma}\nu_{h\varepsilon_c}, \hat{\gamma}_i) - \bar{\gamma}\nu_{h\varepsilon_c}}{\bar{\gamma}(1 - \nu_{h\varepsilon_c})} \Big) - \exp \Big(-\frac{\max(\bar{\gamma}\nu_{h\varepsilon_c}, \hat{\gamma}_{i+1}) - \bar{\gamma}\nu_{h\varepsilon_c}}{\bar{\gamma}(1 - \nu_{h\varepsilon_c})} \Big) \Big]. \quad (8.33)$$

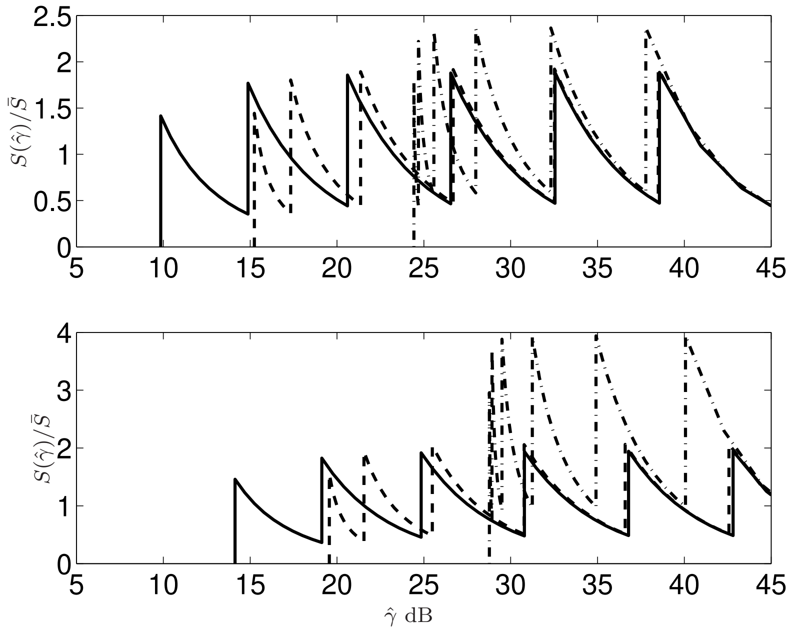

Figure 8.3: Optimum normalized transmit power versus instantaneous predicted SNR of M-QAM schemes for $\bar{\gamma} = 30$ dB. The upper and lower plots correspond to $P_b = 10^{-3}$ and 10^{-7}, and the solid, dashed and dashed-dotted lines correspond to $\sigma^2_{\varepsilon_p} = 0.001$, 0.01, and 0.1, respectively.

8.6 Results

We assume six different M-QAM signal constellations corresponding to 4-QAM, 16-QAM, 64-QAM, 256-QAM, 1024-QAM and 4096-QAM, are available at the transmitter. Also, a flat Rayleigh fading channel with $\sigma_h^2 = 1$ is presumed. The optimal region boundaries for the cases where the required instantaneous BERs are 10^{-3} and 10^{-7} and the average received SNR $\bar{\gamma} = 30$ dB are evaluated and shown in Figure 8.3, for the prediction error variances $\sigma^2_{\varepsilon_p} = 0.001$, 0.01 and 0.1 respectively (that is the prediction NMSE are $0,0005$, $0,005$ and $0,05$). It is observed that the optimal region boundaries are increased when the prediction error increases, as expected. The maximum spectral efficiency for $P_b = 10^{-3}$ and 10^{-7} and $\sigma^2_{\varepsilon_p} = 0.001$, 0.01 and 0.1 are illustrated in Figure 8.4. The curves show that the gain in the spectral efficiency using good predictors is considerable as compared to the poor predictors. The reduction in spectral efficiency with an increasing

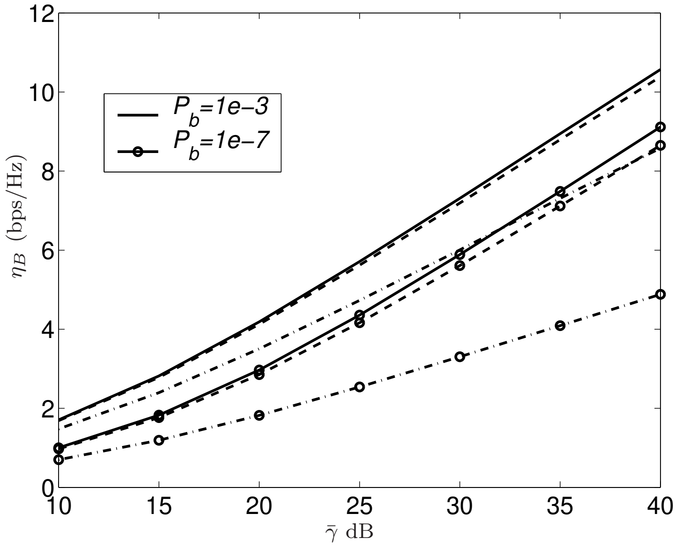

Figure 8.4: M-QAM Spectral efficiency versus average received SNR for $P_b = 10^{-3}$ *and* 10^{-7}. *The solid, dashed and dashed-dotted lines correspond to* $\sigma^2_{\varepsilon_p} = 0.001$, 0.01 *and* 0.1, *respectively.*

prediction inaccuracy becomes larger when the required BER is decreased. Compare $P_b = 10^{-7}$ to $P_b = 10^{-3}$ in Figure 8.4.

The lowest predicted SNR at which the system transmits increases with increasing prediction NMSE. This is due to that the relative uncertainty is higher in the fading dips as seen in Section 7.4.3. A large portion of the time this single user system will not transmit. It is thus ideal to combine the optimized adaptive modulation with a multiuser diversity scheme, scheduling the radio resources between the users.

8.7 Conclusion

The optimum design of an adaptive modulation scheme based on M-QAM modulation has been investigated. The transmitter adjusts the transmission rate and power based on the predicted SNR to maximize the spectral efficiency while satisfying the instantaneous BER and average transmit power constraints. The instantaneous BER as a function of predicted SNR, average SNR and predicted error variance has been evaluated. Optimum solutions for the adaptive rate and transmit power were derived.

The analytical results show that:

- The predicted SNR boundaries for a given constellation size are increased when the prediction error increases.
- The spectral efficiency decreases as the variance for the prediction error increases and (or) the required BER decreases.
- The loss in spectral efficiency can be significantly reduced by using a good predictor.
- The difference in spectral efficiency for different prediction error variances decreases when the BER requirement is made less tight.

A combination with a multiuser diversity scheme would be very beneficial especially when the prediction NMSE is quite high as then the link adaptation avoids transmission at low predicted SNR and leaves the radio resource open.

8.A Pdfs for the Predicted SNR

The pdf for the predicted SNR and the conditional pdf for the true SNR conditioned on the predicted SNR are needed to solve the optimization problem for the rate regions and the power. In Section 7.4.1 the corresponding distributions for the true and predicted power are derive for a flat Rayleigh fading channel. These expressions will be used to obtain the pdfs for the SNR.

