Duc Van Duong

Analysis and Optimization of Pilot-Aided Adaptive Coded Modulation Under Noisy Channel State Information and Antenna Diversity

Doctoral thesis
for the degree of philosophiae doctor

Trondheim, August 2006

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics
and Electrical Engineering
Department of Electronics and Telecommunications



NTNU
Norwegian University of Science and Technology

Doctoral thesis
for the degree of philosophiae doctor

Faculty of Information Technology, Mathematics and
Electrical Engineering
Department of Electronics and Telecommunications

© Duc Van Duong

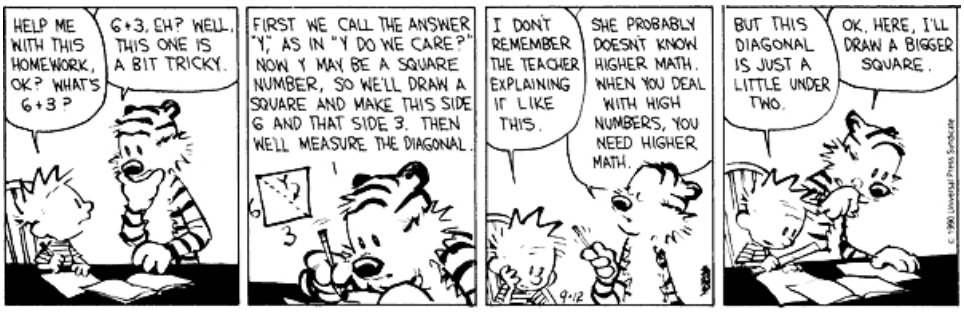
ISBN 82-471-8081-2 (printed version)
ISBN 82-471-8080-4 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2006:157

Printed by NTNU-trykk

Kính gởi Ba Má

To the teachers
For my loved ones



Calvin and Hobbes by Bill Watterson
Copyright © 1985–1996 by Universal Press Syndicate

Abstract

The thesis is largely built on a collection of published and submitted papers where the main focus is to analyze and optimize single-carrier adaptive coded modulation systems with and without antenna diversity. Multidimensional trellis codes are used as component codes.

The majority of the analysis is done with both estimation and prediction errors being incorporated. Both channel estimation and prediction are performed using a pilot-symbol-assisted modulation scheme. Thus, known pilot symbols (overhead information) must be transmitted; which consumes power and also degrades system spectral efficiency. Both power consumption and pilot insertion frequency are optimized such that they are kept at necessary values to maximize system throughput without sacrificing the error rate performance. The results show that efficient and reliable system performance can be achieved over a wide range of the considered average channel quality.

Going from a single-input single-output system to both spatially uncorrelated and correlated single-input multiple-ouput (SIMO) systems, and further to an uncorrelated multiple-input multiple-output (MIMO) diversity system, is the evolution of the thesis. In the SIMO case, maximum ratio combining is used to combine the incoming signals, whereas the signals are space-time combined in the MIMO diversity system. The multiple-input single-output system comes out as a special case of a MIMO system.

Besides the spatially uncorrelated antenna array, the effect of spatial correlation is also considered in the SIMO case. In this case, only prediction error is considered and channel estimation is assumed to be perfect. At first, the impact of spatial correlation in a predicted system originally designed to operate on uncorrelated channels is quanitifed. Then, a maximum a posteriori (MAP)-optimal "space-time predictor" is derived to take spatial correlation into account. As expected, the results show that the throughput is still lower than the uncorrelated system, but the degradation is decreased when the MAP-optimal space-time predictor is used. Thus, by exploiting the correlation properly, the degradation can be reduced.

By numerical examples, we demonstrate the potential effect of limiting the predictor complexity, of fixing the pilot spacing, as well as of assuming perfect estimation. The two first simplifications imply lower system complexity and feedback rate, whereas the last assumption is usually made to ease the mathematical analysis. The numerical examples indicate that all the simplifications can be done without serious impact on the predicted system performance.

Preface

This dissertation is submitted in partial fulfillment of the requirements for the degree of *Philosophiae doctor* (PhD) at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU). My main and co-supervisor have been Professor Geir E. Øien and Associate Professor Lars Lundheim, respectively. Both of them are with the Department of Electronics and Telecommunications, NTNU.

The studies have been carried out in the period from August 2002 to September 2006. The work includes the equivalent of approximately one and a half semester course studies and one year of being a teaching assistant in various graduate classes, as well as being an advisor for some master theses and term projects. I have spent most of the time at NTNU, but during the academic period August–November 2005 I visited Professor Lajos Hanzo and his research group at the School of Electronics and Computer Science, University of Southampton, UK. I have also spent the period January–July 2006 at University Graduate Center (UniK) in Kjeller, Norway. From the end of July 2006 I joined Norwegian Defence Research Establishment (FFI) located in Kjeller, Norway.

The work included in this thesis is funded by the Norwegian Research Coucil through the "Bandwidth-Efficient and Adaptive Transmission Schemes for Wireless Multimedia Communications" (BEATS) project (www.iet.ntnu.no/projects/beats/), while the assistantship was financed by Department of Electronics and Telecommunications, NTNU.

Acknowledgments

After 4 years being a PhD student and doing research, the first person I would like to thank is my supervisor, Professor Geir E. Øien. He was always there when I needed help and his way of being friendly, as well as a supervisor, made it easy to do research. Also, he had an open mind

and positive attitude when discussing problems—both research and non-research issues.

Moreover, I would like to thank Professor Kjell J. Hole and Doktor ingeniør Bengt Holter for our discussions and collaboration. In addition, I want to thank Bengt Holter for proof reading my manuscript. Furthermore, I value the discussions I had with Associate Professor Lars Lundheim and Professor Tor Ramstad. My thank also goes to the colleagues in the Signal Processing Group, NTNU. Especially, I want to thank Fredrik Hekland, Greg Håkonsen, and Vidar Markhus for the fruitful discussions we had—both on scientific and non-scientific issues. I appreciate other colleagues/friends who are not mentioned by name for our not-so-regularly "friday-beers" in Trondheim, a social event and also a relaxing time.

I am grateful to Professor Lajos Hanzo for his guidance and his flexibility upon accepting my visit, since it was public holidays most of the time during my stay in Southampton. The hospitality of UniK is also thanked. Furthermore, I would like to thank FFI for letting me finish the thesis while being their employee.

Finally, I would like to thank my parents, my sisters, and my dear—Hanh Bich Thi Tran—for their support and unconditional love. Especially, I am grateful to Hanh who always has faith and confidence in me since the day we first met. Also, I appreciate my friends (in the so-called HND-group) who have made my student-life in Trondheim (in the years 1998–2005) more colorful.

Duc V. Duong
Oslo, August 2006

Contents

Abstract i

Preface iii

Contents v

Abbreviations ix

Notation and Symbols xi

1 Introduction to Adaptive Wireless Communications 1
- 1.1 Spectral Efficient and Adaptive Transmission Using Feedback 2
- 1.2 Key Elements of Adaptive Coded Modulation Systems Using Pilot-Aided Channel Monitoring 4
  - 1.2.1 Prediction and Estimation of the Channel Condition 6
  - 1.2.2 Selection Rule for the Modulation and System Parameters 7
  - 1.2.3 System Performance Measurements 8
- 1.3 Characterizations of Fading Channels 8
  - 1.3.1 Effects of Space Diversity and Diversity Combining 11
- 1.4 Outline 14

2 Preliminaries of Our Adaptive Coded Modulation Systems 17
- 2.1 Background 17
- 2.2 System Description 19
- 2.3 ASE Performance 21
- 2.4 BER Performance 22
- 2.5 Transmit Powers 23
- 2.6 MAP-Optimal Channel Estimators and Predictors 25
  - 2.6.1 Channel Estimator 25
  - 2.6.2 Channel Predictor 27

2.6.3 Visualization of MAP-optimal Prediction 28
2.7 System Parameters Used in Numerical Example 30

3 Adaptive Coded Modulation With Imperfect Channel State Information: the SISO Case 33
3.1 Introduction 33
3.2 System Model 34
3.3 BER Analysis 36
3.3.1 BER in the Presence of Channel Estimation Errors . 36
3.3.2 BER in the Presence of Both Channel Estimation and Prediction Errors 37
3.4 ASE Analysis 40
3.5 Numerical Example 42
3.6 Concluding Remarks 46

4 Adaptive Coded Modulation With Receive Antenna Diversity and Imperfect Channel Knowledge 47
4.1 Introduction 47
4.2 Description of the System 48
4.3 BER Performance Analysis 49
4.3.1 BER Analysis in the Presence of Estimation Errors . 49
4.3.2 BER Analysis in the Presence of Both Estimation and Prediction Errors 50
4.3.3 Overall Average BER Performance Analysis 53
4.4 Optimization of ASE 53
4.4.1 ASE Performance Analysis 53
4.4.2 Optimization 53
4.5 Average Fade Region Duration 54
4.6 Numerical Analysis 54
4.7 Conclusions 61

5 Impact of Spatial Correlation 65
5.1 Introduction 65
5.2 Space-Time Correlation Model 66
5.2.1 Spatial Correlation Models 67
5.3 Link Adaptation in Spatially Correlated Antenna Diversity: Independently Predicted Branches 68
5.3.1 The Combined Signal Statistics 68
5.3.2 System Performance Analysis 70
5.3.3 Numerical Calculation 71

5.4 Link Adaptation in Spatially Correlated Antenna Diversity: Jointly Predicted Branches . . . 77
5.4.1 System Performance . . . 80
5.5 Conclusions . . . 85

6 Orthogonal Space-Time Block Coded Rate-Adaptive Systems With Imperfect CSI 87
6.1 Introduction . . . 87
6.2 System Model . . . 89
6.3 Channel Estimation and Prediction . . . 91
6.3.1 Estimation . . . 92
6.3.2 Prediction . . . 92
6.4 Evaluation of BER Performance . . . 93
6.5 Optimization of ASE . . . 96
6.6 Numerical Example and Discussion . . . 97
6.7 Concluding Remarks . . . 102

7 Effects of Simplifications and Suboptimalities 103
7.1 Fixing the Pilot Spacing . . . 104
7.2 Effect of Complexity-Limited Predictor . . . 105
7.3 Effect of the Assumption on Perfect Estimation . . . 107

8 Conclusions 113
8.1 Results and Contributions . . . 113
8.1.1 Contributions of the Thesis . . . 114
8.2 Suggestions for Further Research . . . 115

A Derivation of the MAP-Optimal Predictor and Estimator with the Corresponding MMSE 119
A.1 MAP-optimal Predictor and MMSE of the Prediction Error 119
A.2 Derivation of the Predictor Used in Other Papers . . . 121

B Derivation of BER(M_n) in Chapter 3 125
B.1 Useful Integration Rules and Identities . . . 125
B.2 Derivations . . . 126
B.2.1 Calculation of $\mathcal{T}(0,\infty)$ . . . 127
B.2.2 Calculation of $\mathcal{T}(0,|h_e|_{n,T})$ . . . 127
B.2.3 Calculation of $\mathcal{F}(0,|h_e|_{n,T})$ . . . 128
B.3 BER$(M_n|\hat{\gamma})$ . . . 129
B.4 BER(M_n) . . . 129
B.4.1 Calculation of $\mathcal{B}1(M_n)$ . . . 129

B.4.2 Calculation of $\mathcal{B}21(M_n)$ 130
B.4.3 Calculation of $\mathcal{B}22(M_n)$ 130

C An Alternative Proof of the Expression for $\mathsf{Var}\left(h_{e;\mu}|h_{p;\mu}\right)$ **131**

D Calculation of the Correlation Coefficient ρ in Equation (5.24) **133**

Bibliography **137**

Abbreviations

1-D	1-dimensional
2-D	2-dimensional
3-D	3-dimensional
4-D	4-dimensional
ACM	Adaptive coded modulation
ASE	Average spectral efficiency
AWGN	Additive white Gaussian noise
BER	Bit-error-rate
BS	Base station
CDF	Cumulative distribution function
CSI	Channel state information
CSNR	Channel-signal-to-noise ratio
DA	Data aided
dB	Decibel
LOS	Line-of-sight
MAP	Maximum a posteriori
MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
ML	Maximum likelihood
MRC	Maximum ratio combining
MS	Mobile station
MSE	Mean square error
MMSE	Minimum mean square error
M-QAM	M-ary (or multilevel) quadrature amplitude modulation
NDA	Non data aided
OFDM	Orthogonal frequency division multiplexing
PDF	Probability density function
QAM	Quadrature amplitude modulation
RP	Random process

RV	Random variable
SIMO	Single-input multiple-output
SISO	Single-input single-output
SM	Spatial multiplexing
STBC	Space-time block coding
TCM	Trellis code modulation
WSS	Wide sense stationary
XIXO	Any combination of the number of transmit and receive antennas

Notation and Symbols

Vectors are usually written in bold, upright, lowercase, whereas matrices are in bold, upright uppercase. Bold faced italic lower case Greek characters are also used for vectors. For differences, compare "w", "$\mathbf{w}$", "W", "$\mathbf{W}$", "ω", and "$\boldsymbol{\omega}$". Vectors are column vectors, unless otherwise explicitly defined.

In the following notation list, the dimension of vectors and matrices are also given.

$\otimes$	Kronecker product
w^*	Complex conjugate of w
$\mathbf{w}^{\mathsf{T}}$	Transpose of vector $\mathbf{w}$
$\mathbf{w}^{\mathsf{H}}$	Hermitian (conjugate transpose) of vector $\mathbf{w}$
$\mathbf{W}^{-1}$	Matrix inversion
$\lvert w \rvert$	Absolute value
$\lVert \mathbf{w} \rVert_{\mathsf{F}}$	Frobenius norm of a vector, $\sqrt{\mathbf{w}^{\mathsf{H}}\mathbf{w}}$
$\lfloor w \rfloor$	Taking the integer part of w only
$[\mathbf{W}]_{mn}$	Element in column m and row n of the matrix $\mathbf{W}$
$\jmath$	$\sqrt{-1}$
$\mathsf{Cov}(\cdot,\cdot)$	Covariance
$\mathsf{E}[\cdot]$	Expectation
$\mathsf{tr}\{\cdot\}$	Trace operator
$\mathsf{Var}(\cdot)$	Variance
$\mathsf{vec}(\cdot)$	Vectorized operator

α	The variable which determines how power is distributed between pilot and data symbols
β, β_μ	Absolute value of fading gain, and of the μth branch
γ	Total instantaneous true channel signal-to-noise ratio (CSNR)
γ_d	Total instantaneous detected CSNR assuming perfect estimation
$\gamma_{n,T}$	The lowest CSNR value for which the exponential approximation of the BER–CSNR relationship is relevant when code n is used
$\bar{\gamma}$	Expected value of the total true CSNR
$\bar{\gamma}_b$	Expected subchannel (*one receive branch*) CSNR
$\bar{\gamma}_\mu$, $\bar{\gamma}_{\mu\nu}$	Average CSNR on receive antanna μ, and on any subchannel $\mu\nu$
$\bar{\gamma}_d$	Total average detected CSNR assuming perfect estimation
$\hat{\gamma}$	Total instantaneous predicted CSNR
$\hat{\gamma}_n$	The lowest predicted CSNR attaining the target BER when code n is used
$\bar{\hat{\gamma}}$	Expected value of the total predicted CSNR
$\Gamma(\cdot)$	Gamma function
$\Gamma(\cdot,\cdot)$	Incomplete gamma function
$\overline{\Gamma}(\cdot,\cdot)$	Normalized incomplete gamma function
$\epsilon_e(k;l)$	Estimation error
$\epsilon_{e;\mu}(k;l)$, $\epsilon_{e;\mu\nu}(k;l)$	Estimation error of the μth receive branch, and of any μ-ν transmit-receive antenna pair
$\epsilon_p(k;l)$	Prediction error
$\epsilon_{p;\mu}(k;l)$, $\epsilon_{p;\mu\nu}(k;l)$	Prediction error of the μth receive branch, and of any μ-ν transmit-receive antenna pair

$\theta, \theta_1, \theta_2$	Scale factors of the gamma probability density function (PDF)
λ_κ	Eigenvalue number κ of the channel covariance matrix
μ	Transmit antenna branch index
ρ	Correlation between true and predicted CSNR
$\rho_{h,s}$	Spatial correlation of the fading gain
$\rho_{h,t}$	Temporal correlation of the fading gain
ρ_s	Power correlation between any two adjacent antenna branches
ϱ	Normalized correlation between the estimation error and the predicted channel
$\sigma_e^2(l)$	Variance of the estimation error
$\sigma_{e;\mu}^2(l)$, $\sigma_{e;\mu\nu}^2(l)$	Variance of the estimation error of the μth receive branch, and of the $\mu\nu$th subchannel
$\sigma_p^2(l)$	Variance of the prediction error
$\sigma_{p;\mu}^2(l)$, $\sigma_{p;\mu\nu}^2(l)$	Variance of the prediction error of the μth receive branch, and of the $\mu\nu$th subchannel
$\sigma_{h_e\|h_p}^2(l)$	Variance of estimated channel given the predicted channel
$\sigma_{h_{e;\mu}\|h_{p;\mu}}^2$, $\sigma_{h_{e;\mu\nu}\|h_{p;\mu\nu}}^2$	Variance of estimated channel given the predicted channel: on the μth receive branch, and on the $\mu\nu$th subchannel
τ	System delay
$\bar{\tau}_n$	Average fade region duration (average time the fading remains between two specified levels)
ν	Receive antenna branch index
χ_n	Level crossing rate for the nth level
Φ	Collection of constants and variables
Ω	Average fading power

$\{a_n\}_{n=1}^{N=8}$	Set of parameter in exponential approximation to the BER–CSNR relationship for code no. n
$\{a_n(\ell)\}_{\ell=1,\ n=1}^{\mathcal{L}=3,\ N=8}$	Another set of parameter in exponential approximation to the BER–CSNR relationship for code no. n
A_n, $A_n(\ell)$, $A'_n(\ell)$	Collections of constants and variables
ASE	Average spectral efficieny
$\{b_n\}_{n=1}^{N=8}$	Set of parameter in exponential approximation to the BER–CSNR relationship for code no. n
$\{b_n(\ell)\}_{\ell=1,\ n=1}^{\mathcal{L}=3,\ N=8}$	Another set of parameter in exponential approximation to the BER–CSNR relationship for code no. n
$\mathrm{BER}(M_n)$	Average BER of the constellation M_n
$\mathrm{BER}(M_n\|\hat{\gamma})$	BER of the constellation M_n given the predicted CSNR $\hat{\gamma}$
$\mathrm{BER}\,(M_n\|h_e)$	BER of the constellation M_n given estimated channel
$\mathrm{BER}\left(M_n\middle\|\left\{h_{e;\mu}\right\}\right)$	BER of the constellation M_n given the set of estimated receive branches
$\mathrm{BER}\left(M_n\middle\|\left\{h_{e;\mu\nu}\right\}\right)$	BER of the constellation M_n given the set of estimated subchannels
$\mathrm{BER}\left(M_n\|h_p\right)$	BER of the constellation M_n given predicted channel
$\mathrm{BER}\left(M_n\middle\|\left\{h_{p;\mu}\right\}\right)$	BER of the constellation M_n given the set of predicted receive branches
$\mathrm{BER}\left(M_n\middle\|\left\{h_{p;\mu\nu}\right\}\right)$	BER of the constellation M_n given the set of predicted subchannels
$\overline{\mathrm{BER}}$	Average BER
BER_0	Target BER
c	Speed of light, $c = 3 \cdot 10^8$ m/s

$\mathbb{C}$	Set of complex numbers
$\mathcal{E}$	Average transmit power each pilot and data symbol can have
$\mathcal{E}_d$	Actual transmit data power
$\bar{\mathcal{E}}_d$	Average transmit data power
$\mathcal{E}_{\mathrm{pl}}$	Actual transmit pilot power
D	Number of frames ahead in time to be predicted
$\boldsymbol{\mathcal{D}}(\mathbf{x})$	Diagonal matrix with vector $\mathbf{x}$ on its diagonal (dimension: $K_e \times K_e$ or $K_p \times K_p$ depending on the matrix is used for estimation or prediction)
f_c	Carrier frequency
f_d	Maximum Doppler frequency
f_s	Pilot sampling frequency
$\mathcal{G}(m_d, \theta)$	Shorthand notation for gamma distribution with shape parameter m_d and scale parameter θ
$h(k;l)$	Complex fading gain at the kth frame and the lth symbol in that frame
$h_e(k;l)$	Estimated channel
$h_{e;\mu}(k;l)$, $h_{e;\mu\nu}(k;l)$	Estimated channel of the μth receive branch, and of the $\mu\nu$th subchannel
$h_p(k;l)$	Predicted channel
$h_{p;\mu}(k;l)$, $h_{p;\mu\nu}(k;l)$	Predicted channel of the μth receive branch, and of the $\mu\nu$th subchannel
$\mathbf{h}$	Array of the channel gains (dimension: $n_{\mathrm{R}} \times 1$)
$\mathbf{h}_p$	Vector of jointly predicted channel gains (dimension: $n_{\mathrm{R}} \times 1$)
$\mathbf{h}_{\mathrm{pl};\mu}$	Vector of pilot symbols received on the μth receive branch (dimension: $K_e \times 1$ or $K_p \times 1$ depending on if the vector is used for estimation or prediction)

$\underline{\mathbf{h}}_{\mathrm{pl}}$	Stacked vector of the pilot symbols received from all the receive branches (dimension: $K_p n_{\mathrm{R}} \times 1$)
$\mathbf{H}$	Channel gain matrix (dimension: $n_{\mathrm{R}} \times n_{\mathrm{T}}$)
$\boldsymbol{\mathcal{H}}$	Channel gain matrix corresponding to the buffered pilot symbols from all the subchannels (dimension: $K_p \times n_{\mathrm{R}}$)
$I_\nu(\cdot)$	Modified Bessel function of the first kind and of order ν
$\mathbf{I}_K$	$K \times K$ identity matrix
$J_\nu(\cdot)$	Bessel function of the first kind and of order ν
K	Rician K-factor
K_e	Number of pilot symbols the estimation is based upon; estimator filter length
K_p	Number of pilot symbols the prediction is based upon; predictor filter length
$\mathcal{L}$	Number of exponential functions which approximate the simulated BER performance
L	Distance (counted in number of channel symbols) between two consecutive pilot symbols
L_b	Distance (counted in number of channel symbols) between two consecutive pilot symbols on *a single antenna branch* after space-time block coding (STBC)
m_d	Shape parameter of the gamma distribution; $\mathcal{G}(m_d, \theta)$
M, M_n	Number of symbols in a signal constellation, and in constellation no. n
n	1. Code number 2. Predicted CSNR bin number
$n(k;l)$	Additive white Gaussian noise (AWGN) at frame index k and symbol index l in that frame
$\mathbf{n}$	Vector of AWGN (dimension: $n_{\mathrm{R}} \times 1$)

$\underline{\mathbf{n}}_{\mathrm{pl}}$	Stacked vector of the received channel noise corresponding to the received pilot symbol matrix (dimension: $K_p n_{\mathrm{R}} \times 1$)
n_{T}, n_{R}	Number of transmit/receive antennas
N	Number of codes available in an adaptive transmission system
$\boldsymbol{\mathcal{N}}$	Channel noise matrix corresponding to the buffered pilot instants from all the subchannels (dimension: $K_p \times n_{\mathrm{R}}$)
N_0	Noise spectral density
$\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_4$	Orthogonal design of STBC for 1, 2, and 4 transmit antennas
P_n	Probability that code n is used
P_{out}	Outage probability or probability of no transmission
$Q_K(\cdot,\cdot)$	Generalized Marcum Q-function
r	Ratio between the total average predicted and the total average actual CSNR
r_d	Ratio between the total average detected and the total average actual CSNR when assuming perfect estimation
$\mathbf{r}_e$, $\mathbf{r}_p$	Vector containing the covariance between the fading at pilot symbol instants and the fading to be estimated/predicted (dimension: $K_e \times 1$ and $K_p \times 1$)
$R(\cdot)$	Normalized autocorrelation function
R_n	Spectral efficiency (the number of information bits per second per unit bandwidth) of code no. n
R_n^{STBC}	Spectral efficiency (the number of information bits per second per unit bandwidth) of code no. n after STBC
R_s	Rate of the employed space-time block code