8.A.1 The pdf for the predicted SNR

For a flat Rayleigh fading channel the distribution of the predicted power, using the unbiased quadratic predictor (7.17), is given by (7.39) as

$$f_{\hat{p}}(\hat{p}(t+L|t)) = \frac{\mathrm{U}\big(\hat{p}(t+L|t) - \sigma^2_{\varepsilon_c}\big)}{\sigma_h^2 - \sigma^2_{\varepsilon_c}} \exp\left(-\frac{\hat{p}(t+L|t) - \sigma^2_{\varepsilon_c}}{\sigma_h^2 - \sigma^2_{\varepsilon_c}}\right), \tag{8.34}$$

where $\sigma^2_{\varepsilon_c}$ is the variance for the complex channel prediction error. The predicted SNR is obtained from the predicted power as (8.4)

$$\hat{\gamma}(t|t-L) = \bar{\gamma}\frac{\hat{p}(t|t-L)}{\sigma_h^2}. \tag{8.35}$$

For notational convenience the time dependence is dropped in the following. The Jacobian for the transformation from predicted power to predicted SNR is

$$\frac{d\hat{\gamma}}{d\hat{p}} = \frac{\bar{\gamma}}{\sigma_h^2} \tag{8.36}$$

The pdf for the predicted SNR is then obtains from (8.34) as

$$f_{\hat{\gamma}}(\hat{\gamma}) = \frac{f_{\hat{p}}(\hat{\gamma}\sigma_h^2/\bar{\gamma})}{\left|\frac{d\hat{\gamma}}{d\hat{p}}\right|} = \frac{\sigma_h^2}{\bar{\gamma}} \frac{\mathrm{U}\big((\sigma_h^2\hat{\gamma}/\bar{\gamma}) - \sigma^2_{\varepsilon_c}\big)}{\sigma_h^2 - \sigma^2_{\varepsilon_c}} \exp\left(-\frac{(\sigma_h^2\hat{\gamma}/\bar{\gamma}) - \sigma^2_{\varepsilon_c}}{\sigma_h^2 - \sigma^2_{\varepsilon_c}}\right). \tag{8.37}$$

Let the ratio between the variance for the complex channel prediction error and the variance of the channel be denoted

$$\nu_{h\varepsilon_c} = \frac{\sigma^2_{\varepsilon_c}}{\sigma_h^2}. \tag{8.38}$$

The pdf for the predicted power can thus be obtained as

$$f_{\hat{\gamma}}(\hat{\gamma}) = \frac{\mathrm{U}(\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c})}{\bar{\gamma}(1-\nu_{h\varepsilon_c})} \exp\left(-\frac{\hat{\gamma} - \bar{\gamma}\nu_{h\varepsilon_c}}{\bar{\gamma}(1-\nu_{h\varepsilon_c})}\right). \tag{8.39}$$

8.A.2 The conditional pdf for true and predicted SNR

The condition pdf for the true power conditioned on the predicted power is obtain in (7.46) as

$$f_{p\hat{p}}(p|\hat{p}) = \frac{\mathrm{U}(p)\mathrm{U}(\hat{p}-\sigma^2_{\varepsilon_c})}{\sigma^2_{\varepsilon_c}} \exp\left(-\frac{p+\hat{p}-\sigma^2_{\varepsilon_c}}{\sigma^2_{\varepsilon_c}}\right) \mathrm{I}_0\left(\frac{2}{\sigma^2_{\varepsilon_c}}\sqrt{p(\hat{p}-\sigma^2_{\varepsilon_c})}\right). \quad (8.40)$$

Here $\hat{p}$ is not seen as a stochastic variable and can thus be substituted by $\hat{\gamma}\sigma_h^2/\bar{\gamma}$ without changing the cfpd. The normalization of (8.40) is only performed with respect to $p = \gamma\sigma_h^2/\bar{\gamma}$ as

$$\begin{aligned} f_{\gamma\hat{\gamma}}(\gamma|\hat{\gamma}) &= \frac{1}{\left|\frac{d\gamma}{dp}\right|} f_{p\hat{p}}(\gamma\sigma_h^2/\bar{\gamma}|\hat{\gamma}\sigma_h^2/\bar{\gamma}) \\ &= \frac{\mathrm{U}\left(\frac{\gamma\sigma_h^2}{\bar{\gamma}}\right)\mathrm{U}\left(\frac{\hat{\gamma}\sigma_h^2}{\bar{\gamma}}-\sigma^2_{\varepsilon_c}\right)}{\bar{\gamma}\sigma^2_{\varepsilon_c}/\sigma_h^2} \exp\left(-\frac{\frac{\gamma\sigma_h^2}{\bar{\gamma}}+\frac{\hat{\gamma}\sigma_h^2}{\bar{\gamma}}-\sigma^2_{\varepsilon_c}}{\sigma^2_{\varepsilon_c}}\right) \mathrm{I}_0\left(\frac{2}{\sigma^2_{\varepsilon_c}}\sqrt{\frac{\gamma\sigma_h^2}{\bar{\gamma}}\left(\frac{\hat{\gamma}\sigma_h^2}{\bar{\gamma}}-\sigma^2_{\varepsilon_c}\right)}\right) \\ &= \frac{\mathrm{U}(\gamma)\mathrm{U}(\hat{\gamma}-\bar{\gamma}\nu_{h\varepsilon_c})}{\bar{\gamma}\nu_{h\varepsilon_c}} \exp\left(-\frac{\gamma+\hat{\gamma}-\bar{\gamma}\nu_{h\varepsilon_c}}{\bar{\gamma}\nu_{h\varepsilon_c}}\right) \mathrm{I}_0\left(\frac{2}{\bar{\gamma}\nu_{h\varepsilon_c}}\sqrt{\gamma(\hat{\gamma}-\bar{\gamma}\nu_{h\varepsilon_c})}\right). \end{aligned} \quad (8.41)$$

Chapter 9

Concluding Remarks

9.1 Channel Models

The understanding of the mobile radio channel and the process behind the small scale fading is crucial for success in designing predictors. The linearized model where the tap in a mobile radio channel is described as a sum of weighted complex sinusoids seem to suggest that the tap should be easy to predict. The changing angles to the reflectors and scatterers that contribute to the multipath result in an adjusted model where the frequencies drift. How fast the frequencies drift depend on the distances and angles towards the secondary sources. This is an important effect if there are strong scatterers close to the transceiver and it will limit the performance of a predictor based on the sinusoidal model.

The assumption of perfect distant reflectors can be questioned. When a building or a group of objects act as a cluster of scatterers the corresponding contribution to the channel can be described as a narrowband filtered Gaussian noise. There is a mapping from the geometrical distribution of a cluster to the angular distribution of power around the receiver. The correlation function of the contribution to the channel of a distant cluster is a damped (not necessarily exponentially damped) complex sinusoid.

A Cauchy like distributed cluster of scatters result in a sharp peak in the angular distribution. The contribution of such a cluster to the channel can be modeled as an AR1-process. When a number of clusters contribute to the channel a suitable model is given by an ARMA-process. The properties of this ARMA-process, and the noise it is observed in, limits the performance of the channel predictor. A linear predictor is suitable for prediction of such processes.

9.2 Channel Estimation Error

In a block based LS-estimation the channel is estimated using received symbols at baseband sample rate. The channel estimation error is both due to measurement noise and the channel variation during the estimation interval. The error that is caused by the time variation can be decomposed in two uncorrelated errors: The excess error due to the weighted averaging of the channel and the bias error due to the curvature of the channel. For a channel described by the Jakes model these two errors are uncorrelated. The variance of the error due to the time variation within the estimation interval can be parameterized using the time-frequency product (the length of the estimation interval measured in wavelengths) and the number of samples used for identification. When the time-frequency product is small, the error due to the time variation will also be small. The estimation error due to the measurement noise is proportional to the inverse of the number of of samples used for identification. Instead of increasing the estimation interval to average away the measurement noise, better channel estimates can be obtained by filtering of the channel estimates obtained from short estimation intervals.

The variance for the estimation error forms a noise floor in the power delay profile or the average Doppler spectrum of the estimated channels.

9.3 Noise Reduction on Estimated Channels

Exploiting simple models of the dynamics of the channel noise reduction using Wiener-smoothers can be performed on the estimated channels. When evaluated on simulated Jakes channels an IIR smoother based on a Chebyshev type one low-pass filter of degree 4 achieves a smoothing MSE similar to that of an optimal FIR-smoother with 128 coefficients designed for the Jakes channel. The interpolation problem is thus not that sensitive to the choice of model dynamics as long as it has the appropriate low-pass structure.

The IIR smoothers are numerically sensitive to errors on the coefficients when the model filters are very narrowband (that is at oversampling ratios above 50). The FIR-smoothers are not numerically sensitive but need on the order of 100 coefficients. The FIR smoother should have a memory that cover at least one wavelength of traveled distance.

When sub-sampled predictors are used the noise reduction acts as the necessary anti-aliasing filter.

9.4 Channel Tap Prediction

The deployment of noise reduction on the observed taps of a channel reduces the number of predictor coefficients that has to be estimated from the observations. For a tap described by a Jakes model the increase in performance when going from 8 to 32 coefficients in the predictor is insignificant if noise reduction has been applied in a first step.

The performance of a direct FIR-predictor depends on the delay spacing. A robust choice of the delay spacing is about a tenth of a wavelength for a predictor with 8 to 32 coefficients.

The estimated AR and ARMA-models indicate that there is more signal power at high frequencies than there actually is. This is due to the estimation errors. This is compensated for in the design of the Kalman predictor, where a noise model with pronounced power at higher frequencies is used. This reduces the amplification at these frequencies.

Evaluation on measured channels

On measured taps the direct sub-sampled FIR-predictor that is designed to minimize the prediction error for a given range has roughly the same performance as a Kalman predictor using a sub-sampled AR model for the dynamics.

The performance is highly dependent on the variance of the channel estimation error but it also depends on the dynamics of the individual taps. Taps with a low Doppler spread are generally more predictable than taps with high Doppler spread.

9.5 Power Prediction

The power prediction is obtained as the squared magnitude of the complex prediction of a tap, together with a bias compensation. A prediction of the total power is obtained as the sum of the predicted powers for the taps.

Evaluation on measured channels

The NMSE of the power prediction of the individual taps is on average quite close to that of a tap described by the Jakes model. For prediction ranges above 0.3 wavelengths the observed performance is actually better than on the Jakes model. Just as for the complex tap predictions the performance depends both on the variance of the channel estimation error and the dynamics of the tap.

For power control the total power of the communication link is adjusted. When prediction is not applied the last available observation of the power is signaled back to the base station. The power control is then performed on outdated information. For the total power the proposed unbiased quadratic predictor using 8 coefficients and smoothed regressors clearly outperforms the last observed power sample as predictor. Up to a range of 0.2 wavelengths the proposed predictor is ten times better than the last observed power sample as predictor. Up to a quarter of a wavelength the prediction NMSE is below 0.01, as compared to the last sample predictor that for such a long range is close to the average power as predictor in performance. At a carrier frequency of 1880 MHz, 0.2 wavelengths is 32 mm of traveled distance, which at a vehicle speed of 50 km/h corresponds to 2 ms.

9.6 Link Adaptation

As a final example of an application of the power predictor and the theory developed in Chapter 7, a system with link adaptation is described. The proposed system utilizes the unbiased quadratic predictor and maximizes the spectral efficiency while satisfying a certain BER requirement.

The modulation is chosen depending on the the predicted SNR. The region boundaries for the predicted SNR where the modulation is changed is highly dependent on the power prediction NMSE. A combination with a multiuser diversity scheme would be very beneficial especially when the prediction NMSE is quite high as then the link adaptation avoids transmission at low predicted SNR and leaves the radio resource open.

The taps of the measured channel can give an indication of how predictable a flat fading channel could be. The power prediction NMSE is assumed to be 0.0005, 0.005 and 0.05 in Figure 8.3 and Figure 8.4. Figure 7.17 summarizes the power prediction NMSE for the measured taps for different ranges and signal-to-estimation error ratios. An NMSE of 0.0005 is not achieved at any range and 0.005 is achieved only for the taps with a signal-to-estimation-error ratio in the range 32 to 38 dB for a range of less than 0.15 wavelengths. A power prediction NMSE of 0.05 can be achieved for ranges up to a quarter of a wavelength when the signal-to-estimation-error ratio is above 20 dB. For a range of 0.1 wavelengths it is achieved already for 8 dB. If the signaling in the feedback path to the base station is fast enough, it is feasible to deploy this type of link adaptation.

9.7 The Predictor Design

- Estimate the channel with the uttermost care. All estimation errors result in degraded performance of the predictor.
- Use Wiener-smoothers on the estimated taps to reduce the estimation error in a separate step.
- Estimate a sub-sampled AR-model for the dynamics of the taps. Alternatively, estimate the coefficients of a direct FIR-predictor, using smoothed regressors. A delay spacing of about a tenth of a wavelength is appropriate.
- Keep the number of estimated coefficients low, as estimation errors reduce the prediction performance.
- Design a Kalman predictor based on the estimated AR-model and a noise model that emphasizes high frequencies to compensate for the estimation errors in the AR-model.
- Power prediction is obtained as the squared magnitude of the complex prediction with an added bias compensation.

9.8 Topics for Future Research

The problem is to design robust prediction algorithms that can cope with both sudden and gradual changes of the impulse response and the underlying statistics. Algorithms for both the time and the frequency domain are needed to cover the large span of applications. The use of channel prediction in antenna array systems must be given special attention. There is a large potential for adaptive MIMO systems facilitated by channel prediction.

The algorithms must have low complexity as the time constraints are rigid and the computing power in a mobile terminal is limited. The time used for calculation must be compensated for by a correspondingly longer prediction range.

Next we give some directions and recommendations for future research:

- Study the performance and complexity of predictions in the time domain mapped to the frequency domain. The high temporal resolution in the channel impulse response causes each tap to contain contributions from a lower number of paths, and thus makes them more predictable. Predictions in the time domain can then be mapped to the frequency domain where it can be used for link adaptation in an OFDM like system.

- Develop descriptions of how the statistics of the channel changes over time. How long time intervals the models for the dynamics of a channel are valid is of great importance to the prediction performance. It also governs how long the estimation intervals can be.

- Develop robust adaptive channel prediction algorithms. When the dynamics of the channel changes gradually, adaptive methods can be used to obtain continuous updates of the models or prediction coefficients [27]. Block based methods can be seen as a special case of the adaptive methods, as the models are frequently re-estimated.

- Merge channel prediction with antenna array systems. The combination of these techniques can give adaptive beam forming that do not suffer from outdated channel estimates [57].

- Derive performance limits for different types of environments. The predictability of the channel will differ depending on the type of environment.

- Closer integration with the applications. Not all applications that use channel predictions have the same performance criteria. The prediction algorithms should be tailored to the specific needs of the application.

Appendix A

Visualized Channels

To further understand the mobile radio channel, visualization of a number of different aspects of the measured impulse responses is of great importance. In the following the estimated channels from three different measurements will be described and the following properties will be illustrated.