$\mathbf{R}_e$, $\mathbf{R}_p$	Normalized correlation matrix for estimation and prediction cases (dimension: $K_e \times K_e$ and $K_p \times K_p$)
$\mathbf{R}_{\underline{\mathbf{h}}}$	Block-Hermitian autocovariance matrix of the stacked vector $\underline{\mathbf{h}}$ (dimension: $K_p n_{\mathrm{R}} \times K_p n_{\mathrm{R}}$)
$\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}$	Cross-covariance matrix of the stacked vector $\underline{\tilde{\mathbf{y}}}$ and the vector $\mathbf{h}$ (dimension: $K_p n_{\mathrm{R}} \times n_{\mathrm{R}}$)
$\mathbb{R}^+$	Set of nonnegative real numbers
$s(k;l)$	Transmitted symbols (both data and pilot; depending on the index l)
S	Number of symbols used by the space-time block encoder
T	Length of the signal sequence after space-time block coding
T_s	Channel symbol duration
$\mathcal{T}(\cdot,\cdot)$	Component in the expression for BER$(M_n\|h_p)$
$\mathcal{F}(\cdot,\cdot)$	Component in the expression for BER$(M_n\|h_p)$
v	Terminal speed
$\mathbf{w}_e$, $\boldsymbol{\omega}_e$	Maximum a posteriori (MAP) optimal estimator of length K_e
$\mathbf{w}_p$, $\boldsymbol{\omega}_p$	Maximum a posteriori (MAP) optimal predictor of length K_p
$y_d(k;l)$	Received data symbol (at frame k and symbol number l in that frame)
$y_{d;\mu}(k;l)$, $y_{d;\mu\nu}(k;l)$	Data symbol received on the branch μ, and on the subchannel $\mu\nu$
$y_{\mathrm{pl}}(k;l)$	Received pilot symbol (at frame k and symbol number l in that frame)
$y_{\mathrm{pl};\mu}(k;l)$, $y_{\mathrm{pl};\mu\nu}(k;l)$	Pilot symbol received on the branch μ, and on the subchannel $\mu\nu$

$\mathbf{y}_{\mathrm{pl}}$, $\mathbf{y}_{\mathrm{pl};\mu}$, $\mathbf{y}_{\mathrm{pl};\mu\nu}$	Vector of buffered pilot symbols received in *one branch*, in the μth receive branch, and in the $\mu\nu$th subchannel (dimension: $K_e \times 1$ or $K_p \times 1$ depending on the vector is used for estimation or prediction)
$\tilde{\mathbf{y}}$	Vector of pilot symbols scaled with its symbol and power (dimension: $K_e \times 1$ or $K_p \times 1$ depending on the vector is used for estimation or prediction)
$\underline{\tilde{\mathbf{y}}}$	Stacked vector of pilot symbols scaled with its symbol and power (dimension: $K_p n_{\mathrm{R}} \times 1$)
$\boldsymbol{\mathcal{Y}}$	Matrix containing the buffered pilot symbols from all the subchannels (dimension: $K_p \times n_{\mathrm{R}}$)
$\mathbf{z}_d$, $\mathbf{z}_d(k;l)$	Vector of received, noisy, and faded data symbols used in MIMO system (dimension: $n_{\mathrm{R}} \times 1$)
$\mathbf{z}_{\mathrm{pl}}$, $\mathbf{z}_{\mathrm{pl}}(k;l)$	Vector of received, noisy, and faded pilot symbols used in MIMO system (dimension: $n_{\mathrm{R}} \times 1$)

Chapter 1

Introduction to Adaptive Wireless Communications

The demand for reliable high-rate data communication over wireless channels gives rise to the need for spectrally efficient transmission schemes. This is due to the fact that the bandwidth is scarce, and both spectrum and power usage are strictly regulated. The basic idea of spectrally efficient transmission is to transmit many information bits per second per unit bandwidth on the average while maintaining a certain quality.

One way of realizing such spectrally efficient communication is by *adaptive transmission*. The success of such an adaptive transmission scheme is strongly dependent on the knowledge of the channel at the transmitter. Thus, one challenge is to extract the best possible estimate[1] of the channel at any given time. However, nature is seldom kind when it comes to wireless communications. The wireless signal, when travelling from a source to a destination, is obstructed by different objects. These objects can be small or big as e.g. leaves, trees, cars, buildings, and so on. Depending on the size of these obstacles, the signal may be reflected, scattered, and/or diffracted. Reflection of signals occurs when the signal is met by obstacles which are comparable in size to the wavelength of the signal or larger. When the obstructing object is less than the wavelength of the signal, scattering will occur. Diffraction happens when signal "bends around" obstructing objects with irregular edges [Rappaport, 2002].

[1] In general, the term estimate of the channel here can be any kind of channel measurements which are used for adaptation purposes. The predicted/estimated channel gain or channel-signal-to-noise ratio (CSNR) is one possible measurement. Thus, it is noted that, we sometimes only use the term channel estimation or prediction for both estimation and prediction. Later on, in this thesis, we will use the channel estimates for decoding and detection while the prediction is used for system adaptation.

A common result of these three phenomena is that the signal will arrive at the receiver from different paths and with different delays. The different replica of the signal are then summed constructively or destructively, causing fluctuation of the signal level such that the signal can be severely attenuated. This phenomenon is generally known as fading. Besides the restricted spectrum and power usage, the fading also gives rise to the need of robust and efficient transmission.

1.1 Spectral Efficient and Adaptive Transmission Using Feedback

In fixed-rate and fixed-power systems, the transmission must be dimensioned relative to the worst case scenario of the channel—i.e. the system must be designed to perform acceptably in deep fades. This results in poor performance and inefficient use of spectrum when the channel condition is good. As opposed to this, spectrally efficient adaptive transmission schemes take the advantage of having a good channel by sending more bits. The rate is decreased as the channel is getting worse, and most often goes into idle when the channel is below a certain quality (then the system is said to be in outage). In order to do so, the system needs *feedback* information to assist what the transmitter should do.

Adaptive transmission over a fading channel using feedback is not a new idea and was first proposed by Hayes in 1968. In that paper, transmit power adaptation was introduced. A disadvantage of the usage of a variable-power scheme to combat fading in deep fades is that the potential co-channel interference may increase. Some years later, in 1972, Cavers suggested an adaptive symbol duration scheme. This scheme works at the expense of a variable bandwidth, which is difficult to implement in practice. As an answer to both problems, an uncoded adaptive star-QAM scheme was proposed by Webb and Steele in 1995 and a coding scheme by Goldsmith and Chua in 1998. Also, a combination of power and rate adaptation was introduced by Goldsmith and Chua in 1997.

Although using feedback in adaptive communications was already proposed, the interest in these techniques was not restored until 1997 when Goldsmith and Varaiya derived the theoretical Shannon capacity under idealized assumptions for single-user single-input single-output (SISO) fading channels using an optimal power adaptation and a variable-rate scheme. Since then many works have been conducted within the field of link adaptation. The work of Goldsmith and Varaiya was extended by Alouini and Goldsmith in their 1997 and 1999 papers to take different

spatial diversity combining techniques into account. Both [Goldsmith and Varaiya, 1997] and [Alouini and Goldsmith, 1997, 1999] assumed no feedback delay and perfect channel state information (CSI) knowledge at both transmitter and receiver.[2]

However, practical adaptive modulation schemes designed based on such *perfect* CSI are sensitive to CSI imperfections induced by channel estimation errors and feedback delay; meaning that these system works well only when channel estimation errors and/or feedback delays are sufficiently small [Goldsmith and Chua, 1997; Alouini and Goldsmith, 2000]. The system in the above two references will not operate satisfactory for normalized Doppler $\tau f_d > 0.01$ and estimation error greater than 1 dB. Both criteria are to ensure a bit-error-rate (BER) of 10^{-3}. Such a delay in combination with such a low estimation error might be hard to meet, unless the channel fading is sufficiently slow.

When the channel is slowly varying, outdated fading estimates are sufficient to track the channel for robust (in the sense that it does not require any knowledge about the Doppler frequency or the exact shape of the autocorrelation function of the channel) and reliable adaptive coding [Goeckel, 1999]. An adaptive coding scheme for slowly varying channels was also analyzed in [Vucetic, 1991]. Here the channel was modelled as a finite-state Markov process and the emphasis lies on the adaptation of the error protection. However, for fast fading, outdated channel estimates fed back to the transmitter become less useful for adaptive signaling applications. To overcome this problem, we need to predict the channel further ahead in time and, thus, long range fading prediction becomes more important. This relaxes the delay constraint considerably [Duel-Hallen, Hu, and Hallen, 2000; Hu and Duel-Hallen, 2001].

Until recently, however, research contributions to the field of adaptive communications are still based on either perfect CSI at both transmitter and receiver [Goldsmith and Chua, 1997, 1998; Alouini and Goldsmith, 2000; Chung and Goldsmith, 2001], or imperfect CSI at transmitter through the predicted channel or delayed estimated channel fed back from the receiver, while the CSI is still assumed perfect at the receiver [Goeckel, 1999; Hu, Duel-Hallen, and Hallen, 2000; Øien, Holm, and Hole, 2004; Falahati, Svensson, Ekman, and Sternad, 2004]. To the best of the author's knowledge, [Cai and Giannakis, 2005][3] is the only one paper prior to the present

[2] In addition, [Goldsmith and Varaiya, 1997] also considers the case when perfect CSI is available only to the receiver.

[3] Note that the mean $\bar{\gamma}_c$ of the PDF of [Cai and Giannakis, 2005, Eq. (31)] is wrong. The correct value is $\bar{\gamma}_c = (\tilde{\sigma}_{\hat{h}}^2 + \check{\gamma}|\rho|^2 N_0/\bar{\mathcal{E}}_d)\mathcal{E}_{di}/(N_0 + g\mathcal{E}_{di}\sigma_\epsilon^2)$.

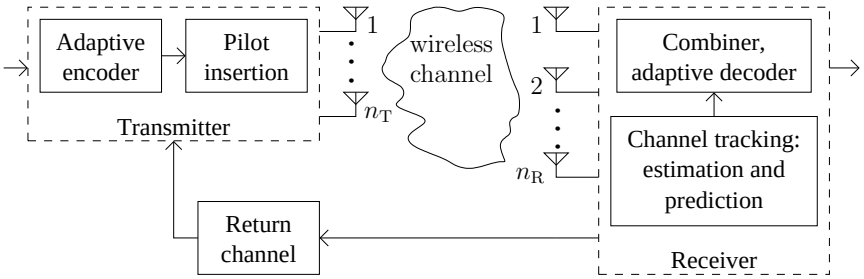


FIGURE 1.1: Generic figure of an ACM system with multiantenna transmitter and receiver. System adaptation is based on information fed back from the receiver. Channel estimation and prediction is performed using a PSAM scheme.

work dealing with both estimation error at the receiver and prediction error at the transmitter.

Adaptive systems operating on SISO channel models are considered in [Ue, Sampei, Morinaga, and Hamaguchi, 1998; Alouini, Tang, and Goldsmith, 1999; Hole, Holm, and Øien, 2000; Lau and Macleod, 2001]. Systems incorporating multiantenna transmissions are studied in [Hole, Holm, and Øien, 2001; Øien, Holm, and Hole, 2002a; Ko and Tepedelenlioglu, 2003, 2004, 2006; Zhou and Giannakis, 2004a]. For further research contributions to adaptive modulation and coding techniques, look e.g. to the tutorial paper by Hole and Øien [2001] and to the book by Hanzo, Wong, and Yee [2002].

1.2 Key Elements of Adaptive Coded Modulation Systems Using Pilot-Aided Channel Monitoring

The generic block diagram of an adaptive coded modulation (ACM) system using a pilot-symbol-assisted modulation (PSAM) channel tracking scheme is depicted in Figure 1.1. At the transmitter, the adaptive encoder generates coded bits from data information using any kind of error protecting codes. The coded bits then determine the transmit symbol taken from the appropriate modulation. For the purpose of tracking the channel using a PSAM scheme, pilot symbols with a certain power are inserted. Note that both the encoder and the pilot insertion procedure are based on *predicted* information fed back from the receiver. In SISO ($n_T = 1, n_R = 1$) and SIMO[4] ($n_T = 1, n_R > 1$) systems, the data stream is ready to be sent

[4]Single-input multiple-output

over the fading channel. One more operation needs to be done in the case of MIMO[5] ($n_T > 1, n_R > 1$) diversity or MISO[6] ($n_T > 1, n_R = 1$) diversity systems; namely the *space-time block coding* (STBC) operation before the signal is launched on the wireless channel. This operation makes it possible to transmit data simultaneously from all of the transmit antennas and, at the same time, provide full diversity. MIMO systems can also be used to maximize the data rate by sending independent data on different antennas. In this case, they are called MIMO spatial multiplexing (SM) systems. The subchannels in the multiantenna transmitter and/or receiver can be spatially correlated or uncorrelated depending on the scattering environment and/or the spacing between the antenna elements.

At the receiver, the receive signals are detected and combined (in case of multireception) and decoded. The detection and combining scheme can be done either in a coherent, partially coherent, or a noncoherent manner [Simon and Alouini, 2005]. The adaptive decoder tries to retrieve the transmitted signal adaptively. Here, both detection and decoding operations are relying on the *estimated* information which is provided from the channel tracking unit.

A common assumption in adaptive communication systems is error-free and noiseless feedback. The former assumption can be realized in practice due to the fact that the feedback information rate is small so that the information can be protected with a strong code and possibly also with automatic repeat request signaling. An example of the latter assumption is communication between a mobile station (MS) and base station (BS). In the uplink (MS-to-BS) the power is quite restricted, while in downlink (BS-to-MS) the power constraint can be more relaxed. Thus the second link can, in some cases, be considered as a noiseless link. However, perfect CSI at transmitter is still hard to obtain in practical systems since the feedback information is sent over a wireless channel which is subject to noise and fading.

The design of ACM systems to react to the channel variations is strongly dependent on the following factors:

- *Prediction and estimation of the channel conditions:* This is necessary in order to determine what code/constellation with proper parameters to be used in the future transmission. The estimation is used for detection.

- *Choice of modulation and selection rule:* Based on the predicted infor-

[5]Multiple-input multiple-output
[6]Multiple-input single-output

mation, this operation ensures that both transmitter and receiver are operating on the same set of parameters, and, at the same time, it assures a certain system quality.

- *Performance measurement:* To quantify the feasibility and quality of the adaptive scheme, some system measurement metrics must be defined and evaluated.
- Another factor like power control is also important, but it is outside of scope of the thesis.

1.2.1 Prediction and Estimation of the Channel Condition

In adaptive transmission systems, we have noted that one challenging task is to provide the transmitter the best possible and accurate information of the channel, since it is a necessary requirement for the system to function properly. Hence, reliable estimators/predictors need to be developed and employed to reduce the effect of imperfect CSI. However, another approach to get around this is by designing the adaptive transmitters and receivers which account for CSI errors explicitly. This is the choice in this thesis.

In order to perform such adaptations, information about the channel must be available to the transmitter and can be realized by means of a return channel (feedback channel). The information about the channel is conveyed by estimating/predicting the channel variations which can be done using either non-data-aided (NDA) or data-aided (DA) schemes [Meyr, Moeneclaey, and Fechtel, 1998]. While NDA channel tracking schemes perform their task based on previous correctly detected symbols, the DA scheme is based on training (pilot) symbols known to both transmitter and receiver, and which are sent regularly along with the information. How often they are transmitted is dependent on the rate of time-variance of the channel. Comparing to the NDA schemes, the training-based systems must transmit overhead information which degrades the system's throughput. However, needless to say, the channel can be better tracked with DA methods when the channel is fast varying or undergoes deep fades during which the symbols are most likely to be wrongly detected such that NDA channel tracking becomes unreliable.

The first paper analyzing DA channel estimation using a PSAM scheme for fading channels was presented by Cavers in 1991. In that system the pilot symbols are regularly multiplexed into the datastream prior to transmission over the channel. The pilot symbols are extracted at the receiver side and are used to interpolate and to smooth the channel for channel prediction and estimation, respectively. The interpolating (or smoothing) filter

coefficients are obtained by solving normal equations and they are optimal in the mean square error (MSE) sense. In such a PSAM system, the pilot spacing is kept constant regardless of the quality of the channel. However, uniform spacing pilot symbols regardless of the channel quality is not necessarily an optimal scheme [Cai and Giannakis, 2005]. Fixed pilot spacing implies that the amount of overhead information can be too large when the average channel condition is good and too small at bad average channel conditions. On average, when the channel condition is bad, the system needs densely spaced pilot symbol in order to estimate the channel reliably. In particular, in an uniform PSAM system a considerable pilot oversampling compared to the Nyquist sampling rate is required [Meyr *et al.*, 1998, Sec. 14.2.2]. On the other hand, at favourable channel conditions, the channel is so good that it will be a waste of bandwidth when having frequently pilot transmission. Thus, having improper pilot spacing (and pilot power) will decrease system performance in both cases. One way of getting around of this problem is letting the pilot insertion frequency optimally vary according to the average channel quality, i.e. an adaptive PSAM scheme [Cai and Giannakis, 2005]. In this way, the necessary overhead information is kept at a minimum. This will also be done in this thesis.

1.2.2 Selection Rule for the Modulation and System Parameters

Based on feedback information, a symbol drawn from a certain constellation (available to both transmitter and receiver) is selected for transmission. The feedback information can be channel mean [Zhou and Giannakis, 2004a], covariance [Jafar, Vishwanath, and Goldsmith, 2001], or CSNR-based [Øien *et al.*, 2004]. Depending on what information is available to the transmitter, we can do more or less smart signaling and selection scheme. For example, in a multiple transmit antenna system, if we have the whole channel matrix available, we can optimize the power usage on each transmit antenna and forming the transmit beam such that the average BER is minimized [Zhou and Giannakis, 2002]. Furthermore, we can also combine the beamforming with STBC to increase the rate. On the other hand, in order to keep a low feedback rate, predicted CSNR-based feedback can be used. This is the choice in the present thesis. Based on this, we will select the modulation, the power distribution between pilot and data symbols, and the pilot period such that the throughput is maximized. In all the cases, the parameters are chosen such that a certain quality/requirement must be fulfilled.

1.2.3 System Performance Measurements

This is to answer the question on how the performance of an ACM system is measured. To measure the system's performance, we need some quality criteria in order to decide whether or not the system is functioning satisfactorily. The measure criteria could be e.g. throughput, error performance, outage probability, and so on. These will be more formally defined in Chapter 2.

1.3 Characterizations of Fading Channels

Regardless what building blocks the system are based on, a system's performance metrics like average throughput and average error performance can not be analyzed without knowing the properties of the channel over which the signals are propagated. Hence, we will characterize the fading channel in this section.

A mathematical model for multipath propagation can be obtained by solving Maxwell's equations with appropriate boundary conditions. However, the computational complexity of such an approach makes it impractical. Also, the physical environment must be known. Moreover, if the number of reflectors is large and they are in motion relative to the transmitter and the receiver, the mathematical model will be even more difficult to achieve. In this case, statistical approximations must be applied to characterize the receive signal [Goldsmith, 2005]. Over the years, it has been shown that the rapid fluctuation of the signal in a narrowband system is well described by stochastic distributions; such as Rayleigh, Rice, or Nakagami distributions for the envelope of the fading gain [Stüber, 1996]. A system is said to be narrowband if the signal bandwidth is small compared to the channel's coherence bandwidth, i.e. the inverse delay spread of the channel.

In narrowband systems, all the different frequency components of the signal are similarly affected by the fading. Hence, the fading is said to be frequency-flat or frequency-nonselective; as in e.g. one subchannel (one tone) in a properly designed orthogonal frequency division multiplexing (OFDM) system. This is in contrast to the frequency-selective fading experienced in a wideband system [Rappaport, 2002]. Since, over a flat fading channel, different frequency components of the signal are subject to the same attenuation, the impact of fading is through a multiplicative factor. Thus, in general, the receive signal $y(k)$ in a narrowband system can be written as $y(k) = \sqrt{\mathcal{E}}h(k)s(k) + n(k)$ where $\mathcal{E}$ is the average transmit

power, $h(k)$ is the fading channel gain, $s(k)$ is the symbol to be transmitted, and $n(k)$ is the additive noise.

In general, $h(k)$ is a complex value and can be expressed—using phasor representation—as $h(k) = \beta(k)e^{j\phi(k)}$ where β and ϕ is the fading amplitude and phase, respectively. They are treated as random variables (RVs) with certain statistical properties when analyzing system performance. The phase ϕ plays an important role in modulation systems using e.g. quadrature amplitude modulation (QAM) since a symbol is determined also by the phase (in addition to the amplitude). However, in coherent detection communication systems, where the receiver is able to detect and compensate for the phase of the incoming signal, only the envelope is considered. One way of aiding the receiver to detect the phase is to use differential encoding. This would result in a 3 dB penalty in signal-to-noise ratio but with the gain of easy detection where the receiver only needs to detect the relative phase shifts in the incoming signal [Lee and Messerschmitt, 2000].

In general, the probability density function (PDF) of β is dependent on the physical propagation environment [Simon and Alouini, 2005]. When there is no line-of-sight (LOS) signal between the transmitter and the receiver, the Rayleigh distribution is used to describe the channel fading amplitude. As such, the PDF of β is given by [Nakagami, 1960, Eq. (1)]

$$f_\beta(\beta) = \frac{2\beta}{\Omega}\exp\left(-\frac{\beta^2}{\Omega}\right), \quad \beta \geq 0, \tag{1.1}$$

where $\Omega = \mathsf{E}\left[\beta^2\right]$ is the average fading power. Note that the Rayleigh fading amplitude is simply the envelope of a zero-mean complex-valued Gaussian process with independent and identically distributed (i.i.d.) real and imaginary parts. In presence of LOS component, the PDF is Rice and is given by [Nakagami, 1960, Eq. (5)]

$$f_\beta(\beta) = \frac{2\beta}{\Omega}\exp\left(-\frac{\beta^2+s^2}{\Omega}\right) I_0\left(\frac{2\beta s}{\Omega}\right), \quad \beta, s \geq 0, \tag{1.2}$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind [Gradshteyn and Ryzhik, 2000, Section 8.43]. The Rician K-factor—defined as the ratio of the specular (LOS) power component to the average power of the scattered components—is given as $K = s^2/\Omega$.[7] Different from the Rayleigh distribution, the Rice fading PDF is the envelope of a nonzero-mean complex-valued Gaussian process with i.i.d. real and imaginary parts.

[7]The PDF in (1.2) is also called Nakagami-n, where $n^2 = K$, by Simon and Alouini [2005].

Thus, if K is zero—meaning that there is no LOS component—the function $I_0(0) = 1$ and the PDF in (1.2) is reduced to the one in (1.1).

Equations (1.1) and (1.2) can be respectively calculated and well approximated by means of the Nakagami-m PDF [Nakagami, 1960, Eq. (3)]:

$$f_\beta(\beta) = \frac{2m^m}{\Omega^m \Gamma(m)} \beta^{2m-1} \exp\left(-\frac{m\beta^2}{\Omega}\right), \quad \beta \geq 0,\ m \geq \frac{1}{2}, \tag{1.3}$$

by setting $m = 1$ and $m = (1+K)^2/(1+2K)$ for $K \geq 0$, respectively. Here, $\Gamma(\cdot)$ is the gamma function [Gradshteyn and Ryzhik, 2000, Eq. (8.310-1)]. Note that the parameter m spans the interval $[1/2, \infty\rangle$; corresponding to the one-sided Gaussian distribution for $m = 1/2$ and the Gaussian distribution for $m \to \infty$.