Received power
The total received power changes as a function of time. When the power decreases rapidly, this is called a fade. The received power depends on the channel as

$$p(t) = \sum_k |h_k(t)|^2. \tag{A.1}$$

Frequency response
The instantaneous value of the time-varying channel has a frequency response. That is a discretized version of the frequency responses in equation (2.23). The fading is thus frequency selective. A fade in the received power corresponds to a situation with many frequencies or taps fading simultaneously.

$$H(\omega, t) = \mathcal{F}_k\{h_k(t)\}. \tag{A.2}$$

Power delay profile
The power delay profile is the power as a function of the delay k of the impulse response, averaged over the time t for a time-varying response,

$$P(k) = \frac{1}{N} \sum_t |h_k(t)|^2, \tag{A.3}$$

where N is the number of observations. The power delay profile shows how the received power is distributed over the delays. A peak in the PDP corresponds to a path with the corresponding delay k.

Doppler spectra

The Fourier transform of the complex taps described in equation (2.25) gives the Doppler shift for the different contributing rays.

$$H_k(\Omega) = \mathcal{F}_t\{h_k(t)\} \tag{A.4}$$

The Doppler spectrum thus shows the frequency distribution of the oscillations of the individual taps. It is plotted as a power spectrum, $|H_k(\Omega)|^2$. To increase the resolution, a MASC spectral estimator [54] is used instead of the Fourier transform. The Doppler spectrum gives information about the direction of the incident waves. A peak in the Doppler spectrum close to f_D corresponds to a wave coming from the direction of motion. In the same way a peak close to $-f_D$ is a wave coming from behind the vehicle. Waves from the side correspond to peaks close to 0 Hz.

The key features of the three exemplified channels are listed in Table A.1.

	A	B	C
Velocity [m/s]	25	14	10
Delay spread [μs]	0.4	0.3	0.8
Doppler spread [Hz]	56	43	35
3 dB coherence bandwidth [Hz]	4.9	2.3	0.6

Table A.1: Table of the key features for the displayed channels.

Channel A

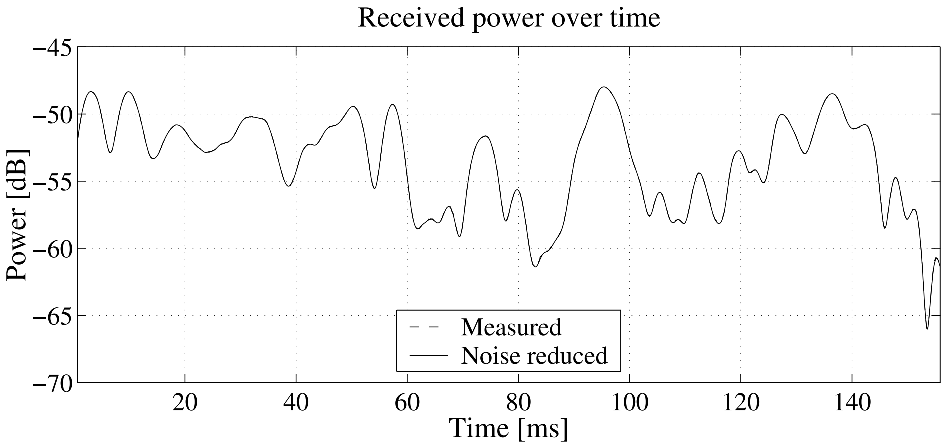

Figure A.1: The fading pattern of channel A changes a lot over time.

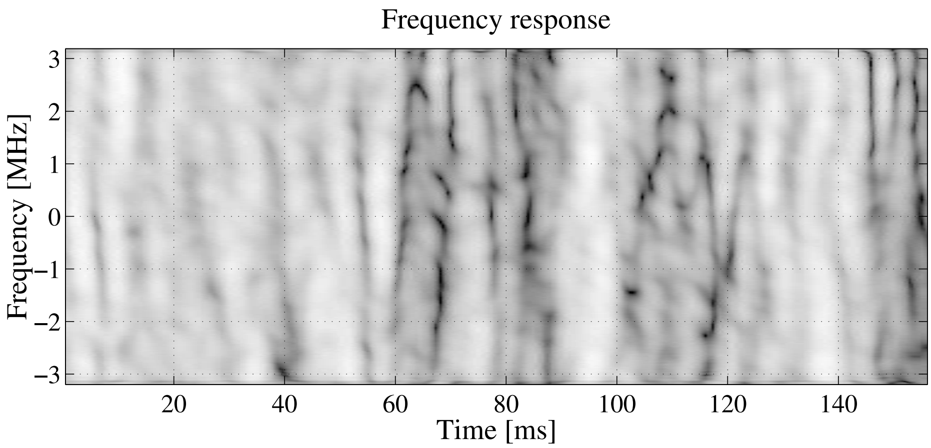

Figure A.2: Frequency response $|H(\omega, t)|$ of the estimated channel A. Dark coloring denotes fades. The dynamic range (black-white) is 40 dB. All frequencies fade simultaneously due to that mainly one tap contributes to the power.

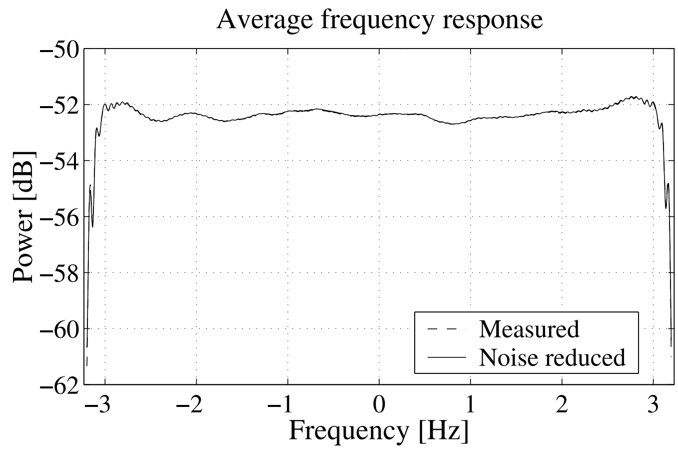

Figure A.3: The average frequency response of channel A is almost flat.

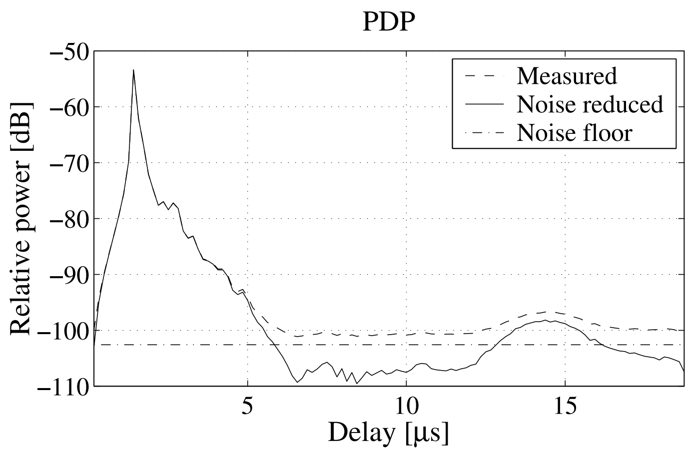

Figure A.4: The PDP of channel A shows that the power comes mainly from one tap. The noise floor is the power of the estimation error.