Defining the instantaneous CSNR as

$$\gamma = \frac{\beta^2 \mathcal{E}}{N_0} \tag{1.4}$$

where $\mathcal{E}$ is the average transmit power and N_0 is the noise variance, the average CSNR is

$$\bar{\gamma} = \frac{\mathsf{E}\left[\beta^2\right] \mathcal{E}}{N_0} = \frac{\Omega \mathcal{E}}{N_0}. \tag{1.5}$$

It can be seen in (1.4) that $\mathcal{E}/N_0$ is a constant and, thus, the distribution of β determines the distribution of γ and vice versa. Applying transformation of RVs (see e.g. [Papoulis and Pillai, 2002, Chap. 7]) to (1.1) we obtain the PDF of the CSNR as

$$f_\gamma(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right) \tag{1.6}$$

for Rayleigh fading. This is an exponential distribution. Recognizing that exponential distribution is a special case of the gamma distribution, and since the gamma distribution is frequently used later on in this thesis, the following definition will be useful:

Definition 1 (The gamma distribution)
X is said to follow a gamma distribution with shape parameter $\psi > 0$ and scale parameter $\chi > 0$ if the PDF of X is given by

$$f_X(x) = \frac{x^{\psi-1}}{\Gamma(\psi)\chi^\psi} \exp\left(-\frac{x}{\chi}\right), \ x \geq 0. \tag{1.7}$$

The short hand notation $X \sim \mathcal{G}(\psi, \chi)$ is used to denote that X follows a gamma distribution with shape factor ψ and scale factor χ. Note that the mean and variance of X is $\bar{x} = \mathsf{E}[X] = \psi\chi$ and $\sigma_x^2 = \mathsf{E}\left[X^2\right] - \bar{x}^2 = \psi\chi^2$, respectively.

Clearly, by comparing (1.6) and (1.7), the exponential distribution is a special case of a gamma distribution with $\psi = 1$, i.e. $\gamma \sim \mathcal{G}(1, \bar{\gamma})$. In this thesis, we will only consider a Rayleigh fading model for the channel. The PDF of γ in the Nakagami-m case will however still be useful since a Rayleigh channel with m antennas can be viewed as a Nakagami-m channel for certain antenna combining methods (cf. Section 1.3.1).

Finally, is noted that the expression in (1.6) corresponds to the case when we only have one antenna on each side of the transmission link, i.e. a SISO system. PDFs for multiantenna diversity-based systems with Rayleigh fading are given in the following subsection.

1.3.1 Effects of Space Diversity and Diversity Combining

The key concept of diversity in general is to create a number of more or less independent transmission "paths", all carrying the *same* information. In such a scenario, different signal paths may undergo independent channel fadings and, thus, independent fading statistics. As such, the probability of having all of them in a deep fade at the same time is small. Various types of diversity techniques together with different basic combining schemes are described in [Jakes, 1994]. In this thesis, we will only consider space (antenna) diversity and, hence, other techniques will not be mentioned.

Diversity combining is different from another popular and important antenna processing technique, beamforming, where the phase of signals from different antenna elements are adjust to point a beam in a desired direction. In the diversity combining technique, the signals are combined to increase the output signal level without affecting the individual antenna pattern. On the other hand, the beamforming technique exploit the differential phase between different antennas to modify the antenna pattern of the whole array. In that way, the whole of the array will have a single antenna pattern once they are combined [Godara, 1997b]. Beamforming is analyzed in e.g. [Godara, 1997a; Jafar and Goldsmith, 2001; Jafar *et al.*, 2001; Zhou and Giannakis, 2002, 2004a] and is beyond of the scope of this thesis.

It is well known that space diversity effectively averages out deep fades and mitigates considerably the effects of imperfect channel prediction/ estimation which, again, helps in improving system performance. This motivates the use of multiple-antenna reception at the receiver. In this case, we have a SIMO system. Furthermore, if the signals received at different an-

tennas are assumed uncorrelated[8] and maximum ratio combining (MRC)[9] is used to combine the receive signals, the combined CSNR is $\gamma = \sum_{b=1}^{n_R} \gamma_b$ where γ_b is the CSNR on *one receive branch* defined in (1.4). Thus, the average CSNR is $\bar{\gamma} = n_R \bar{\gamma}_b$ and the PDF of the total CSNR is [Simon and Alouini, 2005]

$$f_\gamma(\gamma) = \frac{\gamma^{n_R-1}}{\Gamma(n_R)} \left(\frac{1}{\bar{\gamma}_b}\right)^{n_R} \exp\left(-\frac{\gamma}{\bar{\gamma}_b}\right) \tag{1.8}$$

where n_R is the number of receive antennas. Hence, $\gamma \sim \mathcal{G}(n_R, \bar{\gamma}_b)$, i.e. the channel is effectively turned into a Nakagami-n_R channel. Note that we use the notation $\bar{\gamma}_b$ to denote the average CSNR on one receive branch even though the branches are independent and have the same mean CSNR.

In addition, we may have multiple antennas to transmit the *same* data from. In combination with multi reception we then have a MIMO diversity system where, space-time coding must be used to exploit and to achieve the available spatial diversity at the transmitter [Alamouti, 1998; Tarokh, Jafarkhami, and Calderbank, 1999].

Now, let $\bar{\gamma}_{ab} = \Omega\mathcal{E}/(n_T N_0)$ be the average CSNR received in one transmitter–receiver subchannel (when the same power is allocated to different transmit antennas n_T) then $\bar{\gamma}_b = n_T \bar{\gamma}_{ab}$ and $\bar{\gamma} = n_R \bar{\gamma}_b$. Again, when assuming independence between the subchannels and using space-time decoding [Alamouti, 1998],[10] the PDF of the effective CSNR is [Holter, Øien, Hole, and Holm, 2003; Ko and Tepedelenlioglu, 2003]

$$f_\gamma(\gamma) = \frac{\gamma^{n_T n_R-1}}{\Gamma(n_T n_R)} \left(\frac{n_T}{\bar{\gamma}_b}\right)^{n_T n_R} \exp\left(-\frac{\gamma n_T}{\bar{\gamma}_b}\right), \tag{1.9}$$

i.e. $\gamma \sim \mathcal{G}(n_T n_R, \bar{\gamma}_b/n_T)$. Clearly a MISO system is a special case of a MIMO diversity system where $n_R = 1$. In fact, both Eqs. (1.6) and (1.8) can be deduced from the above equation.

In Figure 1.2 we see both the PDF and the cumulative distribution function (CDF) of the received CSNR as functions of normalized CSNR plotted

[8] This can be obtained by sufficient separation of the antenna elements. About half a wavelength is theoretically sufficient in an isotropic scattering environment [Rappaport, 2002].

[9] When the channel is perfectly known, MRC is the theoretical optimal combining scheme when the branches are uncorrelated [Brennan, 2003]. It is still valid when there is correlation between them [Dong and Beaulieu, 2002; Loskot and Beaulieu, 2004]. In practice the channel has to be estimated, which gives rise to imperfect knowledge of the channel. In particular, the channel estimation should be treated as an additional source of noise which might not be white. In that case, the optimal combining scheme can be found in [You, Li, and Bar-Ness, 2005].

[10] Although space-time decoding is slightly different from MRC, it is noted that the effective CSNR is still the same and the distribution using MRC is also applicable here.

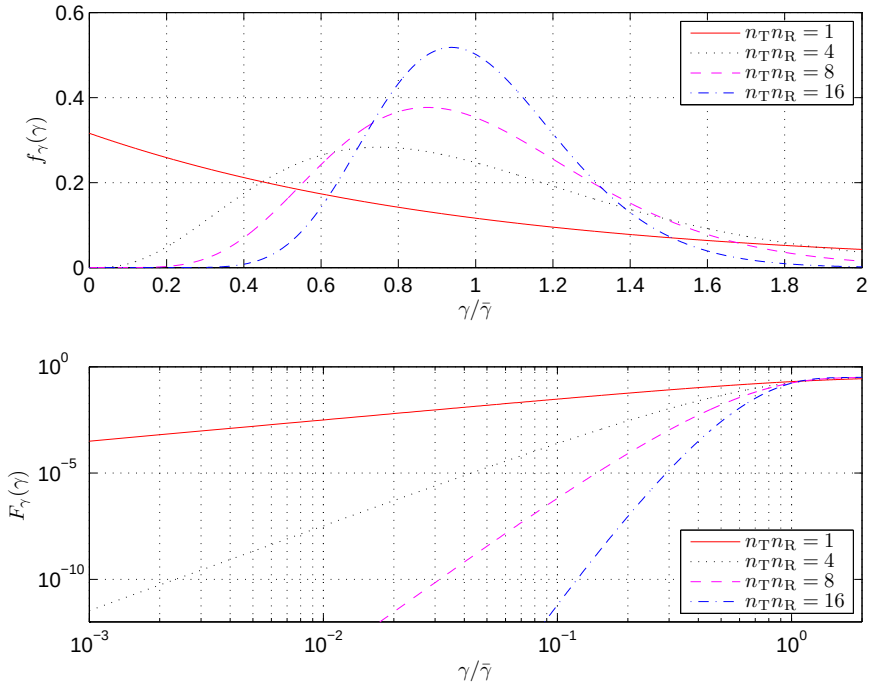


FIGURE 1.2: PDF (top panel) and CDF (bottom panel) of the received CSNR as a function of normalized CSNR. It is plotted for different combinations of transmit-receive antennas using (1.9) when $\bar{\gamma} = n_R \gamma_b = 5$ dB.

for $\bar{\gamma} = 5$ dB. It can be observed that the CSNR variations become smaller when the product $n_T n_R$ is increased. When $n_T n_R$ goes to infinity, the fading channel approaches the additive white Gaussian noise (AWGN) channel since, in this case, the PDF of the CSNR becomes a delta pulse. Thus, spatial diversity is very effective in making the channel *look* like a Gaussian channel [Meyr *et al.*, 1998].

The expressions (1.8) and (1.9) are valid only for spatially uncorrelated antenna branches. When spatial correlation exists in a SIMO model and using MRC at the receiver, the total CSNR is a sum of correlated gamma variates and can not be modelled by a gamma PDF. In this case, the exact PDF of γ in the dual branch case may be written as a type I McKay distribution [Holm and Alouini, 2004, Eq. (20)]. For more than 2 receive antennas, a closed-form expression is given in [Aalo, 1995, Eq. (18)], but only for constant correlation and identically distributed branches. For an arbitrary correlation model and a number of antennas greater than 2, an infinite sum expression is derived and presented in [Alouini, Abdi, and Kaveh,

2001, Eq. (5)]. In the general case of independently distributed Rayleigh branches, it can happen that some of the branches have the same mean and the other remaining branches have different means. In such a case, the PDF of γ is given in [Ho Van Khuong and Kong, 2006].

1.4 Outline

The thesis is largely built on a collection of published/submitted papers, except from Chapter 2 where the common system aspects, the system parameters, and the performance metrics used in the rest of the thesis is introduced. Which papers the materials are taken from are clearly pointed out in the beginning of every chapter. Information of where the papers were published or submitted is also given there. The rest of the thesis is organized as follows.

- *Chapter 2* introduces the preliminaries of our rate-adaptive transmission system. The quality metrics in terms of average spectral efficiency (ASE) and bit-error-rate (BER) are introduced. General expressions for optimal channel estimators and predictors are re-developed and illustrations of the predictor performance are presented. The system parameters used throughout the thesis are also given.

- *Chapter 3* focuses on optimizing the frequency of pilot insertion and the power usage on these pilots in an adaptive transmission system operating on a SISO Rayleigh fading channel and based on PSAM channel monitoring. Trellis codes designed for Gaussian channels are used as component codes. Both the estimation error at the receiver and prediction error at the transmitter are taken into account when performing adaptation. Also here, the optimization algorithm is described. Since this optimization algorithm is used throughout (in every chapter) in the thesis, it is given as a general algorithm.

- *Chapter 4* extends the results in Chapter 3 to a spatially uncorrelated SIMO Rayleigh fading system and presents the optimized results of pilot spacing and power allocation on these pilots. The average time the system remains in one mode before switching to the different mode is also given for both SISO and SIMO cases. Comparison to another system—where estimation is assumed to be perfect and the switching thresholds between different constellations are fixed regardless of average channel quality—is made.

- *Chapter 5* studies a SIMO system operating on identically distributed but correlated Rayleigh fading channels. In this chapter, the receiver estimation is assumed perfect. The chapter is divided into two parts. The first part consists of analyzing the system degradation when the system originally developed for partially uncorrelated subchannels is employed in spatially correlated subchannels: that is, the subchannels are predicted independently of each other and the spatial correlation is not taken into account. In the second part, the subchannels are jointly predicted. Thus the spatial correlation is assumed known so that it can be incorporated in the prediction process. Thus, a "space-time" predictor is needed and derived.

- *Chapter 6* adds multiple transmit antennas to the uncorrelated SIMO case. That is, we have a MIMO diversity system, where the branches are assumed uncorrelated. As in other cases (SISO and SIMO), the feedback is predicted CSNR-based. Thus, without knowing the channel matrix, the transmit power is divided equally between the transmit antennas. This is not necessarily optimal, but based on our feedback it is the only choice we have. Orthogonal space-time block codes are employed to exploit the space diversity at the transmitter and, since the branches are assumed uncorrelated, the diversity order is the product of number of transmit and receive antennas.

- *Chapter 7* demonstrates the system performance under certain simplifications, such as the pilot spacing being fixed to certain values and the predictor order being reduced to a lower one. The approach is contrasted to that of previous chapters where the pilot spacing is selected optimally according to the average CSNR to maximize the ASE, and the length of the predictor is kept large to assure reliable channel prediction. The simplifications enable good trade-offs in practical implementation where it is desired to keep the complexity low. The impact of the widely assumed simplification on perfect channel estimation is also demonstrated.

- *Chapter 8* summarizes the main results and contributions of the thesis and gives an outlook on some further research topics that might be of interest.

- *Appendices A*, *B*, C, and D clarify some mathematical expressions and calculations which are used earlier in the thesis.

Chapter 2

Preliminaries of Our Adaptive Coded Modulation Systems

In this chapter, we introduce the common system aspects and performance metrics used in the rest of the thesis. The chapter starts with Section 2.1 where a summary of previous works and a motivation for our system are given. We will visualize how pilot symbols are multiplexed into the data stream and highlight some particularly important issues in our ACM systems in Section 2.2. Furthermore, we will state the two major performance metrics; namely ASE and average BER performances in Sections 2.3 and 2.4, respectively. In Section 2.5, how transmit power allocation to pilot and data symbols is discussed. We then re-develop the MAP-optimal noncausal estimator and causal predictor with the corresponding minimum mean square errors (MMSE) in Section 2.6. Also, illustrations together with the MSE performances of the predictor are given. An introduction to the common parameters used in the numerical examples throughout the thesis is given in Section 2.7.

2.1 Background

Before going on to describe our systems, we recapitulate some important previous works in order to motivate for our work. In [Goldsmith and Varaiya, 1997], the capacity of a single-user adaptive system communicating over flat-fading channels with arbitrary fading distributions has been derived in terms of the PDF of the fading gain when perfect CSI is available at both transmitter and receiver. Furthermore, Alouini and Goldsmith give closed-form expressions for the capacity of different diversity combining techniques for Rayleigh fading in their 1997 paper and for the general case

of Nakagami-m fading in their 1999 paper. In those papers, the capacity of a channel can be attained using a certain rate/power-adaptation scheme in which the information bits is continuously updated according to the channel quality at any given time. They have also shown that the capacity of the rate-adaptive system with constant transmit power is very close to the one with both rate and power optimally selected.[1]

Average spectral efficiency (ASE) is a measure of, on average, how many bits are transmitted per second per unit bandwidth [bits/s/Hz] with a given transmission scheme. In [Goldsmith and Chua, 1997], the ASE of a more practical discrete-rate multilevel QAM (M-QAM) is derived. The results show that there is a gap between the obtainable ASE and the capacity in Rayleigh fading. To reduce this gap, Goldsmith and Chua [1998] analyzed a trellis coded modulation (TCM) scheme with both rate and power adaptation, where the desired minimum distance of different codes is kept constant. Their adaptive TCM scheme utilizes a set of 2-dimensional (2-D) trellis codes with different spectral efficiencies (SEs) to achieve an additional coding gain relative to the work of uncoded adaptive M-QAM in their 1997 paper. Moreover, by adaptively changing the transmit power, in both continuous and discrete manner, it has been shown in [Gjendemsjø *et al.*, 2005; Gjendemsjø, Øien, and Orten, 2006] that the maximum ASE approaches the capacity even with only a few capacity-achieving codes.

Based on (among other) the work in [Goldsmith and Chua, 1998], Hole *et al.* [2000] were the first ones who designed a rate-adaptive system using an adaptive codec with eight 4-dimensional (4-D) trellis codes utilizing eight nested 2-D M-QAM signal constellations (depicted in Figure 2.1) with $M \in \{M_n\}_{n=1}^{N=8}$ being in the set $\{4, 8, 16, 32, 64, 128, 256, 512\}$. The SE of these codes is given as[2]

$$R_n = \log_2(M_n) - \frac{1}{2} \text{ [bits/s/Hz]}. \tag{2.1}$$

The 4-D TCM is employed in e.g. [Holm, 2002; Øien *et al.*, 2004; Holter, 2005] and in most of the publications made by these authors in the field of ACM.

In all of the above works as well as in most classical papers on adaptive transmission, the assumptions are either instant feedback at both transmitter and receiver, or perfect CSI at the receiver and imperfect CSI available to the transmitter with a certain delay. Inspired by the work on optimizing an

[1] However, it has later been shown that this is not the case when the number of rates is finite and small [Gjendemsjø, Øien, and Holm, 2005].

[2] Actually, the SE is given as $R_n = \log_2(M_n) - 1/G$ for $2G$-D trellis codes. However, since they are using 4-D trellis codes, G is equal to 2.

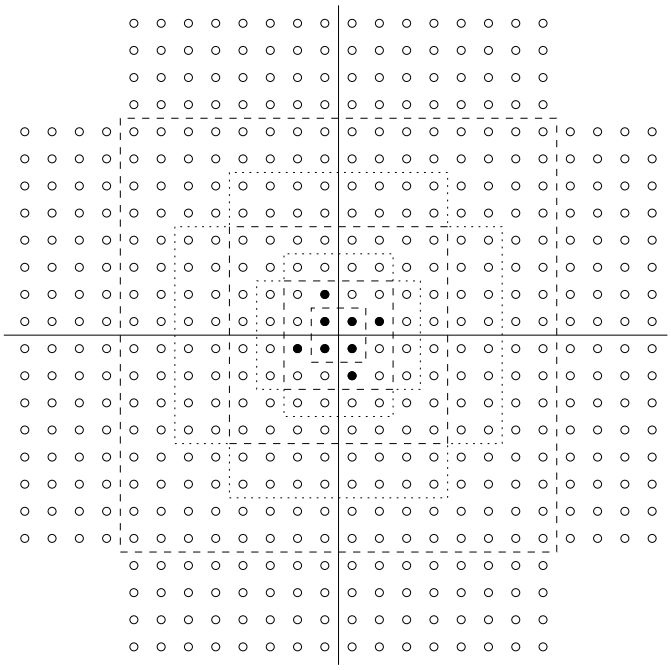


FIGURE 2.1: Nesting of the 4-QAM, 8-STAR, 16-QAM, 32-CROSS, 64-QAM, 128-CROSS, 256-QAM, and 512-CROSS signal constellations. The filled black circles constitute the 8-STAR constellation.

uncoded adaptive system for a SISO channel presented by Cai and Giannakis [2005]—where both channel estimation at the receiver and prediction at the transmitter are incorporated—we, on the other hand, will analyze and optimize an ACM system operating on any combination of number of transmit and receive antennas (XIXO). Also, we will consider non-zero feedback delay. Furthermore, we will keep the average transmit power constant, but the power distribution between pilot and data symbols will be optimized. The pilot spacing is also to be optimally selected.

2.2 System Description

The generic block diagram describing all the systems considered in this thesis is shown in Figure 1.1 where the fading channel is assumed to be slowly varying. This means that the channel is assumed to be more or less constant over many channel symbols. In our realization of this generic system, the adaptive encoder generates coded bits using 4-D trellis codes. The coded bits then determine the transmittable symbol from the proper

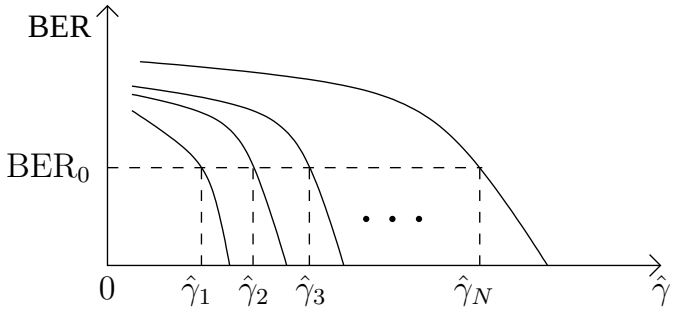


FIGURE 2.2: Instantaneous BER with respect to predicted CSNR along with the switching thresholds for different constellations.

QAM constellation out of the set of $N = 8$ constellations of sizes $M_n = \{4, 8, 16, 32, 64, 128, 256, 512\}$ that is best suited to the predicted channel state. The set of constellations corresponds to a set of SEs $\{R_n\}_{n=1}^{N}$. We have $M_n < M_{n+1}\ \forall n$ and, thus, $R_n < R_{n+1}$. The choice of code is based on the CSI fed back from the receiver: code n is chosen if the overall *predicted* CSNR falls between the switching thresholds $\hat{\gamma}_n$ and $\hat{\gamma}_{n+1}$. The instantaneous predicted CSNR is therefore quantized into $N + 1$ zones, where each quantization zone corresponds to a specific constellation M_n. By letting $\hat{\gamma}_0 = 0$ and $\hat{\gamma}_{N+1} = \infty$ we have $\hat{\gamma}_n < \hat{\gamma}_{n+1}$ for all $n \in \{0, 1, \cdots, N\}$. Since the adaptivity is based on predicted CSNR, the set of thresholds $\{\hat{\gamma}_n\}_{n=1}^{N}$ are determined such that the instantaneous BER (with respect to predicted CSNR) of the selected code n is below BER$_0$; as illustrated in Figure 2.2.

Sometimes in ACM systems the channel is predicted to be so bad that reliable transmission can not be guaranteed and the system goes into an outage (or "no transmission") state. That is when there is no code in the available code set which maintains the desired BER performance. In our case, it occurs when the predicted CSNR is below $\hat{\gamma}_1$. In that case, as long as the predicted CSNR is not greater than $\hat{\gamma}_1$, the system sends nothing but the pilots—in order to keep track of the channel variations—and the data is buffered at the transmitter.[3]

Using a PSAM-based channel monitoring scheme, pilot symbols with a certain power are inserted before everything is launched on the transmit antenna. In the MIMO and MISO diversity case, a pilot symbol is transmitted from each antenna one at a time, and the data symbols are space-time block coded before they are transmitted simultaneously from all of the antennas. For SISO and SIMO cases, similarly to [Torrance and Hanzo, 1995;

[3] An alternative approach would be to keep transmitting and subsequently have to retransmit the outage bits, but this is not considered here.

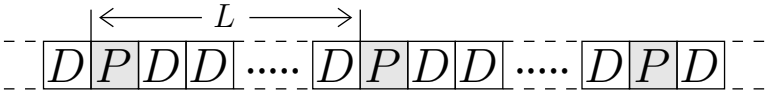


FIGURE 2.3: Transmitted frame structure of the datastream for SISO and SIMO system configuration. Here, P and D denote pilot and data symbol, respectively. The frame structure in the MIMO diversity configuration is given in Figure 6.2.

Tang, Alouini, and Goldsmith, 1999], the datastream has a frame structure as illustrated in Figure 2.3, where each frame starts with a pilot symbol. Due to the fact that every Lth channel symbol is a pilot symbol that does not convey data information, the SE of different constellations becomes [Øien *et al.*, 2004]

$$R_n = \frac{L-1}{L}\left(\log_2(M_n) - \frac{1}{2}\right) \quad \text{[bits/s/Hz]}. \tag{2.2}$$

At the receiver, the pilot symbols are extracted and used for both channel estimation and prediction. Based on the estimated channel, the signal is detected and adaptively decoded. In case of multireception, the signals from different antennas are combined and then decoded. Space-time combining and decoding are performed in MIMO and MISO diversity system, while MRC is used in SIMO systems. Throughout the thesis, we assume coherent combining, meaning that the absolute phase of the incoming signals are detected and compensated for. Thus, we only concentrate on combining the fading gains. In all the XIXO cases, the subchannels are assumed to follow a Rayleigh distribution and for the numerical examples they are assumed to have a Jakes temporal correlation profile. Most of the time, we assume the subchannels to be mutually independent. However, we will also consider SIMO systems with spatial correlation between the branches at the receiver. In this case, we consider receiver channel estimation as perfect, as opposed to the rest of the thesis.

Finally, to have a low feedback rate, the CSI fed back to the transmitter is only the overall predicted CSNR. This information is sent over the feedback channel. The feedback channel is assumed to be error-free, but with non-zero delay.