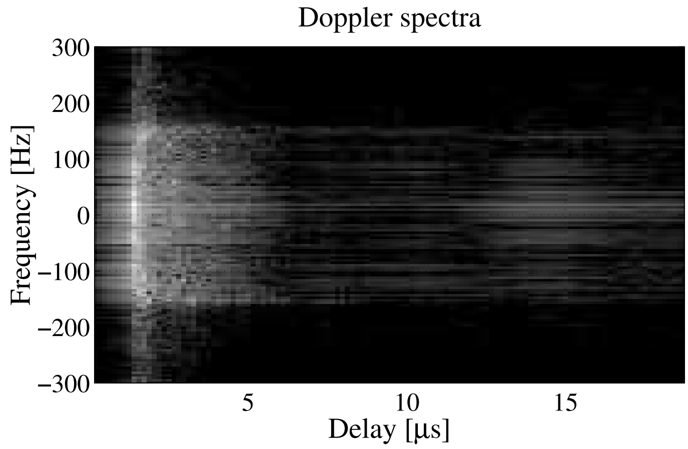

Figure A.5: The Doppler spectra for all the taps of channel A.

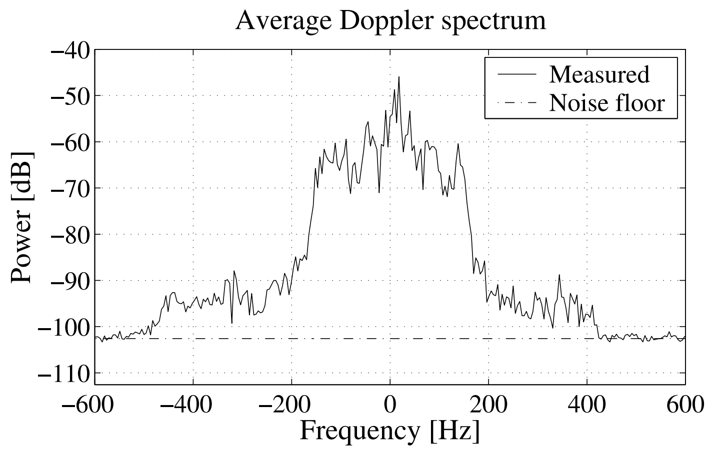

Figure A.6: Average Doppler spectrum of channel A. The strongest wavefronts come from the side as there are peaks close to 0 Hz. Note the small but significant power at frequencies above the Doppler frequency.

Channel B

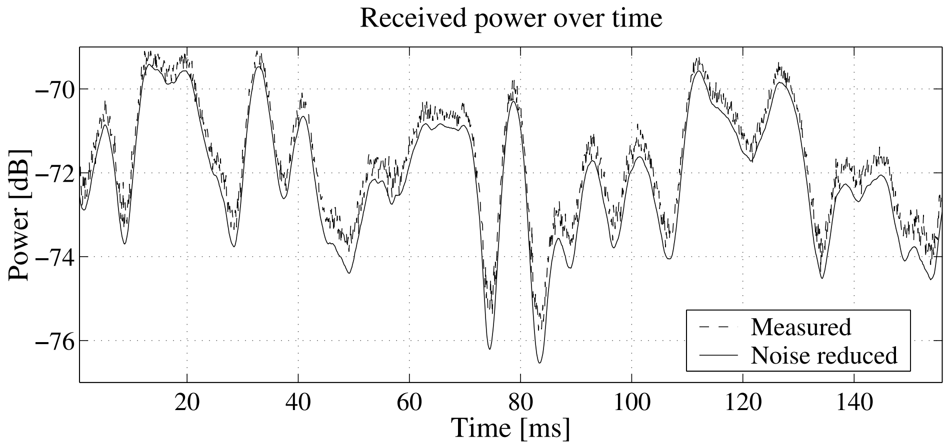

Figure A.7: The the estimated power of channel B. The power decrease when performing noise reduction.

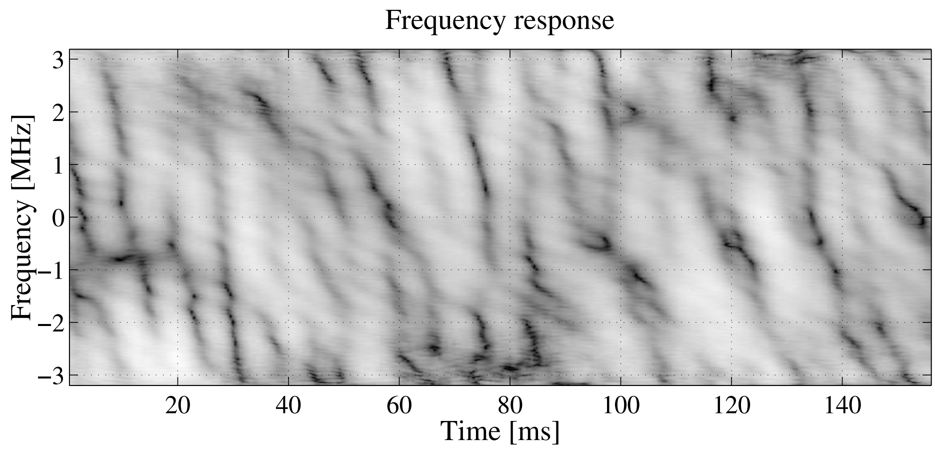

Figure A.8: Frequency response $|H(\omega, t)|$ of the estimated channel B. Dark coloring denotes fades and the dynamic range (black-white) is 40 dB.

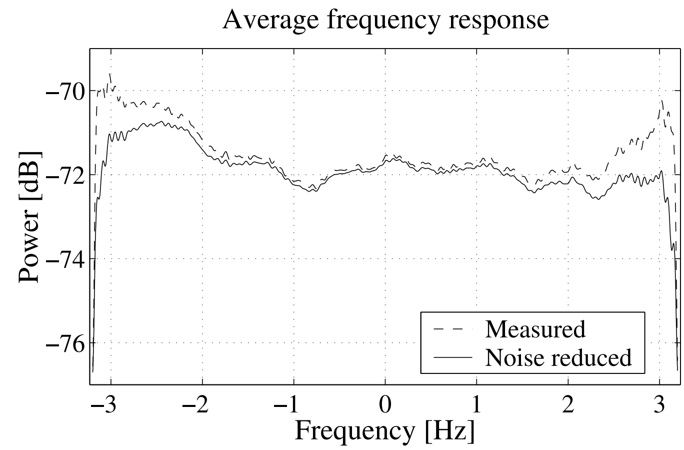

Figure A.9: The average frequency response of channel B is not flat. Noise reduction modifies the estimated frequency response especially at frequencies close to the band-limit.

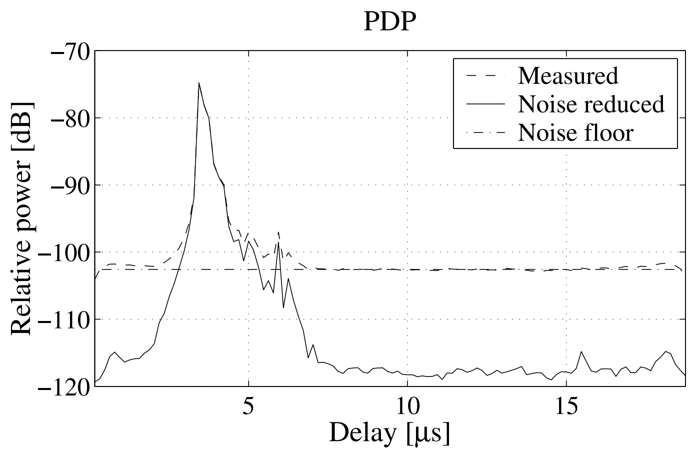

Figure A.10: The PDP of channel B shows that the power mainly comes from a few taps. Noise reduction is very efficient on this channel.

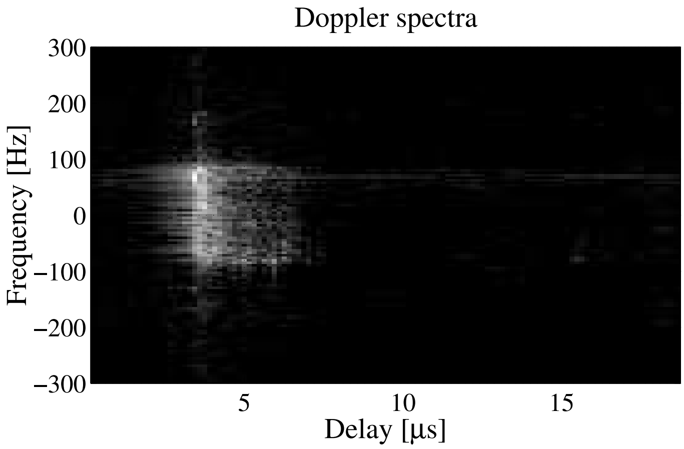

Figure A.11: The Doppler spectra for all the taps of channel B.

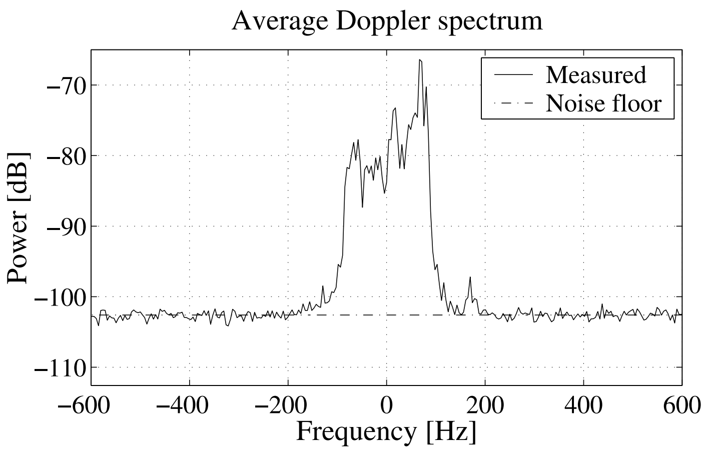

Figure A.12: Average Doppler spectrum of channel B. The strongest wavefronts come from the front.

Channel C

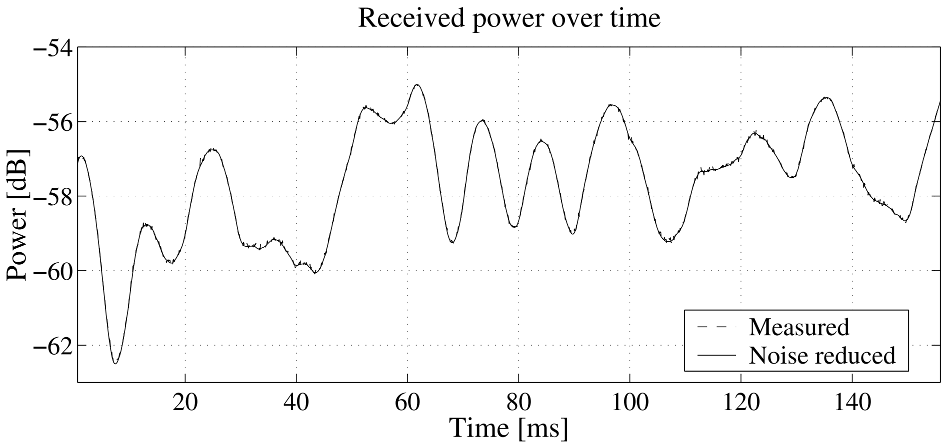

Figure A.13: The fades are not so deep in channel C due to the large delay spread.

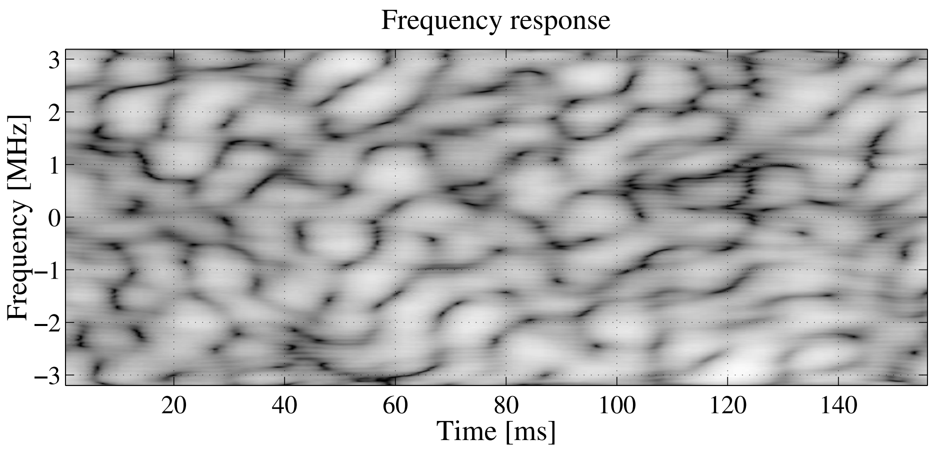

Figure A.14: Frequency response $|H(\omega, t)|$ of the estimated channel. Dark coloring denotes fades and the dynamic range (black-white) is 40 dB. The coherence bandwidth is small as the frequencies fade almost independent of each other.

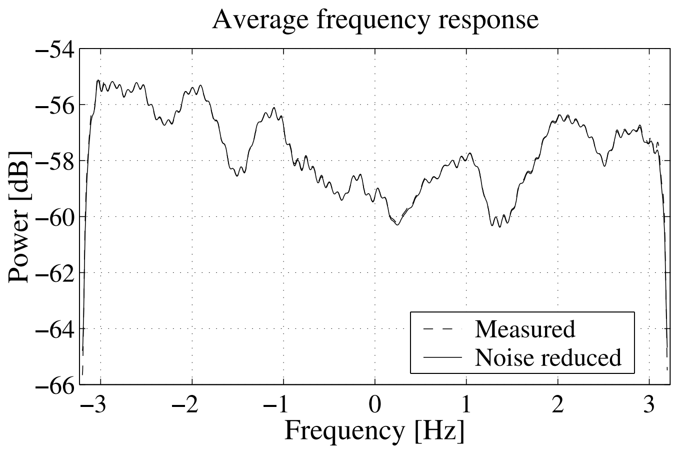

Figure A.15: The average frequency response of channel C.

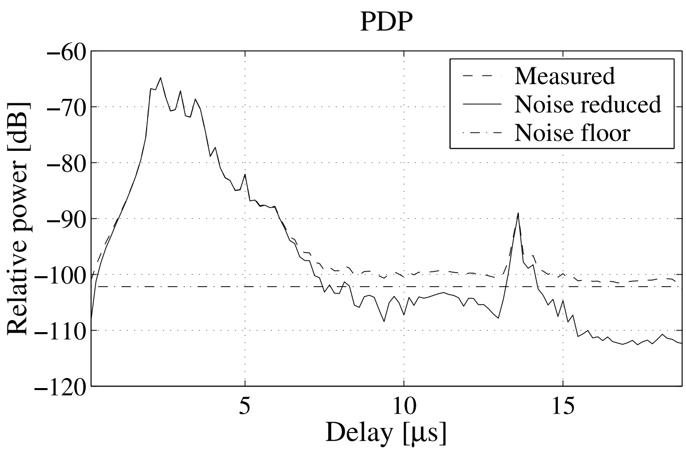

Figure A.16: The PDP of channel C shows that many taps contribute to the total power.