2.3 ASE Performance

Let P_n be the probability that code n is being used. Since the predicted CSNR determines which code will be activated to transmit, P_n is obtained

by integrating the PDF of the predicted CSNR over the interval $[\hat{\gamma}_n, \hat{\gamma}_{n+1}\rangle$. That is,

$$P_n = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma}. \tag{2.3}$$

It is then obvious that the probability of outage (no transmission) is

$$P_{\text{out}} = \int_{0}^{\hat{\gamma}_1} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma}. \tag{2.4}$$

The ASE of the ACM systems encountered in this thesis is thus given as [Ue *et al.*, 1998; Goldsmith and Chua, 1997]

$$\text{ASE} = \sum_{n=1}^{N} R_n \cdot P_n \quad [\text{bits/s/Hz}]. \tag{2.5}$$

where R_n is the SE given by (2.2).

2.4 BER Performance

The overall average BER over all codes, denoted by $\overline{\text{BER}}$, is the ratio between the average number of bits in error, and the number of bits transmitted in total [Ue *et al.*, 1998; Alouini and Goldsmith, 2000; Chung and Goldsmith, 2001]:[4]

$$\overline{\text{BER}} = \frac{\sum_{n=1}^{N} R_n \cdot \text{BER}(M_n)}{\sum_{n=1}^{N} R_n \cdot P_n}, \tag{2.6}$$

where $\text{BER}(M_n)$ is the average BER corresponding to code n.

Applying the 4-D trellis codes mentioned in Section 2.3 over AWGN channels, the BER performance is approximated by

$$\text{BER}(M_n|\gamma) = \begin{cases} a_n \exp\left(-\frac{b_n}{M_n}\gamma\right) & \text{when } \gamma \geq \gamma_{n,T} \\ \frac{1}{2} & \text{when } \gamma < \gamma_{n,T}. \end{cases} \tag{2.7}$$

The BER expression above is obtained by first simulating the BER performance of the 4-D trellis codes on AWGN channels and then using curve

[4] It is noted that [Chung and Goldsmith, 2001] gives two possible definitions for the average BER: 1) $\overline{\text{BER}} = \mathsf{E}[\#\text{ error bits per transmission}] / \mathsf{E}[\#\text{ bits per transmission}]$, and 2) $\overline{\text{BER}} = \mathsf{E}[(\#\text{ error bits per transmission})/(\#\text{ bits per transmission})]$ and argue that the former definition is slightly better than the latter for an ergodic fading process. Moreover, Alouini and Goldsmith [2000] claim that the average BER can be computed *exactly* with the former definition for an adaptive discrete rate system.

fitting to find the exponential expression which fits to the simulated values. Thus, a_n and b_n are code-dependent constants found by curve fitting [Hole *et al.*, 2000, Table I]. Moreover, $\gamma_{n,T} = M_n \ln(2a_n)/b_n$ is the CSNR threshold where the BER expression goes from $1/2$ to the exponential approximation. At high CSNR, the exponential part of (2.7) represents a very good approximation. This BER approximation has been used in, e.g., [Øien, Holm, and Hole, 2002b; Holm, 2002; Holter, 2005] and most of the publications of these authors within the field of ACM. For a graphical illustration of how good the approximation is, have a look into, e.g., [Hole *et al.*, 2000, Fig. 2] or [Øien *et al.*, 2004, Fig. 3].

Whilst the approximation (2.7) is tight, the analysis using this would result in solving more than one integral, and sometimes these integrals can not be solved in closed-form, especially in spatial diversity systems. Therefore, we choose to use a somewhat looser approximation which is similarly to [Falahati, Svensson, Sternad, and Mei, 2003]

$$\mathrm{BER}(M_n|\gamma) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \exp\left(-\gamma \frac{b_n(\ell)}{M_n}\right), \tag{2.8}$$

where $\mathcal{L}$ is the number of exponential functions which approximate the simulated BER (here we use $\mathcal{L} = 3$), and $a_n(\ell)$ and $b_n(\ell)$ are constellation dependent constants obtained by first simulating the codes' BER performance and then use curve fitting with the least square method. These constants are given in Tab. 2.1 and the approximation is illustrated in Figure 2.4 for different constellation sizes. At first sight, the approximation seem to be quite coarse. However, it will be shown later that the ACM performance results of this approximation are very close to those obtained when the tight approximation (2.7) is used.

2.5 Transmit Powers

In this section, the notations for pilot and data symbol powers are introduced. Their interrelation is discussed, under the average power constraint used throughout the thesis.

Let $\mathbb{C}$ be the set of complex numbers and let y, h, and $n \in \mathbb{C}$. In a narrowband system, the noisy and faded input-output relationship for a pilot symbol of any branch can be written as

$$y_{\mathrm{pl}}(k;0) = \sqrt{\mathcal{E}_{\mathrm{pl}}} h(k;0) s(k;0) + n(k;0), \tag{2.9}$$

and for the data symbols as

$$y_d(k;l) = \sqrt{\mathcal{E}_d} h(k;l) s(k;l) + n(k;l), \quad l = [1, \cdots, L-1]. \tag{2.10}$$

TABLE 2.1: The code-dependent constants $\{a_n(\ell)\}_{\ell=1}^{3}$ and $\{b_n(\ell)\}_{\ell=1}^{3}$ for the example 4-D trellis codes.

n	$a_n(1)$	$b_n(1)$	$a_n(2)$	$b_n(2)$	$a_n(3)$	$b_n(3)$
1	233.8034	12.4335	-280.8712	11.4405	51.3394	8.6131
2	210.6415	8.4208	-242.0657	7.9916	34.3732	6.0432
3	246.0565	7.7677	-334.6900	8.1130	89.4924	9.1087
4	99.7887	7.7426	-160.6040	8.3843	61.5091	9.4667
5	78.0083	7.0135	-100.8414	7.4793	23.4319	9.0714
6	86.2181	7.3704	-96.0270	7.6780	10.3583	10.3191
7	87.6912	7.0471	-94.5020	7.2852	7.3344	10.1898
8	89.3099	7.2848	-95.6889	7.4987	6.8972	10.4428

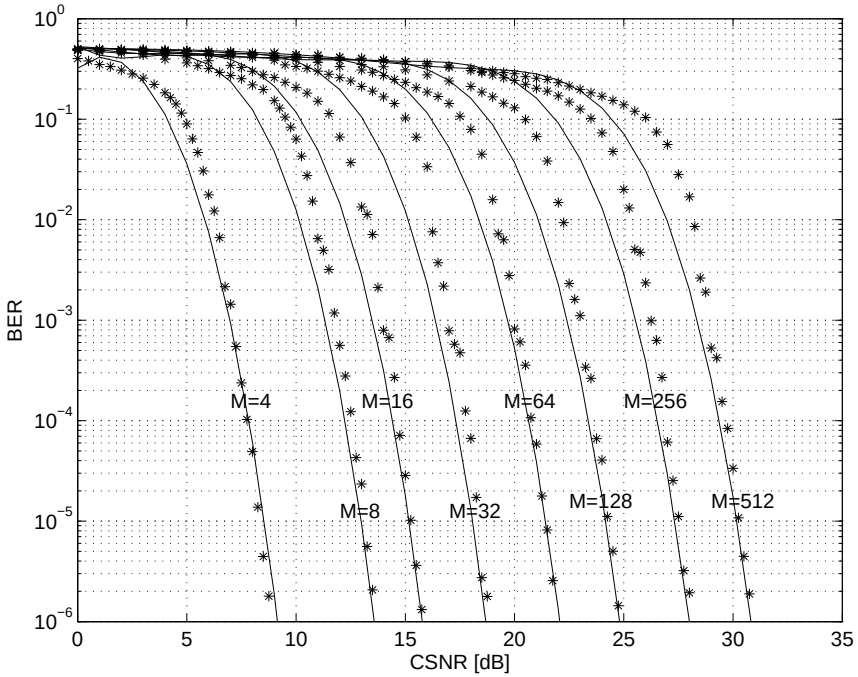


FIGURE 2.4: BER performance of TCM codes on AWGN channels for different M-QAM constellations. The solid lines denote the approximations, while the stars represent the simulated values.

Using the frame-based structure shown in Figure 2.3, the index k counts the frame and l is the symbol index in that frame; such that $y(k;l)$ is the compact way of writing $y(kLT_s + lT_s)$ where T_s is the channel symbol duration. The notations $\mathcal{E}_{\text{pl}}$ and $\mathcal{E}_d$ are used to denote the actual power per pilot symbol and the *actual* power per data symbol, respectively. A pilot is sent at the beginning of a frame denoted by $s(k;0)$ and $\{s(k;l)\}_{l=1}^{L-1}$ are data symbols in the kth frame. Furthermore, $n(\cdot,\cdot)$ denotes zero-mean complex-valued AWGN with variance $N_0/2$ per dimension and dimensions being uncorrelated. The fading gain $h(k;l)$ is assumed to be a stationary complex Gaussian random process with zero mean and variance $\sigma_h^2 = \Omega = 1$.

On average, each symbol (either pilot or data) has a transmit power of $\mathcal{E}$. The average data power and pilot power in a frame can be calculated as

$$\bar{\mathcal{E}}_d = \frac{\alpha L \mathcal{E}}{L-1}, \tag{2.11}$$

and

$$\mathcal{E}_{\text{pl}} = (1-\alpha) L \mathcal{E}, \tag{2.12}$$

respectively. In both equations, $\alpha \in \mathbb{R}^+$ within the interval $\langle 0,1 \rangle$ is the variable which determines how much power should be put on data symbols and on pilot symbols, respectively. Equal pilot and data power, i.e. $\mathcal{E}_{\text{pl}} = \bar{\mathcal{E}}_d = \mathcal{E}$, is obtained when $\alpha = 1 - 1/L$. Moreover, as mentioned in Section 2.3, no data is transmitted when outage occurs—i.e. when the predicted CSNR is below $\hat{\gamma}_1$—and, thus, the actual transmit data power in (2.10) can be set to

$$\mathcal{E}_d = \frac{\bar{\mathcal{E}}_d}{\text{probability of transmission}} = \frac{\bar{\mathcal{E}}_d}{\int_{\hat{\gamma}_1}^{\infty} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma}}. \tag{2.13}$$

2.6 MAP-Optimal Channel Estimators and Predictors

Here, we state the expressions used for our prediction and estimation filters, and give expressions for the corresponding prediction/estimation errors. We impose a wide-sense stationary (WSS) assumption for the channel and the analysis in this section is done assuming that the temporal correlation function of the channel is known.

2.6.1 Channel Estimator

It is noted that *truly optimal* channel estimation calls for noncausal lowpass filtering (smoothing) in the limit $K_e \to \infty$ [Meyr *et al.*, 1998, p. 660], where

K_e is the filter order. For practical purposes, K_e must be finite. The estimator uses a vector of K_e received pilot symbols (half from the past and half from the future) to estimate one sample. Since the channel is complex Gaussian, it is well known that the optimal estimated channel is a linear combination of the observations. Based on the received pilot symbols, the MAP-optimal estimator $\mathbf{w}_e$ obeys the normal equations [Therrien, 1992]

$$\mathbf{R}_y \mathbf{w}_e = \mathbf{r}_y \tag{2.14}$$

where $\mathbf{R}_y = \mathsf{E}\left[\mathbf{y}_{\text{pl}}\mathbf{y}_{\text{pl}}^{\mathsf{H}}\right]$ is the correlation matrix of the buffered receive pilot symbols $\mathbf{y}_{\text{pl}} = \left[y_{\text{pl}}(k - \lfloor K_e/2 \rfloor; 0), \cdots, y_{\text{pl}}(k + \lfloor (K_e - 1)/2 \rfloor; 0)\right]^{\mathsf{T}}$ (dimension $K_e \times 1$) and $\mathbf{r}_y = \mathsf{E}\left[\mathbf{y}_{\text{pl}} h^*(k; l)\right]$ is the covariance vector between the desired channel sample to be estimated and the vector of received pilot symbols. The notation $\lfloor x \rfloor$ represents the integer part of x, and $(\cdot)^{\mathsf{H}}$ and $(\cdot)^*$ denote conjugate transpose and complex conjugate, respectively.

Obviously, the pilot vector is $\mathbf{y}_{\text{pl}} = \sqrt{\mathcal{E}_{\text{pl}}}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{h}_{\text{pl}} + \mathbf{n}_{\text{pl}}$, where $\boldsymbol{\mathcal{D}}(\mathbf{s}) \in \mathbb{C}^{K_e \times K_e}$ is a diagonal matrix with the pilot vector $\mathbf{s}$ on its diagonal and $\mathbf{h}_{\text{pl}} = \left[h(k - \lfloor K_e/2 \rfloor; 0), \cdots, h(k + \lfloor (K_e - 1)/2 \rfloor; 0)\right]^{\mathsf{T}}$ is the corresponding channel gain vector. With the assumption of independence between the additive noise and the channel, it is easy and straightforward to show that

$$\mathbf{R}_y = \mathcal{E}_{\text{pl}}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}_e\boldsymbol{\mathcal{D}}^*(\mathbf{s}) + N_0\mathbf{I}_{K_e}. \tag{2.15}$$

Likewise, $\mathbf{r}_y$ can be written as

$$\mathbf{r}_y = \sqrt{\mathcal{E}_{\text{pl}}}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_e. \tag{2.16}$$

In Eqs. (2.15) and (2.16), $\mathbf{R}_e = \mathsf{E}\left[\mathbf{h}_{\text{pl}}\mathbf{h}_{\text{pl}}^{\mathsf{H}}\right]$ and $\mathbf{r}_e = \mathsf{E}\left[\mathbf{h}_{\text{pl}} h^*(k; l)\right]$ are the normalized covariance matrix and normalized covariance vector of the channel, respectively.[5] $\mathbf{I}_{K_e}$ is the $K_e \times K_e$ identity matrix.

For the Jakes fading model assumed in the numerical examples, the elements of the correlation matrix and correlation vector are real and described by the correlation function:

$$R(m) = J_0(2\pi f_d T_s |m|), \tag{2.17}$$

[5] We use the notations $\mathbf{R}_e$ and $\mathbf{r}_e$ to denote the covariance matrix and covariance vector of the channel for *estimation* case. This is to distinguish it from $\mathbf{R}_p$ and $\mathbf{r}_p$ for the *prediction* case. As a rule of thumb, the variables with a subscript e have something to do with estimation. Likewise, the variables with a subscript p have something to do with prediction.

where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, f_d is the maximum Doppler frequency, m is the lag between the points we want to find the correlation of, and as earlier T_s is the channel symbol duration.

Upon solving (2.14) with respect to $\mathbf{w}_e$ and replacing $\mathbf{R}_y$ and $\mathbf{r}_y$ in the solution with (2.15) and (2.16) we get the optimal linear estimator [Cai and Giannakis, 2005; Meyr *et al.*, 1998, Eq. (14-35)]

$$\mathbf{w}_e = \sqrt{\mathcal{E}_{\text{pl}}}\Big(\mathcal{E}_{\text{pl}}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}_e\boldsymbol{\mathcal{D}}^*(\mathbf{s}) + N_0\mathbf{I}_{K_e}\Big)^{-1}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_e. \tag{2.18}$$

The MMSE of the estimation error is then

$$\sigma_e^2(l) = 1 - \sqrt{\mathcal{E}_{\text{pl}}}\mathbf{r}_e^{\mathsf{H}}\boldsymbol{\mathcal{D}}^*(\mathbf{s})\mathbf{w}_e. \tag{2.19}$$

We note that the MMSE only depends on the estimation lag l (and not on which frame k we are estimating). This is due to the stationarity of the channel, which we have assumed. For notational brevity, we sometimes omit such the time indices wherever it is possible.

2.6.2 Channel Predictor

We assume that the transmitter adaptation occurs no more than once per transmission frame. Thus, the system feedback delay[6] can be set to $\tau = DLT_s$, where D is a positive integer. Since we need to predict the channel ahead in time, the predictor must be causal. The *causal* predictor uses K_p pilot symbols from the past to predict one sample in the set $\{h(k;l)\}_{l=1}^{L-1}$ in the kth frame which is τ seconds ahead in time.

Here, the causal pilot vector buffered at the receiver is

$$\mathbf{y}_{\text{pl}} = \big[y_{\text{pl}}(k-D;0), \cdots, y_{\text{pl}}(k-D-K_p+1;0)\big]^{\mathsf{T}} \in \mathbb{C}^{K_p\times 1}.$$

Based on this vector the correlation matrix $\mathbf{R}_y$ and correlation vector $\mathbf{r}_y$ can be derived. Similarly to the estimation case, the optimal linear predictor is

$$\mathbf{w}_p = \sqrt{\mathcal{E}_{\text{pl}}}\Big(\mathcal{E}_{\text{pl}}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}_p\boldsymbol{\mathcal{D}}^*(\mathbf{s}) + N_0\mathbf{I}_{K_p}\Big)^{-1}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_p. \tag{2.20}$$

The MMSE is equal to

$$\sigma_p^2(l) = 1 - \sqrt{\mathcal{E}_{\text{pl}}}\mathbf{r}_p^{\mathsf{H}}\boldsymbol{\mathcal{D}}^*(\mathbf{s})\mathbf{w}_p. \tag{2.21}$$

[6]The feedback delay here includes both the time needed to perform prediction, the physical delay on the return channel, and the processing time needed by the transmitter to activate and configure itself to the code to be transmitted.

Note that $\mathbf{R}_p$ and $\mathbf{r}_p$ are the correlation matrix and the correlation vector for the prediction case. Furthermore, it should be pointed out that we use the same notation, $\mathbf{y}_{\text{pl}}$, to denote the vector of pilot symbols the estimation and the prediction are based on. Basically, the two vectors are different and they also are different in size. However, it should be clear from the context which case the vector yields.

Finally, we note that the above analysis puts no constraints on the pilot symbols. It means that different pilot symbols can in principle be transmitted in different pilot time instances. When the same pilot symbol is always transmitted, the optimal estimator/predictor and the corresponding MMSE expressions can be expressed in a simpler way. Moreover, they are deduced as special cases of the "space-time" estimator/predictor in Appendix A.

2.6.3 Visualization of MAP-optimal Prediction

Performance of the MAP-optimal predictor was studied in [Duong, 2002; Øien, Hansen, Duong, Holm, and Hole, 2002] when both the pilot and data symbols have equal power. A Rayleigh fading channel with Jakes spectrum correlation profile was generated and the derived MAP-optimal predictor was used to track that channel. The system under study in the two references has a symbol duration of $T_s = 2.5$ μs and a carrier frequency $f_c = 5$ GHz. The Doppler frequency is $f_d = 250$ Hz—corresponding to a terminal velocity of $v = 15$ m/s (or 54 km/h). Figure 2.5 demonstrates how the fading envelope is tracked, while Figures 2.6 and 2.7 show the MSE performance of the prediction error.

In Figure 2.5, it is shown, for certain parameter values, how such an optimal predictor works on a Rayleigh fading channel. Apparently, upon studying the curves, the tracking of the fading amplitude degrades somewhat when we predict 5 pilot symbols (200 channel symbols) instead of 1 pilot symbol (40 channel symbols) ahead in time. Equivalently, it is the same as predicting 0.5 ms and 0.1 ms ahead in time, respectively. This can be attributed to the decreasing fading correlation as the prediction horizon increases.

In [Holm, 2002; Øien *et al.*, 2004] a predictor of 1000 taps was used, something that is clearly too complex for most practical situations. The impact of lowering this order, when both pilot and data symbols have the same transmit power, and L is fixed, has been discussed in [Duong, 2002]. The graphs in Figure 2.6 are results of predicting 5 pilot symbols ahead in time at 5 dB average CSNR, and when the pilot spacing L is taken on some discrete values between 66 and 10 (corresponding to pilot sampling

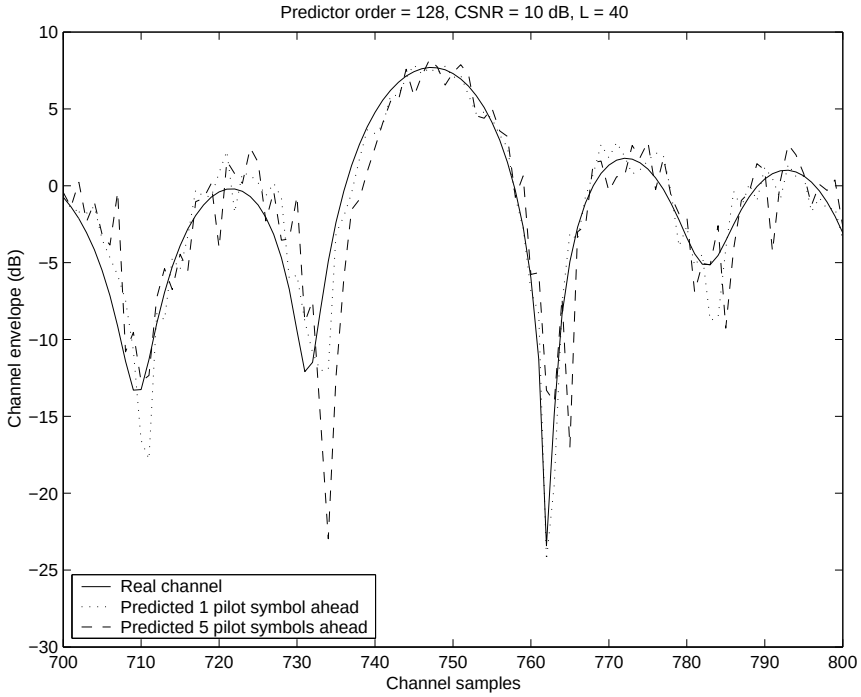


FIGURE 2.5: Predicted and true fading envelope for prediction of 1 and 5 pilot symbols ahead in a SISO system. Carrier frequency $f_c = 5$ GHz; Doppler frequency $f_d = 250$ Hz; predictor order $K_p = 128$; pilot spacing $L = 40$; and average CSNR $\bar{\gamma} = 10$ dB.

frequency between $f_s = 6$ kHz and 40 kHz). From a MSE performance point of view, the channel predictor order can be shortened to a reasonable value without affecting the MSE since it saturates and stabilizes when the order is high enough.

To finalize this subsection, we include a 3-dimensional (3-D) plot of MSE performance as a function of both expected CSNR and the prediction horizon. As shown in Figure 2.7, a predictor with 128 taps in combination with a pilot spacing $L = 40$ is adequate to achieve stable MSE performance, as long as the expected CSNR is not too low or the prediction horizon is not too long. However, when BER performance is used to measure the system quality, the requirement of predictor order K_p and pilot spacing L might be changed.

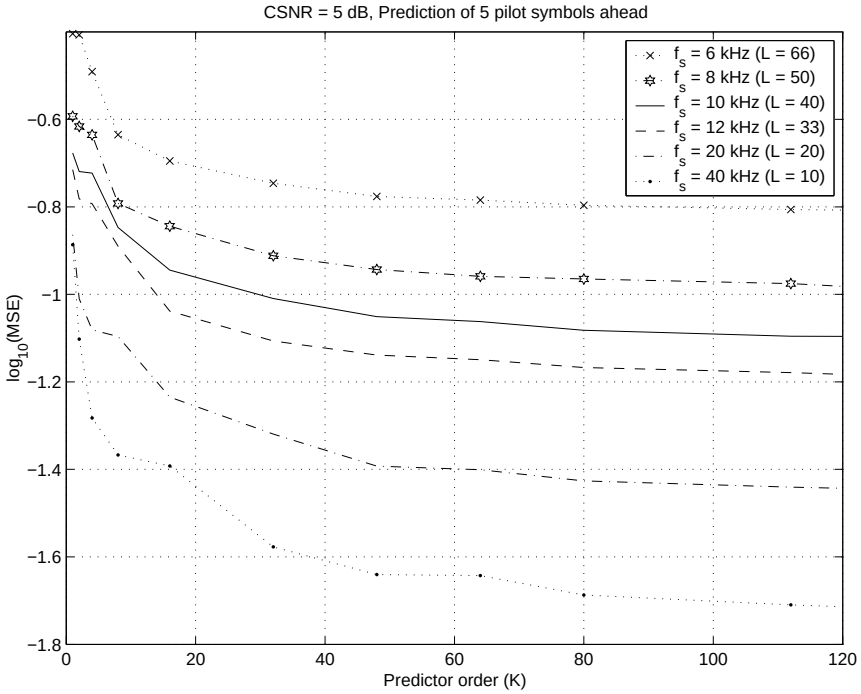


FIGURE 2.6: MSE of the prediction error as a function of predictor order K_p (K was used in [Duong, 2002] to denote the predictor order) at average CSNR $\bar{\gamma} = 5$ dB. Carrier frequency $f_c = 5$ GHz and Doppler frequency $f_d = 250$ Hz. It is plotted for different pilot sampling frequencies f_s corresponding to different pilot spacing L, and prediction of 5 pilot symbols ahead.

2.7 System Parameters Used in Numerical Example

For comparison reasons, most of the system parameters used in this thesis are chosen to be the same as those in [Cai and Giannakis, 2005]. This is because all the results of the thesis can be said to directly generalize the results in [Cai and Giannakis, 2005] in various directions. Also, for easily comparing the results of different systems under considerations we will use the same set of parameters throughout the thesis and they are as follows:

- Set of constellations $\{M_n\} = \{4, 8, 16, 32, 64, 128, 256, 512\}$.
- Carrier frequency $f_c = 2$ GHz.
- Mobile speed $v = 30$ m/s.
- Channel symbol duration $T_s = 5$ μs.

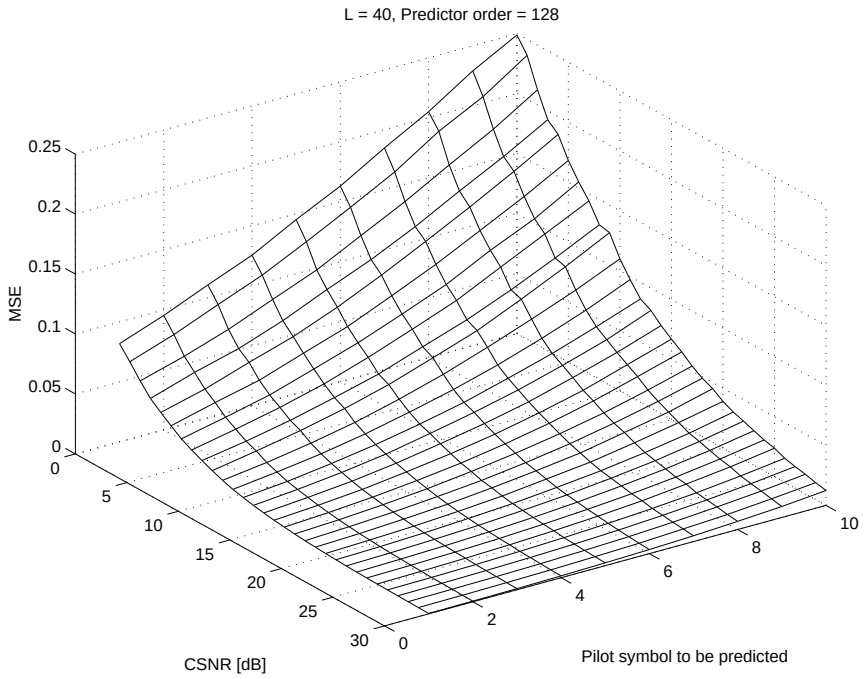


FIGURE 2.7: A 3-dimensional illustration of the MSE of the prediction error as a function of both average CSNR and the number of pilot symbols ahead to be predicted. Carrier frequency $f_c = 5$ GHz; Doppler frequency $f_d = 250$ Hz; $L = 40$; and predictor order $K_p = 128$.

- System delay $\tau = DLT_s = 1$ ms. For simplicity, and due to the fact that D has to be an integer, the pilot spacing L can only take on a certain value in the set $\{1, 2, 4, 5, 8, 10, 20, 25, 40, 50, 100, 200\}$. At the same time $L > n_{\mathrm{T}}$ where n_{T} is the number of transmit antennas. Note that, for SISO and SIMO cases, $L = 1$ implies that SE is zero (cf. Eq. (2.2)), thus $L = 1$ is included here for the sake of completeness only. This is in contrast to [Cai and Giannakis, 2005] where the pilot spacing takes on all possible integer values in the range $[2, \cdots, L_{\max}]$, where $L_{\max}$ is the largest spacing two pilot symbols can have in order to satisfy the Nyquist sampling theorem. Note that, for the given parameters, $L_{\max} = 500 \gg 200$.

- BER requirement $\mathrm{BER}_0 = 10^{-5}$.

- Estimator order $K_e = 20$ and predictor order $K_p = 250$.

- In a space diversity case, unless otherwise is explicitly stated, the sub-

channels are assumed to be mutually independent and the branches CSNRs are the same on every branch.

With this carrier frequency and mobile terminal velocity the maximum Doppler shift is $f_d = v f_c / c = 200$ Hz, where $c = 3 \cdot 10^8$ m/s is the speed of light. Hence, the normalized Doppler frequency is $\tau f_d = 0.2$. This value is very high compared to the critical normalized time delay $\tau f_d = 0.01$ (in a perfect-CSI adaptive system) below which the system operates satisfactory (otherwise, the BER performance can not be guaranteed for such systems).

The choice of $K_p = 250$ leads to a suboptimal but satisfactory predictor [Duong, 2002; Øien *et al.*, 2002]. On the other hand, $K_e = 20$ is sufficient since the estimator uses both information from past and future to estimate the current value and the MSE of the estimation error as a function of estimator order already saturates at this value [Cai and Giannakis, 2005].

Chapter 3

Adaptive Coded Modulation With Imperfect Channel State Information: the SISO Case

This chapter is based on a paper presented at the conference *Nordic Radio Symposium (NRS)* [Duong and Øien, 2004].

3.1 Introduction

In traditional PSAM systems [Cavers, 1991; Sampei and Sunaga, 1993], pilot symbols are multiplexed into the data stream at fixed spacing intervals and there is no special power treatment of these symbols. Cai and Giannakis [2005] analyzed an uncoded adaptive QAM modulation system where both pilot spacing and power allocation on pilot symbols are optimized with respect to the average channel conditions. In this chapter, we extend their concept to include forward error correction coding. We optimize the pilot spacing L and a power allocation variable α in a maximal ASE sense subject to the constraint that the instantaneous BER can not exceed a target value denote BER_0 at any given time. Moreover, we will provide a simpler optimization algorithm than the sequential quadratic programming numerical search algorithm of Cai and Giannakis [2005].

The chapter is organized as follows. Section 3.2 describes the system model where both channel estimation and prediction error variances are derived. Then BER analysis and ASE analysis are presented in Sections 3.3 and 3.4, respectively. A numerical example is given in Section 3.5 and conclusions are drawn in Section 3.6.

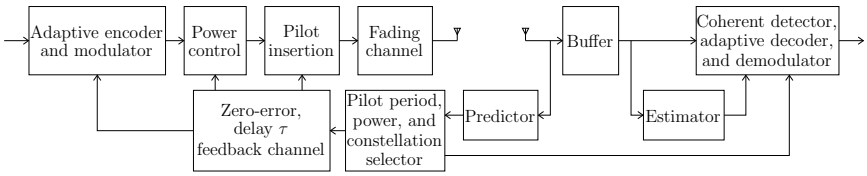


FIGURE 3.1: The adaptive PSAM system model

3.2 System Model

The narrowband system under consideration is depicted in Figure 3.1. We denote the received, noisy and faded pilot symbols in complex baseband as

$$y_{\text{pl}}(k;0) = \sqrt{\mathcal{E}_{\text{pl}}} h(k;0) s(k;0) + n(k;0) \tag{3.1}$$

and the received data symbols as

$$y_d(k;l) = \sqrt{\mathcal{E}_d} h(k;l) s(k;l) + n(k;l), \quad l \in [1, \cdots, L-1]. \tag{3.2}$$

Channel Estimation

Let $h_e(k;l)$ be the linear estimate of $h(k;l)$. The estimation error is given by $\epsilon_e(k;l) = h(k;l) - h_e(k;l)$ and the MSE is defined as $\sigma_e^2(l) = \mathsf{E}\left[|\epsilon_e(k;l)|^2\right]$. Since the channel gain by assumption is a Gaussian random process (RP) with zero mean, $h_e(k;l)$ and $\epsilon_e(k;l)$ are also zero mean Gaussian RVs. The error $\sigma_{h_e}^2(l)$ may then be written as $\sigma_{h_e}^2(l) = 1 - \sigma_e^2(l)$. In what follows we will find analytical expressions for both estimated channel and the corresponding MMSE.

Since the received samples are complex Gaussians, the optimal channel estimate in a MAP sense is a linear combination of the observations [Meyr *et al.*, 1998, p. 741–742], and the estimated channel is given by

$$h_e(k;l) = \mathbf{w}_e^{\mathsf{H}} \mathbf{y}_{\text{pl}}, \tag{3.3}$$

where $\mathbf{y}_{\text{pl}} = \left[y_{\text{pl}}(k - \lfloor K_e/2 \rfloor;0), \cdots, y_{\text{pl}}(k + \lfloor (K_e-1)/2 \rfloor;0)\right]^{\mathsf{T}}$ is a vector containing receive symbols at pilot symbol instances and $\mathbf{w}_e$ is given in (2.18). However, we restrict ourselves to always sending the same pilot symbol, whose absolute value is equal to 1. As a result, $\boldsymbol{\mathcal{D}}(\mathbf{s})$ in (2.18) is always proportional to the identity matrix and it can be shown (see Appendix A) that the optimal estimator is simplified to

$$\boldsymbol{\omega}_e = \left(\mathbf{R}_e + \frac{1}{(1-\alpha)\bar{\gamma}L}\mathbf{I}_{K_e}\right)^{-1} \mathbf{r}_e.$$

Hence the estimated channel becomes

$$h_e(k;l) = \boldsymbol{\omega}_e^{\mathsf{H}} \tilde{\mathbf{y}} \tag{3.4}$$

where $\tilde{\mathbf{y}} = \mathbf{y}_{\text{pl}}/(s\sqrt{\mathcal{E}_{\text{pl}}})$ and it is known as the maximum likelihood (ML) channel estimation. Moreover, it is also shown in Appendix A that the MMSE can be written as [Cai and Giannakis, 2005]

$$\sigma_e^2(l) = 1 - \sum_{\kappa=1}^{K_e} \frac{|\mathbf{u}_\kappa^{\mathsf{H}} \mathbf{r}_e|^2 (1-\alpha) L\bar{\gamma}}{(1-\alpha)L\bar{\gamma}\lambda_\kappa + 1}. \tag{3.5}$$

The vectors $\mathbf{u}_\kappa$ for $\kappa = [1, \cdots, K_e]$ are the eigenvectors corresponding to the eigenvalues λ_κ of the correlation matrix $\mathbf{R}_e$.

Channel Prediction

Let $h_p(k;l)$ be the predicted channel and let the MSE of the prediction error be $\sigma_p^2(l) = \mathsf{E}\big[|\epsilon_p(k;l)|^2\big]$, where $\epsilon_p(k;l) = h(k;l) - h_p(k;l)$. Again, it is easy to show that $\sigma_{h_p}^2(l) = 1 - \sigma_p^2(l)$. Applying the same analysis as in the estimation case, the MAP-optimal linear predicted channel gain is[1]

$$\begin{aligned} h_p(k;l) &= \mathbf{w}_p^{\mathsf{H}} \mathbf{y}_{\text{pl}} \\ &= \mathbf{r}_p^{\mathsf{H}} \left(\mathbf{R}_p + \frac{1}{(1-\alpha)L\bar{\gamma}} \mathbf{I}_{K_p} \right)^{-1} \tilde{\mathbf{y}} = \boldsymbol{\omega}_p^{\mathsf{H}} \tilde{\mathbf{y}}. \end{aligned} \tag{3.6}$$

Now, $\mathbf{y}_{\text{pl}} = \big[y_{\text{pl}}(k-D;0), \cdots, y_{\text{pl}}(k-D-K_p+1;0)\big]^{\mathsf{T}}$, $\tilde{\mathbf{y}} = \mathbf{y}_{\text{pl}}/(s\sqrt{\mathcal{E}_{\text{pl}}})$ is the ML channel estimate, and $\mathbf{w}_p$ and $\boldsymbol{\omega}_p = (\mathbf{R}_p + \mathbf{I}_{K_p}/[(1-\alpha)\bar{\gamma}L])^{-1}\mathbf{r}_p$ are the optimal predictor given by (2.20) and (A.14), respectively.

Similarly to (3.5) the MMSE of prediction error is

$$\sigma_p^2(l) = 1 - \sum_{\kappa=1}^{K_p} \frac{|\mathbf{u}_\kappa^{\mathsf{H}} \mathbf{r}_p|^2 (1-\alpha) L\bar{\gamma}}{(1-\alpha)L\bar{\gamma}\lambda_\kappa + 1}. \tag{3.7}$$

In this case, $\{\mathbf{u}_\kappa\}_{\kappa=1}^{K_p}$ are the eigenvectors corresponding to the eigenvalues $\{\lambda_\kappa\}_{\kappa=1}^{K_p}$ of the matrix $\mathbf{R}_p$.

Comparing our filters to the MAP-optimal filter in [Øien *et al.*, 2004] we see that our filters are slightly different since the factor in front of the identity matrix involves the frame size L and the variable α. This is obvious since we have different pilot and signal power, and both parameters will be optimized to achieve maximum throughput.

[1] If $\mathbf{A}$ is Hermitian, then $\mathbf{A}^{-1}$ is also Hermitian [Råde and Westergren, 2000].

Since the channel is assumed to be zero-mean complex Gaussian and since the prediction error filter has the whitening property, the predicted channel is also complex Gaussian with zero mean. Then the PDF of the amplitude is Rayleigh and the predicted CSNR will follow an exponential distribution or, equivalently, a gamma distribution: $\hat{\gamma} \sim \mathcal{G}(1, \bar{\hat{\gamma}})$, where $\bar{\hat{\gamma}}$ is average of the total predicted CSNR given by (3.18).

Data is not transmitted when $\hat{\gamma} < \hat{\gamma}_1$ and according to (2.13), the actual transmission data power can thus be configured as

$$\mathcal{E}_d = \frac{\bar{\mathcal{E}}_d}{\int_{\hat{\gamma}_1}^{\infty} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma}} = \bar{\mathcal{E}}_d \exp\left(\frac{\hat{\gamma}_1}{\bar{\hat{\gamma}}}\right). \tag{3.8}$$

3.3 BER Analysis

Whether or not the channel estimates are imperfect or assumed perfect, they will be used for detection of the incoming signal. In case the estimates are not perfect, they will be used as if they represent true values. To reduce the complexity of the receiver, we use a suboptimal symbol-by-symbol ML detection

$$\xi(k;l) = \frac{y_d(k;l)}{\sqrt{\mathcal{E}_d} h_e(k;l)}. \tag{3.9}$$

As a result, the instantaneous CSNR at the time of detection is equal to [Cai and Giannakis, 2005]

$$\gamma(k;l) = \frac{\mathcal{E}_d |h_e(k;l)|^2}{N_0 + g\mathcal{E}_d \sigma_e^2(l)}, \tag{3.10}$$

where $g = 1$ for 4-QAM and $g = 1.3$ for higher QAM constellations. Clearly, imperfect estimation degrades the CSNR. If estimation is perfect, $h_e(k;l)$ coincides with the true channel $h(k;l)$ and the MMSE is zero. In this case, $\gamma = \mathcal{E}_d |h|^2 / N_0$, the standard expression in the literature.

3.3.1 BER in the Presence of Channel Estimation Errors

Trellis codes are used as component codes in our analysis, and a tight approximation expression for the BER performance on AWGN channels is given in (2.7). Replacing γ in (2.7) with (3.10), we have the following expression for the BER when having estimated the channel:

$$\text{BER}\left(M_n \middle| |h_e|\right) = \begin{cases} a_n \exp\left(-\dfrac{b_n \mathcal{E}_d |h_e|^2}{M_n (N_0 + g\mathcal{E}_d \sigma_e^2(l))}\right) & \text{when } |h_e| \geq |h_e|_{n,T} \\ \dfrac{1}{2} & \text{when } |h_e| < |h_e|_{n,T}. \end{cases} \tag{3.11}$$

The thresholds $|h_e|_{n,T}$ are easily found by equating the exponential part of (3.11) to the second part of the same equation and solve for $|h_e|$. The result is $|h_e|_{n,T} = \sqrt{\ln(2a_n)M_n(N_0 + g\mathcal{E}_d\sigma_e^2)/(b_n\mathcal{E}_d)}$. With this expression at hand, and since the system adaptation is based on predicted CSNR, we will derive the BER when we have predicted the channel, where the estimation error is also accounted for.

3.3.2 BER in the Presence of Both Channel Estimation and Prediction Errors

Since the predicted and estimated channels respectively are modelled as $h_p(k;l) = h(k;l) - \epsilon_p(k;l)$ and $h_e(k;l) = h(k;l) - \epsilon_e(k;l)$, the estimated channel can be written as

$$h_e(k;l) = h_p(k;l) + \epsilon_p(k;l) - \epsilon_e(k;l). \tag{3.12}$$

This representation is the key to the following analysis. Note that all the RVs in the above equation are complex Gaussians. Evidently, $h_p(k;l)$ and $\epsilon_p(k;l)$ are uncorrelated due to the orthogonality principle [Papoulis and Pillai, 2002, Chap. 7-3] used when designing the MAP-optimal predictor. Moreover, the normalized correlation between $h_p(k;l)$ and $\epsilon_e(k;l)$, denoted by ϱ, can be found based on the MMSE channel estimator and predictor. As a result, when $h_p(k;l)$ is given, $h_e(k;l)$ is Gaussian distributed with mean $\mu_{h_e|h_p} = \mathsf{E}\big[h_e(k;l)|h_p(k;l)\big] = (1-\varrho)h_p(k;l)$ and variance $\sigma^2_{h_e|h_p} = \mathsf{E}\big[|\epsilon_p(k;l) - \epsilon_e(k;l) + \varrho h_p(k;l)|^2\big]$. Since the mean $\mu_{h_e|h_p} \neq 0$, the magnitude $|h_e(k;l)|$ given $h_p(k;l)$ follows a Rice distribution with the Rician factor $K = |(1-\varrho)h_p(k;l)|^2/\sigma^2_{h_e|h_p}$. However, it is shown by Cai and Giannakis [2005] that ϱ typically takes on very small values, and that the BER performance deviation obtained by using the exact value of ϱ instead of setting $\varrho = 0$ is less than 0.2 dB. We therefore choose to set $\varrho = 0$ in our analysis. Hence, K reduces to $K = |h_p(k;l)|^2/\sigma^2_{h_e|h_p}$, where $\sigma^2_{h_e|h_p} = \sigma_p^2(l) + \sigma_e^2(l)$ when assuming that $\epsilon_p(k;l)$ and $\epsilon_e(k;l)$ are uncorrelated.[2]

To obtain the expression characterizing the BER performance given the predicted channel, we average the BER conditioned on the estimated channel over the PDF of estimated channel conditioned on the predicted chan-

[2]Since both the causal and noncausal estimates use partly the same information, there should be a dependence between these two quantities. From next chapter, we will relax this assumption to account for the correlation between the two.

nel. That is,

$$\begin{aligned}\mathrm{BER}(M_n|h_p) &= \int_0^\infty \mathrm{BER}\left(M_n||h_e|\right) f_{h_e|h_p}\left(|h_e||h_p\right) d|h_e| \\ &= \mathcal{T}(0,\infty) - \mathcal{T}\left(0,|he|_{n,T}\right) + \mathcal{F}\left(0,|he|_{n,T}\right), \end{aligned} \tag{3.13}$$

where

$$\mathcal{T}(0,\infty) = \int_0^\infty a_n \exp\left(-\frac{b_n \mathcal{E}_d |h_e|^2}{M_n(N_0 + g\mathcal{E}_d\sigma_e^2)}\right) f_{h_e|h_p}\left(|h_e||h_p\right) d|h_e|, \tag{3.14}$$

$$\mathcal{T}\left(0,|h_e|_{n,T}\right) = \int_0^{|h_e|_{n,T}} a_n \exp\left(-\frac{b_n \mathcal{E}_d |h_e|^2}{M_n(N_0 + g\mathcal{E}_d\sigma_e^2)}\right) f_{h_e|h_p}\left(|h_e||h_p\right) d|h_e|, \tag{3.15}$$

$$\mathcal{F}\left(0,|he|_{n,T}\right) = \int_0^{|h_e|_{n,T}} \frac{1}{2} f_{h_e|h_p}\left(|h_e||h_p\right) d|h_e|. \tag{3.16}$$

Closed-form solutions for these integrals are derived in Appendix B.

The predicted CSNR can be calculated as [Øien *et al.*, 2004; Cai and Giannakis, 2005]

$$\hat{\gamma}(k;l) = \frac{\bar{\mathcal{E}}_d |h_p(k;l)|^2}{N_0}. \tag{3.17}$$

In the above equation, $\bar{\mathcal{E}}_d$ is used because we do not know whether or not the system is in outage at the time it is predicted (i.e. τ seconds ahead). Taking the expectation of both sides of (3.17) gives $\bar{\hat{\gamma}} = \bar{\mathcal{E}}_d(1-\sigma_p^2(l))/N_0$. Upon replacing $N_0 = \mathcal{E}/\bar{\gamma}$ (cf. Eq. (1.5)) and defining $r = \bar{\mathcal{E}}_d(1-\sigma_p^2(l))/\mathcal{E}$ the average CSNR is

$$\bar{\hat{\gamma}} = \frac{\bar{\mathcal{E}}_d}{N_0}\left(1-\sigma_p^2(l)\right) = r\bar{\gamma}. \tag{3.18}$$

Solving (3.17) for $|h_p|$ and inserting the answer into the solution of the

integrals in (3.14)–(3.16) gives

$$\begin{aligned}\mathrm{BER}(M_n|\hat{\gamma}) =& a_n d_n \exp\left(-\frac{\hat{\gamma} A_n d_n \mathcal{E}\mathcal{E}_d}{\bar{\gamma}\bar{\mathcal{E}}_d}\right) \\ &- a_n d_n \exp\left(-\frac{\hat{\gamma}\mathcal{E}}{\bar{\gamma}\bar{\mathcal{E}}_d\sigma^2_{h_e|h_p}}\right) \\ &\quad\times \sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{\hat{\gamma} d_n \mathcal{E}}{\bar{\gamma}\bar{\mathcal{E}}_d\sigma^2_{h_e|h_p}}\right)^m \left(1-\overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{d_n\sigma^2_{h_e|h_p}}\right)\right) \\ &+ \frac{1}{2}\exp\left(-\frac{\hat{\gamma}\mathcal{E}}{\bar{\gamma}\bar{\mathcal{E}}_d\sigma^2_{h_e|h_p}}\right) \\ &\quad\times \sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{\hat{\gamma}\mathcal{E}}{\bar{\gamma}\bar{\mathcal{E}}_d\sigma^2_{h_e|h_p}}\right)^m \left(1-\overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{\sigma^2_{h_e|h_p}}\right)\right),\end{aligned} \tag{3.19}$$

where $\overline{\Gamma}(\cdot,\cdot)$ is the normalized incomplete gamma function (see Definition 2 in Appendix B). For the clarity of the expressions, we have collected some constants and variables and grouped them into the terms

$$A_n = \frac{b_n}{M_n(N_0 + g\mathcal{E}_d\sigma_e^2)}, \tag{3.20a}$$

$$d_n = \frac{1}{A_n\mathcal{E}_d\sigma^2_{h_e|h_p}+1}. \tag{3.20b}$$

Using an approach similar to the one in [Webb and Steele, 1995] the optimal switching thresholds can be determined by solving $\mathrm{BER}(M_n|\hat{\gamma}) = \mathrm{BER}_0$ with respect to $\hat{\gamma}$ for different M_n constellations. This operation can not be done analytically since, as we can see from (3.19), the dependence of $\hat{\gamma}$ is very involved. Thus, a numerical solution must be used.