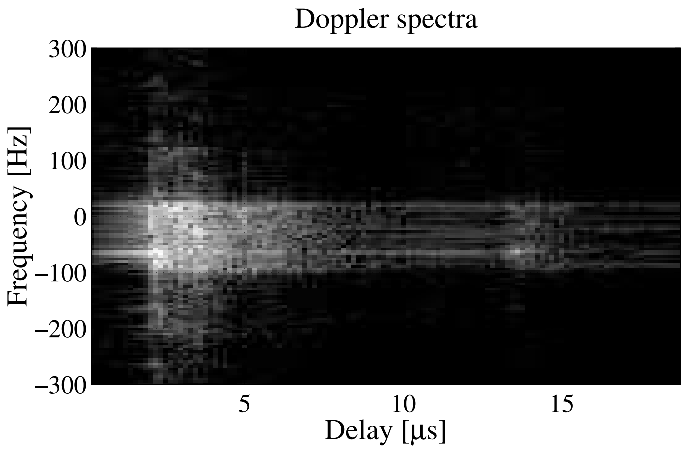

Figure A.17: The Doppler spectra for all the taps of channel C.

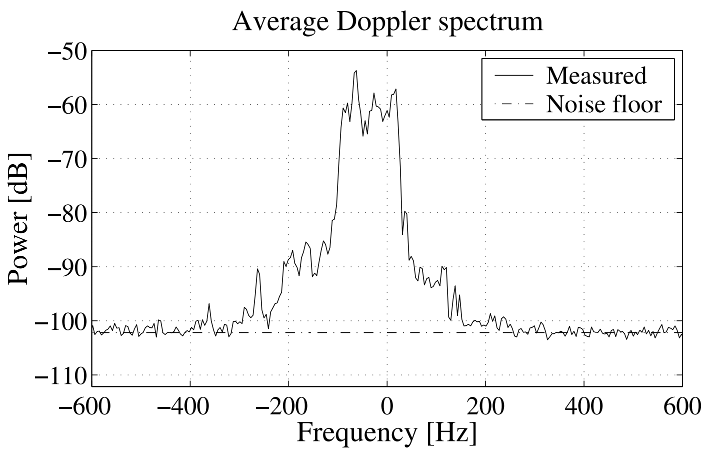

Figure A.18: Average Doppler spectrum of channel C.

Bibliography

[1] N.C. Ericsson, Adaptive modulation and scheduling of IP-trafic over fading channels, In *Proc. IEEE* 50^{th} *Vehic. Tech. Conf. Fall*, Amsterdam, Sep. 1999.

[2] N.C. Ericsson, S. Falahati, A. Ahlén and A. Svensson, Hybrid type-II ARQ/AMS supported by channel predictive scheduling in a multi-user scenario, In *IEEE Vehicular Tech. Conf. (VTC2000-Fall)*, pages 1804–1811, Boston, MA, Sep. 2000.

[3] J.L. Speyer K. Shoarinejad and G.J. Pottie, A distributed scheme for integrated predictive dynamic channel and power allocation in cellular radio networks, In *GLOBECOM '01*, volume 6, pages 3623 – 3627, San Antonio, TX, USA, Nov. 2001.

[4] A. Goldsmith and S. Chua, Variable-rate variable-power MQAM for fading channels, *IEEE Transactions on Communications*, pages 1218–1230, Oct. 1997.

[5] K.N. Lau and M.D. Macleod, Variable rate adaptive trellis coded qam for high bandwidth efficiency applications in rayleigh fading channels, In *IEEE Vehicular Technology Conference VTC98*, pages 348–352, May 1998.

[6] D.L. Goeckel, Adaptive coding for fading channels using outdated channel estimates, *IEEE Trans. Commun.*, **47**:844–855, June 1999.

[7] X. Tang, M.-S. Alouni and A.J. Goldsmith, Effect of channel estimation error on M-QAM BER performance in Rayleigh fading, *IEEE Trans. Commun.*, **47**(12):1856–1864, Dec. 1999.

[8] S.T. Chung and A.J. Goldsmith, Degrees of freedom in adaptive modulation: A unified view, *IEEE Trans. Commun.*, **49**(9):1561–1571, Sep. 2001.

[9] J.G. Proakis, *Digital Communications*, McGraw–Hill Inc., New York, NY, 2nd edition, 1989.

[10] R.H. Clarke, A statistical theory of mobile-radio reception, *The Bell System Technical Journal*, **47**(6):957–1000, July-August 1968.

[11] Jr W.C. Jakes, Multipath interference, In Jr W.C. Jakes, editor, *Microwave Mobile Communications*. IEEE, Piscataway, NY, 1974.

[12] R.B. Ertel, P. Cardieri, K.W. Sowerby, T. Rappaport and J.H. Reed, Overview of spatial channel models for antenna array communication systems, *IEEE Personal Communications*, pages 10–22, Feb. 1998.

[13] A.F. Molisch J. Fuhl and E. Bonek, Unified channel models for mobile radio systems with smart antennas, *IEE Proc.-Radar, Sonar Navig.*, **145**(1):32–41, Feb. 1998.

[14] T. Ekman, Analysis of the LS estimation error on a Rayleigh fading channel, In *Proc. Vehic. Tech. Conf. Fall*, Rhodes, Greece, May 2001.

[15] E. Lindskog, *Space-Time Processing and Equalization for Wireless Communications*, PhD thesis, Uppsala University, Uppsala, Sweden, 1999.

[16] H. Asplund A.A. Glazunov and J.-E. Berg, Statistical analysis of measured short-term impulse response functions of 1.88Ghz radio channels in Stockholm with corresponding channel model, In *Proc. IEEE* 50^{th} *Vehic. Tech. Conf. Fall*, pages 107–111, Amsterdam, Sept 1999.

[17] J.-K. Hwang and J. H. Winters, Sinusoidal modeling and prediction of fast fading processes, In *Proc. IEEE Globecom'98*, pages 892–897, Sidney, Nov 1998.

[18] S.H. Jensen J.B. Andersen, J. Jensen and F. Fredriksen, Prediction of future fading based on past measurements, In *Proc. IEEE* 50^{th} *Vehic. Tech. Conf. Fall*, pages 151–155, Amsterdam, Sep. 1999.

[19] Hao Ling Liang Dong, Guanghan Xu, Prediction of fast fading mobile radio channels in wideband communication systems, In *GLOBECOM '01*, volume 6, pages 3287 – 3291, San Antonio, TX, USA, Nov. 2001.

[20] A. Duel-Hallen T. Eycoz and H. Hallen, Deterministic channel modelling and long range prediction of fast fading radio channels, *IEEE Communications Letters*, **2**(9):254–256, Sept. 1998.

[21] R. Vaughan, P. Teal and R. Raich, Short-term mobile channel prediction using discrete scatterer propagation model and subspace signal processing algorithms, In *Proc. Vehic. Tech. Conf. Fall*, pages 751 –758, Boston, MA, Sep. 2000.

[22] A. Duel-Hallen, S. Hu and H. Hallen, Long-range prediction of fading signals, *IEEE Signal Processing Magazine*, **17**(3):62–75, May 2000.

[23] J.M.A. Tanskanen, Aiping Huang, T.I. Laakso and S.J. Ovaska, Prediction of received signal power in cdma cellular systems, In *Vehicular Technology Conference, 1995 IEEE 45th*, volume 2, pages 922 –926, Chicago, IL, USA, 1995.