Now, in order to calculate $\overline{\mathrm{BER}}$ in Eq. (2.6), we need $\mathrm{BER}(M_n)$ which is the average BER for the actual constellation M_n. To obtain that, we need to average $\mathrm{BER}(M_n|\hat{\gamma})$ over the PDF of the predicted CSNR within the range in which the constellation M_n is employed. As mentioned earlier, the predicted channel also follows a Rayleigh distribution. Then, the predicted CSNR is exponentially distributed, or equivalently, gamma distributed: $\hat{\gamma} \sim \mathcal{G}(1, r\bar{\gamma})$. Hence, $\mathrm{BER}(M_n)$ is expressible as

$$\begin{aligned}\mathrm{BER}(M_n) &= \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathrm{BER}(M_n|\hat{\gamma}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \\ &= \mathcal{B}1(M_n) - \mathcal{B}21(M_n) + \mathcal{B}22(M_n).\end{aligned} \tag{3.21}$$

The three integrals of interest can be expressed as follows (for detailed calculations, look into the Appendix B):

$$\mathcal{B}1(M_n) = \frac{a_n d_n \bar{\mathcal{E}}_d}{rA_n d_n \mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d} \times \left[\exp\left(-\hat{\gamma}_n \frac{rA_n d_n \mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}\bar{\mathcal{E}}_d} \right) - \exp\left(-\hat{\gamma}_{n+1} \frac{rA_n d_n \mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}\bar{\mathcal{E}}_d} \right) \right], \quad (3.22)$$

$$\mathcal{B}21(M_n) = \frac{a_n \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\mathcal{E}} \sum_{m=0}^{\infty} \left(\frac{r d_n \mathcal{E}}{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right)^{m+1} \left(1 - \overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{d_n \sigma^2_{h_e|h_p}} \right) \right) \times \left[\overline{\Gamma}\left(1+m, \hat{\gamma}_n \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) - \overline{\Gamma}\left(1+m, \hat{\gamma}_{n+1} \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) \right], \quad (3.23)$$

and

$$\mathcal{B}22(M_n) = \frac{\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{2r\mathcal{E}} \sum_{m=0}^{\infty} \left(\frac{r\mathcal{E}}{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right)^{m+1} \left(1 - \overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{\sigma^2_{h_e|h_p}} \right) \right) \times \left[\overline{\Gamma}\left(1+m, \hat{\gamma}_n \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) - \overline{\Gamma}\left(1+m, \hat{\gamma}_{n+1} \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) \right]. \quad (3.24)$$

As a check, it can be seen from (3.15) and (3.16) that they are equal if $a_n = 1/2$ and $b_n = 0$ (implying that $d_n = 1$). Thus, (3.24) also can be obtained from (3.23) by setting $a_n = 1/2$ and $d_n = 1$.

Before calculating the average BER given by (2.6), the probabilty P_n has to be calculated. Integrating the exponential PDF (or gamma PDF with shape parameter equal to 1) over the interval $[\hat{\gamma}_n, \hat{\gamma}_{n+1}\rangle$ gives

$$P_n = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} = \exp\left(-\frac{\hat{\gamma}_n}{r\bar{\gamma}} \right) - \exp\left(-\frac{\hat{\gamma}_{n+1}}{r\bar{\gamma}} \right). \quad (3.25)$$

Finally, we are able to find the $\overline{\text{BER}}$ by combining Eqs. (2.2), (3.21)–(3.24), and (3.25).

3.4 ASE Analysis

The goal of this section is to optimize the adaptive system in a maximal ASE sense based on feedback CSI. The system will adapt the code, the pilot

spacing L, and the power allocation variable α in such a way that ASE is maximized while keeping $\text{BER}(M_n|\hat{\gamma}) \leq \text{BER}_0$. As can be seen from the analysis in the previous section, the approach includes both estimation and prediction errors.

It is obvious that the variance of the prediction error is largest when predicting the last symbol in a frame—i.e., $l = L-1$. On the other hand, the variance of the estimation error is almost the same for all l if the order of the estimation filter is $K_e \geq 20$ [Cai and Giannakis, 2005]. Thus we will use the estimation error variance $\sigma_e^2(L-1)$ and the conservative choice of the prediction error variance $\sigma_p^2(L-1)$ when finding the optimal switching thresholds $\{\hat{\gamma}_n\}_{n=1}^N$ as well as in the further optimization process.

Upon combining (2.2) and (3.25) as shown in (2.5) the average ASE is given as

$$\begin{aligned}\text{ASE} &= \sum_{n=1}^{N} R_n P_n \\ &= \frac{L-1}{L}\left[\exp\left(-\frac{\hat{\gamma}_1}{r\bar{\gamma}}\right)\left(\log_2(M_n)-\frac{1}{2}\right)\right. \\ &\qquad \left. + \sum_{n=2}^{N}\exp\left(-\frac{\hat{\gamma}_n}{r\bar{\gamma}}\right)\log_2\left(\frac{M_n}{M_{n-1}}\right)\right]. \end{aligned} \tag{3.26}$$

Let f_s be the pilot sampling frequency. The Nyquist theorem says that f_s must be at least twice the bandwidth of the channel gain process in order to avoid aliasing. In an isotropic scattering environment, the fading is described by the Jakes spectrum and, as a result, the bandwidth is strictly band-limited to the Doppler frequency f_d. When signaling at Nyquist rate, the time between two pilot symbols is $LT_s = 1/(2f_d)$. Hence, the maximum distance between two pilots—measured in number of symbols—is $L_{\max} = \lfloor 1/(2f_dT_s)\rfloor$. Since every Lth symbol is a pilot, then $L = 1$ gives zero SE. Thus, for $L \in [2, \cdots, L_{\max}]$ we have the following optimization problem[3]

$$\begin{aligned} &\max_{\alpha} \ \text{ASE} \\ &\text{subject to } \ 0 < \alpha < 1. \end{aligned} \tag{3.27}$$

For a given mean CSNR, we need to calculate the switching thresholds such that $\text{BER}(M_n|\hat{\gamma}) = \text{BER}_0$. It turns out that $\text{BER}(M_n|\hat{\gamma})$ depends on—among other things—$\mathcal{E}_d$ which, again, is a function of the first switching

[3]Unfortunately, we can not express ASE in such a way that its dependence on α and L becomes clear. This difficulty is due to the fact that both L and α are very involved in both ASE and BER expressions.

threshold $\hat{\gamma}_1$ (cf. Eq. (3.8)). The problem now is reduced to finding the first threshold $\hat{\gamma}_1$, while the other thresholds are easily found once $\hat{\gamma}_1$ is known. The optimization algorithm will be used extensively throughout the thesis and, therefore, we summarize it here in the following algorithm:

Algorithm 1 (The optimization algorithm)
Given an average CSNR level and a value of L within the possible range, we optimize $\alpha \in \mathbb{R}^+$ in the interval $\langle 0, 1\rangle$ such that ASE is maximized. Thus, for each value of α we do the following steps:

- Let $\mathcal{E}_d$ be a function of the whole dynamic range of $\hat{\gamma}$ (note that it is not $\hat{\gamma}_1$).
- With that $\mathcal{E}_d$ we solve $\text{BER}(M_1|\hat{\gamma}) = \text{BER}_0$ with respect to $\hat{\gamma}$, resulting in $\hat{\gamma}_1(\alpha, L)$.
- $\mathcal{E}_d$ is now explicitly given by (3.8) and the other switching thresholds $\{\hat{\gamma}_n(\alpha)\}_{n=2}^{N}$ can easily be found—for the given value of α—by solving $\text{BER}(M_n|\hat{\gamma}) = \text{BER}_0$ for $n = 2, \cdots, N$ with the value $\mathcal{E}_d(\hat{\gamma}_1)$ inserted.
- If ASE reaches its max with this α, then α is declared to be the optimal solution. Otherwise, the algorithm starts over again with another value of α until an optimal solution is obtained.

In this way, the optimization of (3.27) is done simply by letting a numerical search to pick a real-valued $\alpha \in \langle 0, 1\rangle$ which maximizes the ASE. For this purpose, we have used the function `fminbnd` in MATLAB where we have to specify a lower bound and an upper bound of α—where the solution is contained—to find the solution.

After solving (3.27) for all the possible L values, the maximum ASE is found by searching over all L in order to find the α and L values which simultaneously maximize ASE.

Once ASE is maximized, we have an optimal value of α. As a result, the corresponding switching thresholds are also optimal. Note that, since ASE is finite, there is always a solution to the optimization problem. At this point, it is noted that, the problem of maximizing ASE has been reduced from a 2-D function $S(\alpha, \mathcal{E}_d)$ in [Cai and Giannakis, 2005, Eq. (23)] to a 1-dimensional (1-D) function, i.e., $\text{ASE}(\alpha)$. Optimization of a 1-D function is, in general, simpler than the sequential quadratic programing numerical search over a 2-D function (i.e., 2 constraints) which is used in [Cai and Giannakis, 2005].

3.5 Numerical Example

To illustrate the theory analyzed in the previous sections, we run a calculation on the basis of the system parameters given in Section 2.7. The results

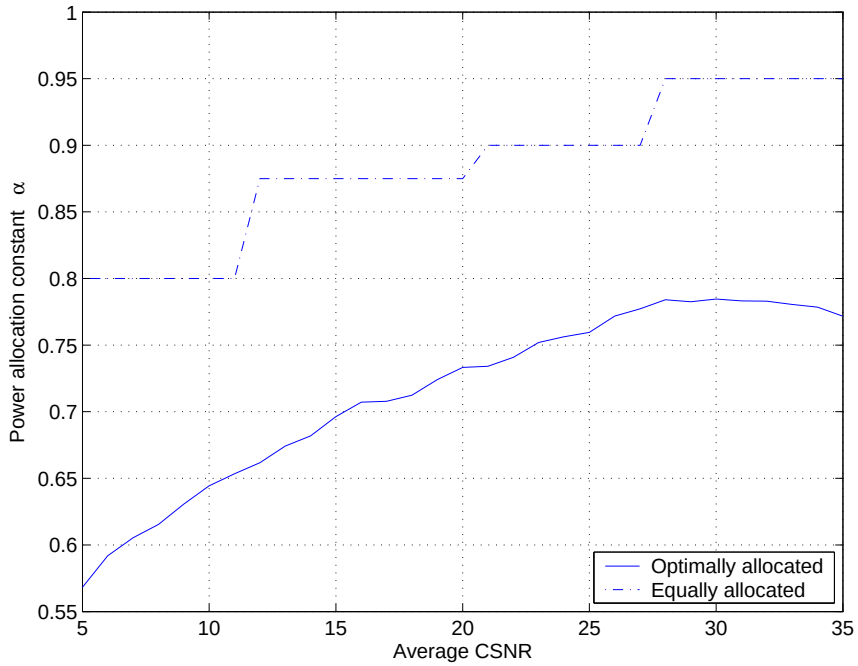


FIGURE 3.2: Fraction of power allocation to data symbols when the pilot period L is optimal. Carrier frequency $f_c = 2$ GHz; Doppler frequency $f_d = 200$ Hz. Both prediction and estimation errors are taken into account and the system delay $\tau = 1$ ms.

of α, L, ASE, and $\overline{\text{BER}}$ are given in Figures 3.2, 3.3, 3.4, and 3.5, respectively. Note that these results are somewhat different to what was presented in [Duong and Øien, 2004]. The changes are done to reflect the fact that the pilot spacing L only can take on certain discrete values; which were introduced in Section 2.7.

Figure 3.2 shows the optimum fraction of power allocated to data symbols when pilot spacing L is optimized. When the power is optimally allocated, less power is put on data symbols compared to when the power is equally allocated. This effect can partially be explained by the fact that we already have protected our data, so less power is needed for the same error performance. Thus, the power can be more freely distributed to the pilots. On the other hand, as the quality of the channel increases in terms of average CSNR, data power increases and less power is put on pilots. In this case, the correlation between the pilot time instances becomes more reliable even though the pilot spacing is larger. This is due to the improvement of the channel quality.

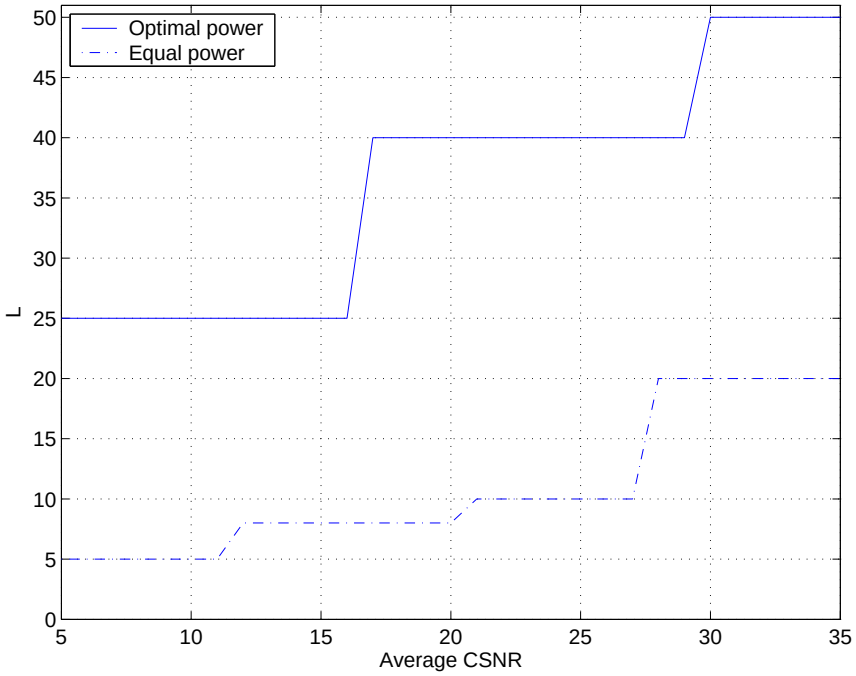


FIGURE 3.3: Optimum pilot spacing L for both optimal and equal power allocation. Carrier frequency $f_c = 2$ GHz; Doppler frequency $f_d = 200$ Hz. Both prediction and estimation errors are taken into account and the system delay $\tau = 1$ ms.

The development of the pilot spacing L with the channel quality is depicted in Figure 3.3. Clearly, as expected, denser pilot symbol spacing is needed when the channel is bad, and it increases with increasing average CSNR. In the optimal power scenario, the distance between pilots is considerably larger than in the equal power scenario. Looking at both Figures 3.2 and 3.3 we see that there is a good match between how the power is distributed and the pilot spacing: long pilot period in combination with high pilot power (optimal power case) versus short pilot period with low pilot power (equal power case). This is necessary in order to maintain a certain correlation between the pilot samples impaired by noise since this correlation is exactly what the system exploits to predict and estimate the channel. The better the correlation is, the more accurate the tracking of the channel becomes and the better is the BER to be maintained.

The optimum ASE is presented in Figure 3.4 as a function of expected CSNR. The capacity of a fading channel is also included. Needless to say, as can be observed, the ASE performance corresponding to optimal power

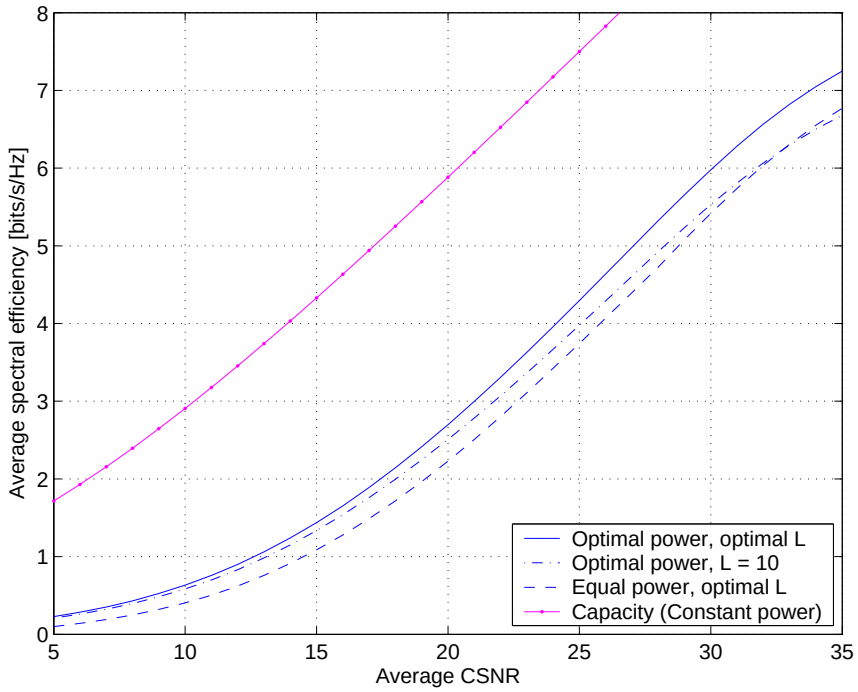


FIGURE 3.4: Average spectral efficiency against expected CSNR. Carrier frequency $f_c = 2$ GHz; Doppler frequency $f_d = 200$ Hz. Both prediction and estimation errors are accounted and the system delay $\tau = 1$ ms.

and optimal L case outperforms both when either power distribution is optimal in combination with fixed L or optimal L in combination with equal power. Moreover, when comparing our results to the uncoded M-QAM performance in [Cai and Giannakis, 2005], we observe that our optimum ASE based on coded QAM constellations is just slightly better than for uncoded M-QAM at very low CSNR. The gain is bigger with increasing CSNR and is almost 1 bit/s/Hz at 30 dB CSNR. This corresponds to a gain of about 3 dB in average CSNR due to coding (and due to the fact that we have one more constellation to use).

Average BER is plotted in Figure 3.5. As can be seen, it is always lower than BER_0. This is as expected, since the switching thresholds are calculated to keep the instantaneous BER (with respect to the predicted CSNR) below the target value BER_0. That is $\text{BER}(M_n|\hat{\gamma}) \leq \text{BER}_0$. The average BER must therefore be lower than BER_0. It is plotted for both optimal and equal power allocation when L is optimal in both cases. The $\overline{\text{BER}}$ curve for optimal power and L is (unnecessary) lower than when only L is optimal. On

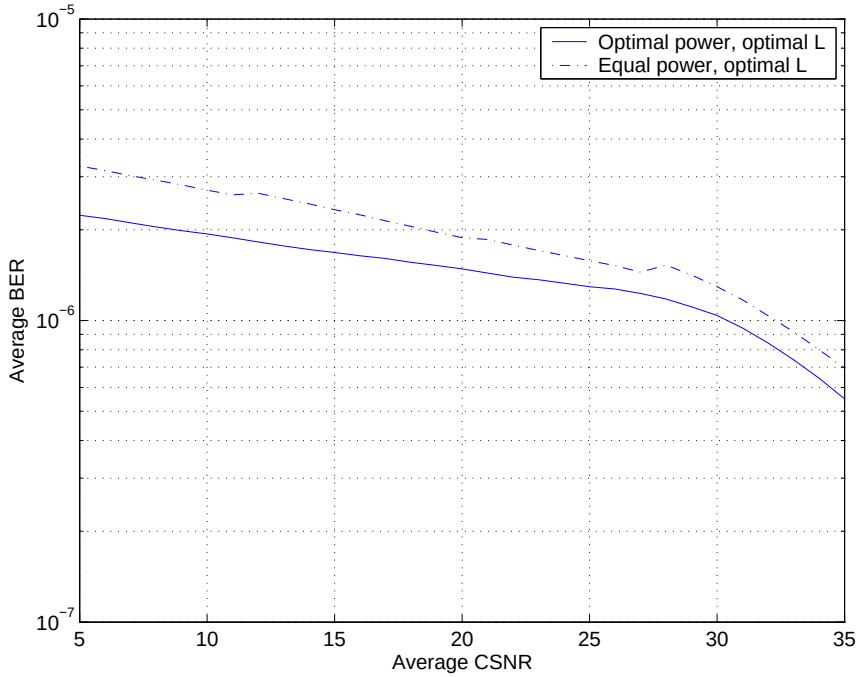


FIGURE 3.5: Average $\overline{\text{BER}}$plotted for optimal power and optimal L, and equal power and optimal L.

the other hand, the thoughput in terms of ASE is still 0.5 bits/s/Hz higher when the power distribution is optimal.

3.6 Concluding Remarks

We have investigated an ACM system where the power allocation between pilot and data symbols is optimized and where the pilot spacing is adaptively changed in an optimal way according to the quality of the channel. Both channel estimation error at the receiver and channel prediction error at the transmitter have been taken into account. The results justify the advantage of having the system parameters optimized, and it was shown that ASE performance is considerably higher than the uncoded case. The gain is up to 1 bits/s/Hz at high CSNR regions. This is achieved without losing BER performance since the $\overline{\text{BER}}$ curves are always below the target BER.

Chapter 4

Adaptive Coded Modulation With Receive Antenna Diversity and Imperfect Channel Knowledge

This chapter is based on a paper presented at the conference *European Signal Processing Conference (EUSIPCO)* [Duong, Øien, and Hole, 2005], a journal paper published in *IEEE Transactions on Vehicular Technology* [Duong, Øien, and Hole, 2006], and a temporary document at *12th Management Committee Meeting, COST Action 273: Towards Mobile Multimedia Broadband Networks* [Duong and Øien, 2005]. The last part of this chapter, dealing with the average time the fading remains within a fading interval, is taken from [Duong and Øien, 2006].

4.1 Introduction

The scheme presented here is a generalization of two papers—one by Cai and Giannakis [2005], in which an adaptive modulation system based on uncoded M-QAM constellations with a single transmit and a single receive antenna is investigated, and one by Øien *et al.* [2004], where an ACM system with MRC reception was investigated assuming perfect receiver CSI. We extend and unify the idea of these papers by analyzing the case when coding is included and when MRC is implemented at the receiver. In case of imperfect channel estimation, MRC is no longer the optimal combining scheme. The optimal combining technique was presented by You

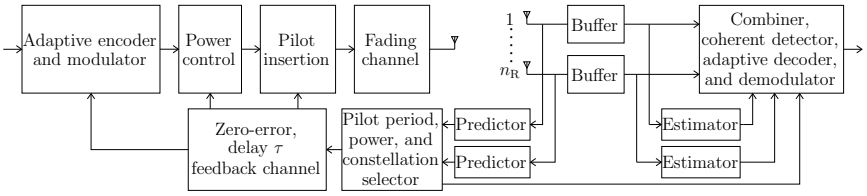


FIGURE 4.1: The ACM system based on adaptive PSAM combined with multi reception of receiving signal.

et al. in 2005 where they considered estimation error as an additional noise source that is not white. However, the results in that paper indicated that the error performance based on the suboptimal combining scheme—where the estimated channels are treated as the true channel state and are used for MRC combining—is very close to that of the optimal combining. Thus, for simplicity, we will use the suboptimal combining technique. The system parameters are optimized, as in the previous chapter. Moreover, we will also address the question on how often, on average, the system needs to switch between different codes.

The chapter is organized as follows. Section 4.2 describes the system under study. The BER performance is analyzed in presence of both estimation and prediction errors in Section 4.3. Formulation of ASE performance is given in Section 4.4, while expressions related to average fade region duration (AFRD) are given in Section 4.5. Numerical analysis is presented in Section 4.6 and, finally, the conclusions are drawn in Section 4.7

4.2 Description of the System

With reference to Figure 4.1, the system now has n_R receive antennas. The signals are combined with the MRC technique, and it is assumed that the branches are mutually uncorrelated. This can be accomplished by separating the antenna elements with at least half a wavelength from each other in an isotropic scattering environment [Rappaport, 2002, p. 385]. However, for low fading correlation, a separation of a quarter of a wavelength is sufficient at the mobile unit [Winters, 1998]. This distance also is sufficient for base station antennas in indoor systems. In outdoor systems, on the other hand, separation of 10–20 wavelengths is required at the base station [Winters, 1998]. This is due to the fact that the base station is often located at high altitude such that the angular spread—the angle over which the signal arrives at the receive antennas—can be very small.

Similar to (2.9) and (2.10), the received, noisy, and faded pilot symbols

of the μth branch are written in complex baseband as

$$y_{\mathrm{pl};\mu}(k;l) = \sqrt{\mathcal{E}_{\mathrm{pl}}} h_\mu(k;l) s(k;l) + n_\mu(k;l), \qquad l = 0;\ \ \mu = 1, \cdots, n_{\mathrm{R}}, \quad (4.1)$$

and the received data symbols as

$$y_{d;\mu}(k;l) = \sqrt{\mathcal{E}_d} h_\mu(k;l) s(k;l) + n_\mu(k;l), \qquad l = 1, \cdots, L-1;\ \ \mu = 1, \cdots, n_{\mathrm{R}}. \quad (4.2)$$

Each subchannel is assumed to be slowly varying so that it remains constant over many channel symbols, and we perform the estimation and prediction independently on each subchannel.

4.3 BER Performance Analysis

Since the subchannels are assumed to be i.i.d. and we do both estimation and prediction independently on each subchannel, the channel estimation and the channel prediction analysis procedures developed in the previous chapter is still applicable on a per branch basis.

4.3.1 BER Analysis in the Presence of Estimation Errors

To reduce the complexity of the receiver, suboptimal symbol-by-symbol ML detection is performed on each receiver branch (cf. Eq. (3.9)). The total CSNR after MRC[1] is the sum of the individual branch CSNRs [Jakes, 1994, Ch. 5; Brennan, 2003, Eq. (26)]. The CSNR on each branch γ_μ is then identical to γ in (3.10) and the total CSNR is therefore

$$\gamma(k;l) = \sum_{\mu=1}^{n_{\mathrm{R}}} \gamma_\mu = \frac{\mathcal{E}_d}{N_0 + g\mathcal{E}_d \sigma_e^2(l)} \sum_{\mu=1}^{n_{\mathrm{R}}} |h_{e;\mu}(k;l)|^2, \quad (4.3)$$

where the last equality is obtained by noting that the variance of the estimation error is the same for all branches, i.e., $\sigma_{e;\mu}^2(l) = \sigma_e^2(l)\ \forall \mu$ and $\sigma_e^2(l)$ is given in Eq. (3.5). This is true, because the branches are assumed to be i.i.d., and the same estimation algorithm is used on all branches.

[1] Diversity combining, in general, is useful for combating fading and is efficient in the absence of interference. In the presence of cochannel interference, however, the array processing methods are able to provide better performance by cancelling interferences [Winters, 1984].

Tight approximations for the BER performance of 4-D trellis codes are given in (2.7) and are used in the previous chapter. However, in order to obtain a closed-form and a mathematically tractable solution when MRC is considered, we use a somewhat looser BER approximation as given in (2.8). Inserting (4.3) into (2.8) the BER becomes

$$\mathrm{BER}\left(M_n \middle| \{h_{e;\mu}\}\right) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \prod_{\mu=1}^{n_\mathrm{R}} \exp\left(-A_n(\ell)\mathcal{E}_d |h_{e;\mu}(k;l)|^2\right). \tag{4.4}$$

This equation will be useful when deriving the BER in the presence of both prediction and estimation errors in the next subsection. Similar to (3.20) we collect some constants and variables and define the term $A_n(\ell)$ and $d_n(\ell)$ as

$$A_n(\ell) = \frac{b_n(\ell)}{M_n(N_0 + g\mathcal{E}_d\sigma_e^2(l))}, \tag{4.5a}$$

$$d_n(\ell) = \frac{1}{A_n(\ell)\mathcal{E}_d\sigma^2_{h_e|h_p} + 1}, \tag{4.5b}$$

respectively. Whilst the constant $A_n(\ell)$ already has appeared in Eq. (4.4), the constant $d_n(\ell)$ is defined for later use.

4.3.2 BER Analysis in the Presence of Both Estimation and Prediction Errors

Following (3.12) we may express the estimated channel as

$$h_{e;\mu}(k;l) = h_{p;\mu}(k;l) + \epsilon_{p;\mu}(k;l) - \epsilon_{e;\mu}(k;l). \tag{4.6}$$

In Chapter 3, it was established that for a single antenna system, the PDF $f_{h_{e;\mu}|h_{p;\mu}}\left(|h_{e;\mu}||h_{p;\mu}\right)$ is Rice. For an MRC system operating on i.i.d. branches, $f_{h_e|h_p}\left(\{|h_e|\} \mid \{h_p\}\right)$ can be written simply as

$$f_{h_e|h_p}\left(\{|h_e|\} \mid \{h_p\}\right) = \prod_{\mu=1}^{n_\mathrm{R}} f_{h_{e;\mu}|h_{p;\mu}}\left(|h_{e;\mu}||h_{p;\mu}\right). \tag{4.7}$$

Each individual PDF still follows a Rice distribution with the Rician factor $K = |(1-\varrho)h_{p;\mu}(k;l)|^2/\sigma^2_{h_{e;\mu}|h_{p;\mu}}$, where $\varrho = E[\epsilon_{e;\mu}(k;l)h^*_{p;\mu}(k;l)]/\sigma^2_{h_{p;\mu}}(l)$ is the normalized correlation between the predicted channel and the estimation error, and

$$\sigma^2_{h_{e;\mu}|h_{p;\mu}} = \mathsf{Var}\big(h_{e;\mu}(k;l)|h_{p;\mu}(k;l)\big) = \mathsf{E}\left[|\epsilon_{p;\mu}(k;l) - \epsilon_{e;\mu}(k;l) + \varrho h_{p;\mu}(k;l)|^2\right].$$

As in the previous chapter, we choose $\varrho = 0$ in our analysis. Since the estimation and prediction procedures partly are based on the same information, there should be a dependence between $\epsilon_{e;\mu}(k;l)$ and $\epsilon_{p;\mu}(k;l)$. Thus, in contrast to [Cai and Giannakis, 2005], we will consider that the two are correlated and the correlation is

$$\mathsf{E}\left[\epsilon_{p;\mu}\epsilon_{e;\mu}^*\right] = \mathsf{E}\left[h_\mu \epsilon_{e;\mu}^*\right] - \mathsf{E}\left[h_{p;\mu}\epsilon_{e;\mu}^*\right] = \sigma_{e;\mu}^2 - \varrho\sigma_{h_{p;\mu}}^2 = \sigma_{e;\mu}^2.$$

By assumption, ϱ is zero and, due to the orthogonality principle, $\mathsf{E}\left[h_\mu \epsilon_{e;\mu}^*\right] = \sigma_{e;\mu}^2$. Hence, $\sigma_{h_{e;\mu}|h_{p;\mu}}^2$ reduces to (see Appendix C for an alternative proof)

$$\sigma_{h_{e;\mu}|h_{p;\mu}}^2 = \mathsf{E}\left[|\epsilon_{p;\mu}(k;l) - \epsilon_{e;\mu}(k;l)|^2\right] = \sigma_{p;\mu}^2(l) - \sigma_{e;\mu}^2(l).$$

As a result, the Rician factor is simplified to $K = |h_{p;\mu}(k;l)|^2/\sigma_{h_e|h_p}^2$ for $\sigma_{h_{e;\mu}|h_{p;\mu}}^2 = \sigma_{h_e|h_p}^2 \; \forall \mu$.

Averaging (4.4) over the PDF of Eq. (4.7) results in the BER conditioned on the set of predicted channels. That is,

$$\begin{aligned}
\mathrm{BER}\left(M_n \middle| \{h_{p;\mu}\}\right) &= \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n_\mathrm{R}-\text{fold}} \mathrm{BER}\left(M_n \middle| \{|h_{e;\mu}|\}\right) \\
&\quad \times f_{h_{e;\mu}|h_{p;\mu}}\left(\{|h_{e;\mu}|\} \mid \{h_{p;\mu}\}\right) d|h_{e;1}| \cdots d|h_{e;n_\mathrm{R}}| \\
&= \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n_\mathrm{R}-\text{fold}} \prod_{\mu=1}^{n_\mathrm{R}} \exp\left(-A_n(\ell)\mathcal{E}_d|h_{e;\mu}|^2\right) \\
&\quad \times f_{h_{e;\mu}|h_{p;\mu}}\left(|h_{e;\mu}||h_{p;\mu}\right) d|h_{e;1}| \cdots d|h_{e;n_\mathrm{R}}|. \qquad (4.8)
\end{aligned}$$

With the aid of [Gradshteyn and Ryzhik, 2000, Eq. (6.633-4)], the solution of (4.8) may be expressed as

$$\mathrm{BER}\left(M_n \middle| \{h_{p;\mu}\}\right) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) d_n(\ell)^{n_\mathrm{R}} \exp\left(-A_n(\ell) d_n(\ell) \mathcal{E}_d \sum_{\mu=1}^{n_\mathrm{R}} |h_{p;\mu}|^2\right), \qquad (4.9)$$

where $A_n(\ell)$ and $d_n(\ell)$ are as in Eq. (4.5).

The predicted CSNR on each branch is $\hat{\gamma}_\mu = \bar{\mathcal{E}}_d|h_{p;\mu}(k;l)|^2/N_0$. Using the developed prediction procedure, it is well known that the MRC of predicted CSNR is a sum of the branch CSNRs and is obtained as

$$\hat{\gamma} = \frac{\bar{\mathcal{E}}_d}{N_0} \sum_{\mu=1}^{n_\mathrm{R}} |h_{p;\mu}(k;l)|^2 = \frac{\bar{\gamma}_b \bar{\mathcal{E}}_d}{\mathcal{E}} \sum_{\mu=1}^{n_\mathrm{R}} |h_{p;\mu}(k;l)|^2, \qquad (4.10)$$

where $\bar{\gamma}_b = \mathcal{E}/N_0$ is the expected CSNR to be received in one branch. Upon solving this with respect to $\sum|h_{p;\mu}|^2$ and inserting the solution into (4.9) we get

$$\mathrm{BER}(M_n|\hat{\gamma}) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) d_n(\ell)^{n_\mathrm{R}} \exp\left(-\frac{\hat{\gamma} A_n(\ell) d_n(\ell) \mathcal{E}\mathcal{E}_d}{\bar{\gamma}_b \bar{\mathcal{E}}_d}\right). \tag{4.11}$$

The combined predicted CSNR, $\hat{\gamma}$, is fed back to the transmitter via the return channel and is used to decide which code to use. If $\hat{\gamma}_n < \hat{\gamma} \leq \hat{\gamma}_{n+1}$, code n (or constellation of size M_n) is selected for transmission. To find the optimal switching thresholds $\{\hat{\gamma}_n\}_{n=1}^{N}$—which will be used to maximize ASE subject to BER and power constraints—we set (4.11) equal to BER_0 and solve for $\hat{\gamma}$ for different constellation indices n. As in the previous chapter, we have to use a numerical approach to obtain the solutions. The optimal switching thresholds are dependent on the expected subchannel CSNR and these thresholds are different from the switching thresholds obtained for perfect CSI receivers [Øien *et al.*, 2004; Holter, 2005].

From (4.10), we can find the average predicted CSNR as

$$\bar{\hat{\gamma}}(k;l) = \frac{\bar{\mathcal{E}}_d}{N_0} \sum_{\mu=1}^{n_\mathrm{R}} \mathsf{E}\left[|h_{p;\mu}(k;l)|^2\right] = r\bar{\gamma}_b n_\mathrm{R} \tag{4.12}$$

where $r = \bar{\mathcal{E}}_d(1-\sigma_p^2)/\mathcal{E}$. In the second equality it is exploited that $\sigma_{p;\mu}^2 = \sigma_p^2 \ \forall \mu$, because the branches are assumed i.i.d., and the same prediction algorithm is used on all branches. The overall predicted CSNR with MRC of n_R branches will follow a gamma distribution [Øien *et al.*, 2004] with the mean specified by (4.12), i.e., $\hat{\gamma} \sim \mathcal{G}(n_\mathrm{R}, r\bar{\gamma}_b)$. Hence, it is seen that the effective channel can be considered as a Nakagami-n_R channel.

Once the first switching threshold is found, the actual transmit data power, accounting for the fact that no data power is used during the occurrence of outage, can be set to (cf. Eq. (2.13))

$$\mathcal{E}_d = \frac{\bar{\mathcal{E}}_d}{\overline{\Gamma}\left(n_\mathrm{R}, \dfrac{\hat{\gamma}_1}{r\bar{\gamma}_b}\right)}. \tag{4.13}$$

4.3.3 Overall Average BER Performance Analysis

The average BER for the nth constellation is found by averaging (4.11) over the gamma PDF of $\hat{\gamma}$. That is:

$$\begin{aligned}\mathrm{BER}(M_n) &= \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathrm{BER}(M_n|\hat{\gamma}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \\ &= \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \left(\frac{d_n(\ell)\bar{\mathcal{E}}_d}{rA_n(\ell)d_n(\ell)\mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d} \right)^{n_{\mathrm{R}}} \\ &\quad \times \left[\overline{\Gamma}\left(n_{\mathrm{R}}, \hat{\gamma}_n \frac{rA_n(\ell)d_n(\ell)\mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}_b\bar{\mathcal{E}}_d} \right) \right. \\ &\quad \left. - \overline{\Gamma}\left(n_{\mathrm{R}}, \hat{\gamma}_{n+1} \frac{rA_n(\ell)d_n(\ell)\mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}_b\bar{\mathcal{E}}_d} \right) \right]. \end{aligned} \tag{4.14}$$

Having $\mathrm{BER}(M_n)$ at hand, the average BER as formulated in (2.6) can be obtained by combining the Eqs. (2.2) and (4.14) together with

$$P_n = \int_{\hat{\gamma}n}^{\hat{\gamma}n+1} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} = \overline{\Gamma}\left(n_{\mathrm{R}}, \frac{\hat{\gamma}_n}{r\bar{\gamma}_b} \right) - \overline{\Gamma}\left(n_{\mathrm{R}}, \frac{\hat{\gamma}_{n+1}}{r\bar{\gamma}_b} \right), \tag{4.15}$$

which is the probability that constellation n is used.

4.4 Optimization of ASE

As in Section 3.4, we will also use the variance of the estimation error $\sigma^2_{e;\mu}(L-1)$ and the conservative choice of the prediction error variance $\sigma^2_{p;\mu}(L-1)$ here.

4.4.1 ASE Performance Analysis

Using the 4-D trellis codes as earlier, the ASE is now given by

$$\begin{aligned}\mathrm{ASE} &= \sum_{n=1}^{N} R_n P_n \\ &= \frac{L-1}{L} \sum_{n=1}^{N} \left(\log_2(M_n) - \frac{1}{G} \right) \left[\overline{\Gamma}\left(n_{\mathrm{R}}, \frac{\hat{\gamma}_n}{r\bar{\gamma}_b} \right) - \overline{\Gamma}\left(n_{\mathrm{R}}, \frac{\hat{\gamma}_{n+1}}{r\bar{\gamma}_b} \right) \right]. \end{aligned} \tag{4.16}$$

4.4.2 Optimization

To find the optimal solutions, we run the maximization of ASE described in Eq. (3.27). Hence, the optimization algorithm in Algorithm 1 on page 42 is used.

4.5 Average Fade Region Duration

One interesting performance measure in a rate adaptive system is how long time, on average, each different code/constellation is being used. *On average*, this reflects how often the system has to reconfigure its transmission mode. The larger the time between two different modes, the less frequently we need to feed the CSI back to the transmitter, since the channel is in the same fading region within that time interval. Thus, this can help to determine the feedback rate in such a system.

Following the approach in [Goldsmith and Chua, 1997] and using $\bar{\tau}_n$ to denote the nth *average fade region duration* (AFRD), the AFRD can be expressed as

$$\bar{\tau}_n = \frac{P_n}{\chi_{n+1} + \chi_n}, \tag{4.17}$$

where χ_n is the level-crossing rate (LCR) for the nth level; which is given in Eq. (4.18). In our ACM system, (4.17) represents the average time duration of $\hat{\gamma}$ within the interval $\left[\hat{\gamma}_n < \hat{\gamma} \leq \hat{\gamma}_{n+1}\right\rangle$.

For i.i.d. Rayleigh fading subchannels, the LCR of the receive CSNR after MRC is given by Beaulieu and Dong [2003]

$$\chi_n = \frac{\sqrt{2\pi} f_d}{\Gamma(n_{\mathrm{R}})} \left(\frac{\hat{\gamma}_n}{r\bar{\gamma}_b n_{\mathrm{R}}}\right)^{n_{\mathrm{R}}-\frac{1}{2}} \exp\left(-\frac{\hat{\gamma}_n}{r\bar{\gamma}_b n_{\mathrm{R}}}\right). \tag{4.18}$$

If we have perfect prediction, i.e. $\hat{\gamma}$ coincides with γ, and $n_{\mathrm{R}} = 1$, then (4.18) simplifies to [Goldsmith and Chua, 1997, Eq. (35)].

4.6 Numerical Analysis

We will consider the case where we have a number of receive antennas $n_{\mathrm{R}} \in \{1, 2, 4\}$. Otherwise, the other parameters are the same as those in Section 2.7. Even though MRC is employed at the receiver, the plots are generated as functions of per branch CSNR. This is done since we 1) want to make a comparison between schemes based on fixed transmit power rather than fixed receive power, and 2) we can more easily compare the results of this work to other previous works. Also, we can see directly the improvement for a given overall transmit power.

Figures 4.2 and 4.3 show the optimum pilot symbol period and the optimum fraction of power allocation to the pilot symbols, respectively, for different number of branches. As can be seen in Figure 4.2, when the power distribution between pilot and data symbol is optimized, fewer pilot symbols are needed. The spacing between two pilot symbols is also larger when

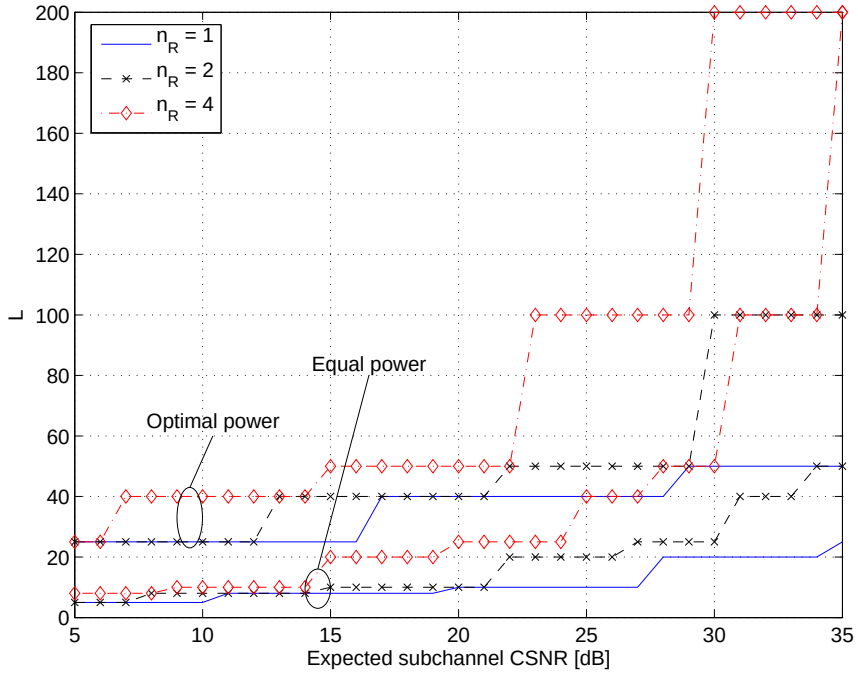


FIGURE 4.2: Optimum pilot spacing L as a function of expected subchannel CSNR when the power is equally and optimally allocated between pilot and data symbols. The number of antennas is $n_{\mathrm{R}} = 1, 2$, and 4. Carrier frequency $f_c = 2$ GHz; Doppler frequency $f_d = 200$ Hz; symbol duration $T_s = 5$ μs; system delay $\tau = 1$ ms; and $\mathrm{BER}_0 = 10^{-5}$.

there are more antennas available to combine. The power allocated to pilot symbols is lower for higher diversity orders for both equal and optimal power regimes (cf. Figure 4.3). This is due to the array gain and to the fact that MRC is optimal in the sense that it maximizes the output CSNR. Thus, less pilot power is needed when more antennas are available. As a consequence of a larger pilot spacing, the optimal power allocation scheme puts more power on the pilot than when power is equally allocated.

The pilot power increases again at very high average CSNR for the optimal case. This corresponds to the steeply increasing pilot period in the same CSNR region. In order to have a good channel prediction and estimation such that the system can rely on—so that the BER requirement is maintained—more power is allocated to the pilots in this case.

Because of the finite number of codes, the ASE reaches a ceiling when the CSNR grows. As expected, the ASE is higher when the transmit power and the pilot spacing are optimized and when we have more antennas

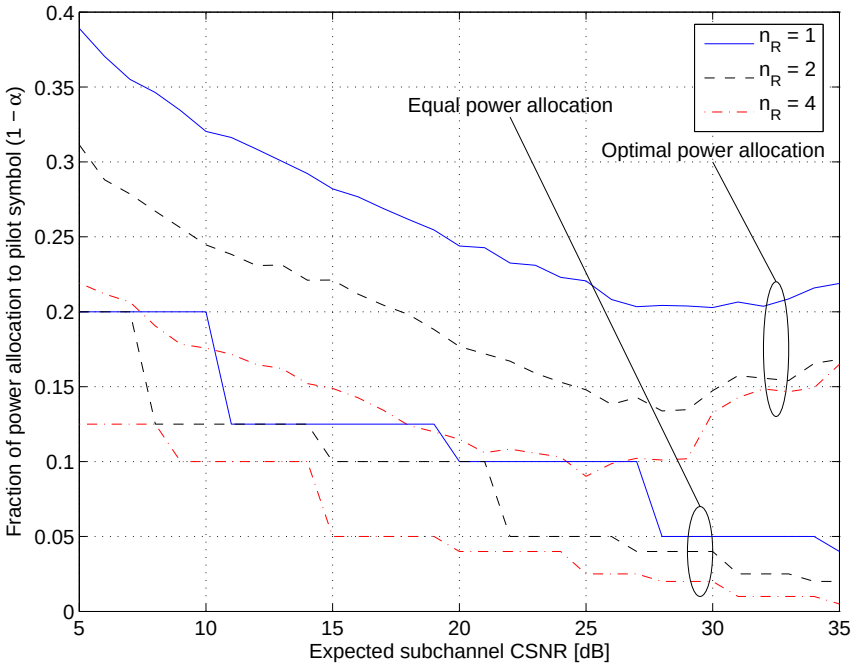


FIGURE 4.3: Fraction of power allocation to pilot symbols (i.e., $1-\alpha$) versus expected CSNR on each branc when the period L is optimal. The number of antennas is $n_R = 1, 2$, and 4. Carrier frequency $f_c = 2$ GHz; Doppler frequency $f_d = 200$ Hz; symbol duration $T_s = 5$ μs; system delay $\tau = 1$ ms; and $\text{BER}_0 = 10^{-5}$.

available. This is due to the fact that the pilot period is larger for higher n_R as well as to the array gain. This is shown clearly in Figure 4.4.

The analysis in this chapter is based on a looser approximation of BER performance on AWGN channels. However, the ASE performance is very close to that when the tight approximation is used (as in the previous chapter). This is seen by comparing Figure 3.4 and Figure 4.4 for $n_R = 1$.