[24] R.J. Lyman, Optimal prediction of the mobile-radio fading envelope, In *ICASSP 2002*, Orlando, Florida, May 2002.

[25] T. Ekman, G. Kubin, M. Sternad and A. Ahlén, Quadratic and linear filters for mobile radio channel prediction, In *Proc. IEEE* 50^{th} *Vehic. Tech. Conf. Fall*, pages 146–150, Amsterdam, Sep. 1999.

[26] T. Ekman and G. Kubin, Nonlinear prediction of mobile radio channels: Measurements and MARS model designs, In *Proc. Int. Conf. Acoust. Speech Sign. Process.*, Phoenix, AZ, March 1999.

[27] T. Ekman, *Prediction of mobile radio channels*, PhD thesis, Uppsala University, Uppsala, Sweden, 2001.

[28] F.C. Lau and W.M. Tam, Novel sir-estimation-based power control in a cdma mobile radio system under multipath environment, *IEEE Transactions on Vehicular Technology*, **50**(1):314–320, Jan. 2001.

[29] Szu-Lin Su Yu-Che Su Jen-Fa Huang, Grey-based power control for ds-cdma cellular mobile systems, *IEEE Transactions on Vehicular Technology*, **49**(6):2081 – 2088, Nov. 2000.

[30] Boon-Hee Soong Moh Lim Sim, E. Gunawan and Cheong-Boon Soh, Performance study of close-loop power control algorithms for a cellular cdma syste, *IEEE Transactions on Vehicular Technology*, **48**(3):911 – 921, May 1999.

[31] X.M. Gao, J.M.A. Tanskanen and S.J. Ovaska, Comparison of linear and neural network-based power prediction schemes for mobile DS/CDMA systems, In *Proc. IEEE* 46^{th} *Vehic. Tech. Conf.*, pages 61–65, Atlanta, GA, May 1996.

[32] X.M. Gao, X.Z. Gao, J.M.A. Tanskanen and S.J. Ovaska, Power control for mobile DS/CDMA systems using a modified Elman neural network, In *Proc. IEEE* 47th *Vehic. Tech. Conf.*, volume 2, pages 750–754, Phoenix, AZ, 1997.

[33] T. Chulajata et. al. S. Choe, Linear prediction at base station for closed loop power control, In *Proc. IEEE* 49th *Vehic. Tech. Conf. Spring*, Houston, Texas, May 1999.

[34] H. Holm K.J. Hole and G.E. Øien, Adaptive multidimensional coded modulation over flat fading channels, *IEEE Journal on Selected Areas in Communications*, **18**(7):1153–1158, July 2000.

[35] T. Ekman S. Falahati, A. Svensson and M. Sternad, Effect of prediction errors on adaptive modulation systems for wireless channels, In *Radiovetenskap och Kommunikation RVK 02*, Stockholm, Sweden, June 2002.

[36] J.D. Parsons, *The Mobile Radio Propagation Channel*, John Wiley & Sons, New York, NY, second edition, 2000.

[37] R. Janaswamy, *Radiowave Propagation and SMart Antennas for Wireless Communication*, Kluwer Academic Publischers, Norwell, MA, 2001.

[38] M. Cedervall B.C. Ng and A. Paulraj, A structured channel estimator for maximum-likelihood sequence estimation, *IEEE Communications Letters*, **1**(2):52–55, March 1997.

[39] G.B. Giannakis and C. Tepedelenlioğlu, Basis expansion of models and diversity techniques for blind identification and equalization of time-varying channels, *Proc. IEEE*, **86**(10):1969–1998, Oct. 1999.

[40] T. Kawano Y. Okamura, E. Ohmori and K. Fukuda, Field strength and its variability in the "vhf" and "uhf" land mobile radio service, *Rev. Elec. Com. Lab*, **16**:825–873, September-October 1968.

[41] M. Toeltsch, J. Laurila, A.F. Molisch et.al., Spatial characterization of urban mobile radio channels, In *Proc. Vehic. Tech. Conf. Fall*, Rhodes, Greece, May 2001.

[42] A. Papoulis, *Probability,Random Variables, and Stochastic Processes*, McGraw–Hill Inc., New York, NY, third edition, 1991.

[43] M.J. Gans, A power-spectral theory of propagation in the mobile-radio environment, *IEEE Trans. on Vehicular Technology*, **21**(1):27–38, Feb. 1972.

[44] T. Aulin, A modified model foir the fading signal at a mobile radio channel, *IEEE Trans. on Vehicular Technology*, **28**(3):182–203, Aug. 1979.

[45] J.D. Parsons and A.M.D. Turkmani, Characterisation of mobile radio signals: model description, *IEE Proceedings-1*, **138**(6):549–556, Dec. 1991.

[46] A.F. Molisch J. Laurila and E. Bonek, Influence of the scattered distribution on power delay profiles and azimuthal power spectra of mobile radio channels, In *Proc. IEEE* 5^{th} *ISSSTA*, pages 267–271, Sun City, Southafrica, Sep. 1998.

[47] S.O. Rice, Statistical properties of a sine wave plus random noise, *The Bell System Technical Journal*, **27**:109–157, Jan. 1947.

[48] U. Martin, Spatio-temporal radio channel characteristics in urban macrocells, *IEE Proc.-Radar, Sonar Navig.*, **145**(1):42–49, Feb. 1998.

[49] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International (UK) Ltd, Hertfordshire, UK, 1989.

[50] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, NJ, second edition, 1999.

[51] J. Homer, I. Mareels, R.R. Bitmead, B. Wahlberg and F. Gustafsson, LMS estimation via structural detection, *IEEE Trans. on Signal Processing*, **46**(10):2651–2663, Oct. 1998.

[52] A. Ahlén, L. Lindbom and M. Sternad, Tracking of time-varying systems. Part II: Analysis of stability and performance of adaptive algorithms with time-invariant gains, *IEEE Trans. on Signal Processing*, 2001, Submitted for publication.

[53] F. Bowman, *Introduction to Bessel Functions*, Dover Publications Inc., New York, NY, 1958.

[54] T. Ekman, A. Jakobsson and P. Stoica, On efficient implementation of the Capon algorithm, In *EUSIPCO2000*, Tampere, Suomi, September 2000.

[55] M. Sternad, T. Ekman and A. Ahlén, Power prediction on broadband channels, In *Proc. Vehic. Tech. Conf. Fall*, Rhodes, Greece, May 2001.

[56] A. Ahlén and M. Sternad, Wiener filter design using polynomial equations, *IEEE Trans. Signal Processing*, **39**:2387–2399, Nov. 1991.

[57] A. Arredondo, K.R. Dandekar, Guanghan Xu , Vector channel modeling and prediction for the improvement of downlink received power, *IEEE Transactions on Communications*, **50**(7):1121– 1129, 2002.

[58] L.L. Scharf, *Stochastical Signal Prosessing, Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, MA, 1991.

[59] M.H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, New York, NY, 1996.

[60] K.J. Åström and B. Wittenmark, *Computer Controlled Systems: Theory and Design*, Prentice Hall, Inc., Englewood Cliffs, NJ, second edition, 1990.

[61] T. Ekman, M. Sternad and A. Ahlén, Unbiased power prediction of rayleigh fading channels, In *Proc. Vehic. Tech. Conf. VTC Fall 2002*, Vancouver, Canada, Sept 2002.

[62] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Co., New York, 1984.

[63] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, New Jersy, 3rd edition, 1996.

[64] W.B. Davenport. and W.L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.

[65] J.G. Proakis, *Digital Communications*, McGraw–Hill Inc., New York, NY, 4th edition, 2001.