We can not directly compare our results with the results in [Øien *et al.*, 2004] since the two systems are running on different sets of parameters. To make the comparison possible, we have run the system of Øien *et al.* with our system parameters and the plots are depicted in Figure 4.5. Using the notation in that paper, the normalized delay is: $j \cdot f_D T_s$ where f_D is the maximum Doppler frequency and j is the index of the symbol to be predicted. If this normalized delay is equal to $\tau f_d = 0.2$ as in our paper, j must be 200. That delay is utilized in the subfigure (a) of Figure 4.5. With that delay, the system of Øien *et al.* will not be able to operate acceptably

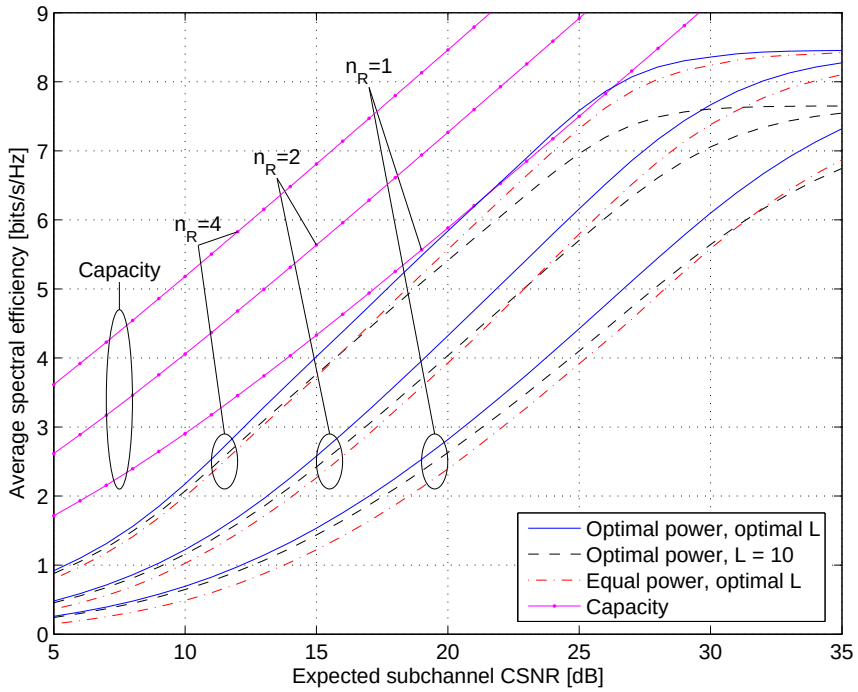
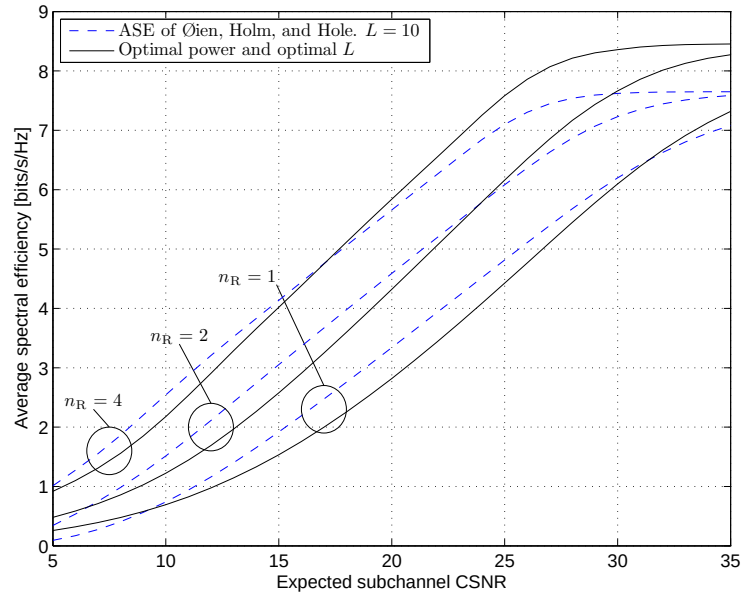


FIGURE 4.4: Average spectral efficiency for both optimal parameters and some fixed parameters together with the channel capacity of Rayleigh fading plotted against subchannel's CSNR. n_{R} is the number of receive antennas and is set to 1, 2, and 4. Carrier frequency $f_c = 2$ GHz; Doppler frequency $f_d = 200$ Hz; symbol duration $T_s = 5$ μs; system delay $\tau = 1$ ms; and $\mathrm{BER}_0 = 10^{-5}$.

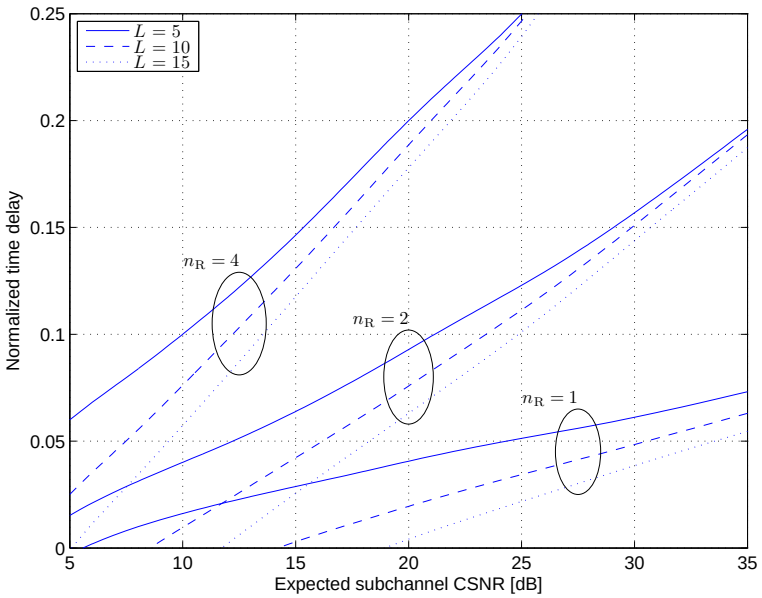
with one or even with two receive antennas. Combining four antennas, the system start functioning at an average CSNR = 21 dB for $L = 10$ (cf. subfigure (b) of Figure 4.5). To conclude, our system operates satisfactorily and is superior to the one of Øien *et al.* for the whole range of CSNRs under consideration and for the considered delay. This holds especially at the high-CSNR region where fixing of pilot spacing clearly is a disadvantage, because the channel can also be satisfactorily tracked with a larger pilot period (see also Figure 4.4).

As can be seen in Figure 4.6, the theoretical $\overline{\mathrm{BER}}$ curves—produced from (2.6) with the derived parameters in this chapter—are always lower than $\mathrm{BER}_0 = 10^{-5}$. The average BER performance corresponding to optimal L and equal power allocation are almost the same as those in the figure. Hence, we do not plot them here.

Depicted in Figure 4.7 is the probability of outage. Clearly, as can be



(a) ASE of our system and of [Øien *et al.*, 2004]



(b) Regions for which system performance in [Øien *et al.*, 2004] is acceptable: the region below and to the right of the curves

FIGURE 4.5: Comparison of our system with the system in [Øien *et al.*, 2004]. Carrier frequency $f_d = 2$ GHz; Doppler frequency $f_d = 200$ Hz; symbol duration $T_s = 5$ μs; $\text{BER}_0 = 10^{-5}$; and number of antennas is 1, 2, and 4. Whilst we consider both estimation and prediction errors the system of Øien *et al.* assumes perfect channel estimation.

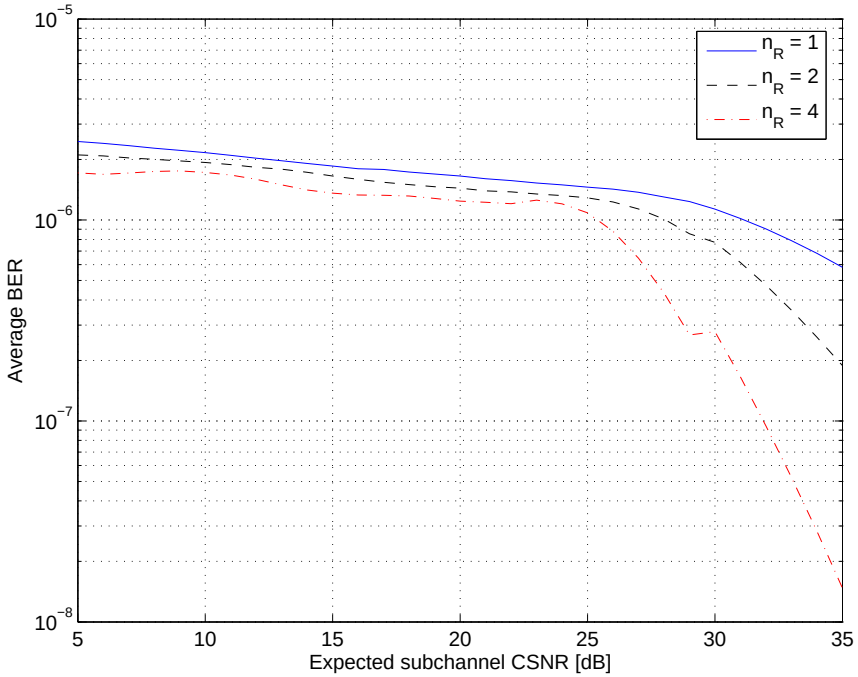


FIGURE 4.6: $\overline{\text{BER}}$ for optimal power and optimal L with different number of receive antennas ($n_R = 1, 2$, and 4).

observed, the outage probability is somewhat lower when both power allocation and pilot symbol spacing are optimal compared to when only pilot spacing L is optimal and the power is equally distributed between pilot and data symbols. Also, the curves demonstrate that diversity mitigates fading such that deep fades are averaged out, such that outage is occurred with lower probability.

In Figure 4.8, the constellation selection probabilities of different constellations are plotted as function of average subchannel CSNR. With increasing diversity order, the curves become less overlapped with each other, and the larger constellations are chosen with higher probabilities at lower average CSNR. Over the whole range of average CSNR under consideration, and even with $n_R = 4$, there are always constellations that are used with non-negligible probabilities, compared to the dominant constellation (the one which is selected with highest probability). There is an inherent trade-off between the use for diversity and the need of adaptivity. However, the results show that there still is benefit of having both adaptation and diversity in the same system.

The results for the AFRD are found in Tab. 4.1. The parameters needed

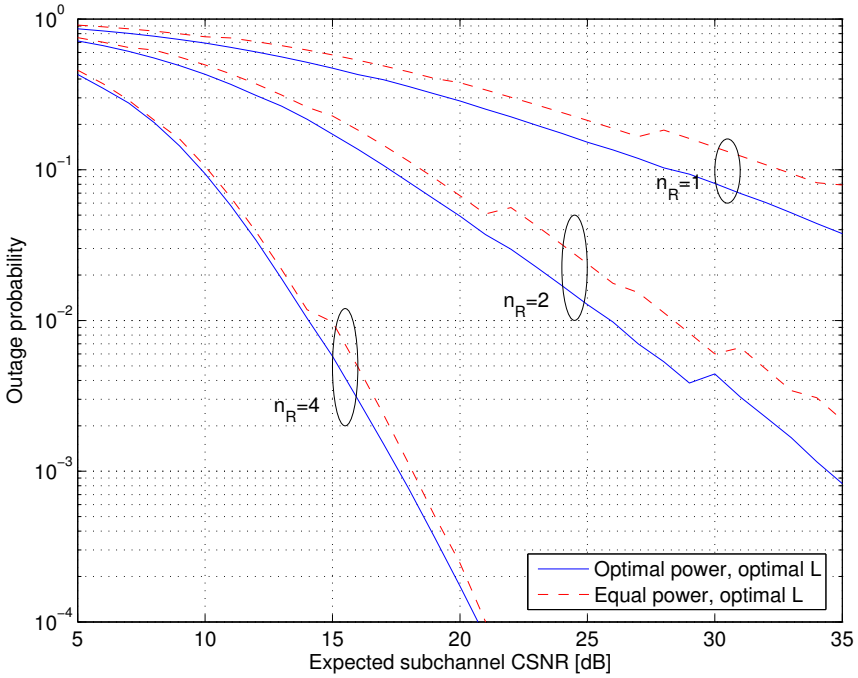


FIGURE 4.7: Probability of no transmission as a function of average subchannel CSNR for $n_R \in 1, 2, 4$ and for both optimal power, optimal L and equal power, optimal L cases.

for calculation of the values in the fields denoted by ∘ are very small; hence, due to the numerical precision, they can not be calculated correctly. Thus, we set those to zero. On the other hand, the • denotes an undefined number. This is due to a zero-by-zero fraction (cf. Eq. (4.17)) since the probability of outage in this channel quality level is essentially zero, and thus the switching threshold $\hat{\gamma}_1 = 0$. As expected, the numerical example indicates that, for a given number of receive antennas, more time is spent on larger constellations when the channel quality is getting better. The time is remarkably greater with multiple reception of receive signals since the channel becomes more stable. At the extreme case ($n_R = 4$ and $\bar{\gamma}_b = 30$ dB), the system uses almost only the largest constellation. Note that one millisecond (ms) corresponds to 200 channel symbols. With this in mind, the values in Tab. 4.1 seem to indicate that the channel is slow enough for our model to be quite accurate in most of the cases.

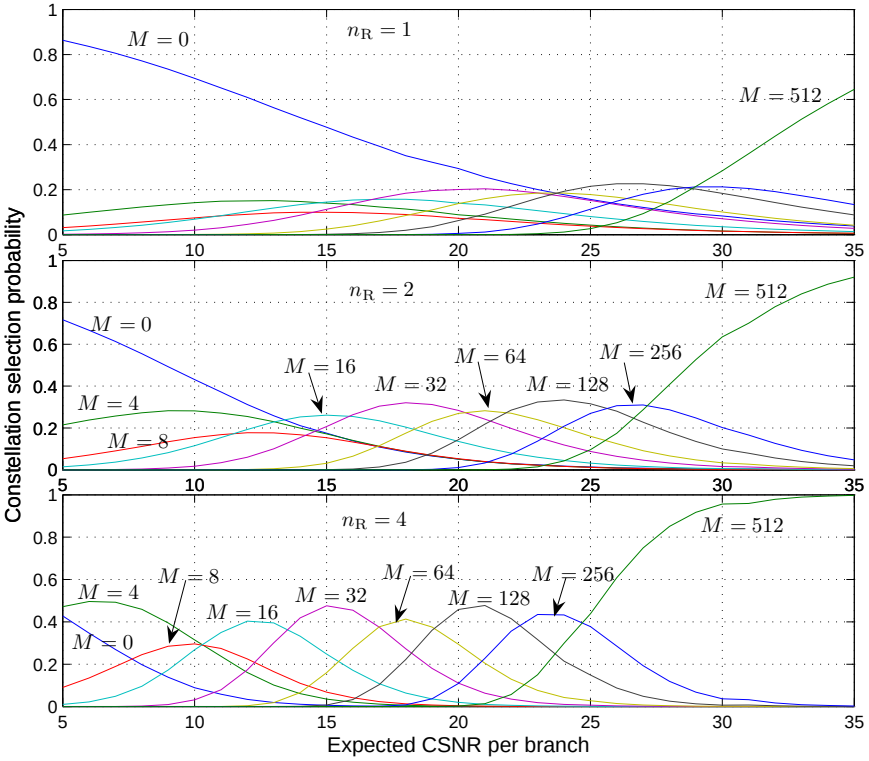


FIGURE 4.8: Constellation selection probability as a function of average subchannel CSNR for $n_R = 1$ (top panel), $n_R = 2$ (middle panel), and $n_R = 4$ (bottom panel). This is plotted for optimal L and optimal power allocation. Note that the curve for $M = 0$ corresponds to outage probability.

4.7 Conclusions

We have investigated an ACM system where receive diversity is implemented by means of MRC. The ASE is a result of an optimization of the pilot symbol period, and the power distribution between pilot and data symbols. The ASE is substantially higher with spatial diversity and is considerably increased compared to when the pilot spacing is fixed, and when the power is equally allocated to pilot and data symbols. The increased ASE is obtained without losing BER performance; especially in the high CSNR regions. This gain is due to the fact that the rate of the pilot symbols is substantially reduced in that region compared to previous fixed solution. Thus, fixing of the pilot symbol rate is clearly a disadvantage.

The optimal power allocation scheme is the one which puts more power

TABLE 4.1: Average time duration (measured in millisecond) each code is being used for different subchannel qualities $\bar{\gamma}_b$. $M_n = 0$ corresponds to no transmission. The number of receive antennas is $n_R = 1, 2$, and 4.

		$n_R = 1$			$n_R = 2$			$n_R = 4$		
n	M_n	$\bar{\gamma}_b = 10$	$\bar{\gamma}_b = 20$	$\bar{\gamma}_b = 30$	$\bar{\gamma}_b = 10$	$\bar{\gamma}_b = 20$	$\bar{\gamma}_b = 30$	$\bar{\gamma}_b = 10$	$\bar{\gamma}_b = 20$	$\bar{\gamma}_b = 30$
0	0	4.1973	1.3727	0.6093	4.3155	1.3373	0.6384	3.8549	1.0934	•
1	4	0.5392	0.2239	0.0605	1.5816	0.5552	0.1986	4.4477	1.2524	0.4122
2	8	0.4343	0.1808	0.0534	1.2140	0.4234	0.1289	3.3198	0.9192	0.6593
3	16	0.7831	0.3645	0.1094	1.9985	0.8355	0.2372	5.3265	1.6140	0.8472
4	32	0.9312	0.6543	0.1972	1.8354	1.5230	0.3937	3.8248	2.8328	0.9919
5	64	0.7141	0.7742	0.2538	1.2303	1.8196	0.4948	2.2393	3.7765	1.0316
6	128	0.5386	1.0139	0.4465	0.8844	2.3699	0.8748	1.5291	6.4712	1.5797
7	256	0.3808	0.8277	0.5884	○	1.7339	1.1909	○	4.6739	2.1234
8	512	○	0.6161	1.7492	○	1.2036	5.7723	○	2.8481	42.1085

on pilot symbols (compared to the equal power allocation strategy) as a response to the decreased pilot symbol rate. The pilot spacing increases, and the pilot power decreases for higher diversity orders and for both optimal and non-optimal power distribution schemes.

Chapter 5

Impact of Spatial Correlation

This chapter takes a step further to include spatial correlation in the system analyzed in the previous chapter. It is based on a paper published in *Norwegian Signal Processing Symposium (NORSIG)* [Duong, Wingar, and Øien, 2005], a paper to appear in *IEEE Vehicular Technology Conference (VTC-Fall)* [Duong, Holter, and Øien, 2006b], and a journal letter submitted to *IEEE Transactions on Vehicular Technology* [Duong, Holter, and Øien, 2006a].

5.1 Introduction

In order to have uncorrelated subchannels, the receive antennas must be spaced at least half a wavelength apart [Rappaport, 2002]. However, from an experimental point of view, about 10–20 wavelengths separation between the antenna elements is sometimes required to provide sufficient spatial decorrelation at the outdoor base station [Winters, 1998]. Similarly, sufficient spatial decorrelation between the receive antennas at the MS or between the antennas of an indoor BS is obtained by separating the antennas by quarter of a wavelength [Winters, 1998].

Due to physical size limitations, the antennas might however be spaced quite close to each other. This implies that there exists some correlation between these antennas. Insufficient scattering around the base station also leads to correlated branches [Paulraj, Nabar, and Gore, 2003]. In such a system, spatial correlation is playing an important role when analyzing the system performance, since it is well known that correlation degrades performance [Ratnarajah, 2006]. However, this degradation can be reduced if the correlation is exploited properly.

In this chapter, we assume that the branches are identically distributed but spatially correlated. The chapter is in general divided into two parts.

Both parts are dealing with a spatially correlated fading environment, but the effect of spatial correlation is not taken into account and not exploited in the prediction process in the first part. I.e. we are implicitly assuming that a system designed for a spatially uncorrelated subchannels environment in a correlated subchannels environment. Then, in the second part, a "space-time predictor" is derived where the spatial correlation is assumed known and, thus, is taken into account.

Only channel prediction is considered here, since it is necessary for system adaptation and since channel estimation can be achieved with high accuracy using an optimal noncausal estimator with reasonable delay [Meyr *et al.*, 1998, p. 660]. In particular, according to [Øien and Hole, 2001, Eq. (21)], the estimation error variance of a single noncausal Wiener filter operating on a slowly varying Rayleigh fading channel is equal to

$$\sigma_e^2 = \frac{\Omega 2WLN_0}{\mathcal{E}\Omega + 2WLN_0} \tag{5.1}$$

where $\mathcal{E}$ is again the average transmit power and $W = f_d T_s$. When channel estimation is done on a per branch basis and applying Nyquist condition on pilot spacing we get $\sigma_e^2 \leq \Omega/\bar{\gamma}_b = 1/\bar{\gamma}_b$ [Holm, 2002, Eq. (3.7)]. Thus, unless the CSNR is very low, the imperfection of channel estimation is negligible compared to the prediction error and channel estimation can be assumed to be perfect. Latter on, in Chapter 7, we will in fact explicitly discuss how this simplification affects a the system's performance for the spatially uncorrelated case.

The rest of the chapter is organized as follows. Section 5.2 describes the space-time correlation model used in the rest of the chapter. Section 5.3 introduces the statistics of the combined signals. Mathematical and numerical analysis of the system performance when the subchannels are independently predicted are also given there. Then the optimal "space-time predictor" is derived in Section 5.4 and the numerical results obtained by using this predictor are also presented. Concluding remarks are given in Section 5.5.

5.2 Space-Time Correlation Model

When the mobile terminal moves away from (or toward) the base station along the mobile position angle ϖ (as illustrated in Figure 5.1)—i.e. when $\zeta = \varpi$ (or $\zeta = \varpi + \pi$)—the space-time correlation function of the fading gain is approximately equal to the product of the spatial and temporal correlation function [Chen, Fitz, Li, and Zoltowski, 2004]. According

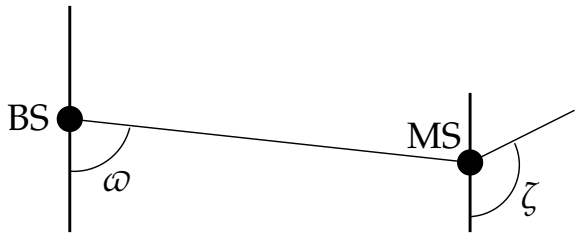


FIGURE 5.1: Illustration of the mobile moving angle ζ and the mobile position angle ϖ.

to [Smith and Shafi, 2004], such a model is adequate for gauging *average* system behavior. On the other hand, the exact and general space-time correlation function can be found in [Lee, 1970; Abdi and Kaveh, 2000; Byers and Takawira, 2004]. Here we are interested only in average behavior, and thus, with reference to [Chen *et al.*, 2004], we assume that the space-time correlation between antenna branch μ and ν at lag m is

$$\begin{aligned}\mathsf{E}\left[h_\mu(n)h_\nu^*(n+m)\right] &\approx \mathsf{E}\left[h_\mu(n)h_\nu^*(n)\right]\mathsf{E}\left[h_\mu(n)h_\mu^*(n+m)\right] \\ &= \rho_{h,s}(|\mu-\nu|)\cdot\rho_{h,t}(m),\end{aligned} \tag{5.2}$$

where $\rho_{h,s}$ and $\rho_{h,t}$ denotes the spatial and temporal correlation of the fading gain, respectively.

It should be noted that, when using the Jakes scattering model, $\rho_{h,t}$ is real-valued and given by the zeroth order Bessel function of the first kind: $J_0(2\pi f_d m)$ [Stüber, 1996].

5.2.1 Spatial Correlation Models

Let the instantaneous CSNR of the μth branch be $\gamma_\mu = \mathcal{E}_d\beta_\mu^2/N_0$ where $\beta_\mu \triangleq |h_\mu(k;l)|$ and let the power correlation (which is real) between any two adjacent branches be

$$\rho_s = \frac{\mathsf{Cov}\left(\gamma_{\mu+1},\gamma_\mu\right)}{\sqrt{\mathsf{Var}\left(\gamma_{\mu+1}\right)\mathsf{Var}\left(\gamma_\mu\right)}} = \frac{\mathsf{Cov}\left(\beta_{\mu+1}^2,\beta_\mu^2\right)}{\sqrt{\mathsf{Var}\left(\beta_{\mu+1}^2\right)\mathsf{Var}\left(\beta_\mu^2\right)}}. \tag{5.3}$$

In general, the correlation of the fading gains between the branches is complex. Following [Alouini *et al.*, 2001; Hasna, Alouini, and Simon, 2001], the power correlation is equal to absolute square of the correlation of the fading gain. Thus, ρ_s can be calculated as

$$\rho_s = |\rho_{h,s}(1)|^2 = |c-\jmath d|^2, \tag{5.4}$$

where c and d are the normalized real and imaginary parts of $\rho_{h,s}(1)$, respectively.

Given the power correlation ρ_s, the two spatial correlation models used in this chapter are:

1. Constant correlation

$$\rho_{\mu\nu} = \begin{cases} 1 & \text{for } \mu = \nu \\ \rho_s & \text{for } \mu \neq \nu. \end{cases} \tag{5.5}$$

2. Exponential correlation

$$\rho_{\mu\nu} = \rho_s^{|\mu-\nu|}. \tag{5.6}$$

In (5.5) and (5.6) μ and ν are the antenna indices: $\mu, \nu \in [1, \cdots, n_{\mathrm{R}}]$. The exponential correlation model may be applied to an equidistant array (linear array) of antenna elements, while the constant correlation model corresponds to an array of closely spaced receive antennas or three antennas placed on an equilateral triangle [Simon and Alouini, 2005].

5.3 Link Adaptation in Spatially Correlated Antenna Diversity: Independently Predicted Branches

In this part of the chapter, we apply the system designed for spatially uncorrelated receive antennas to the correlated one. This is clearly suboptimal, but it gives a kind of benchmark quantifying how much we can gain by predicting the branches jointly (which is done in the next section). It also gives performance to be expected when the spatial correlation is unknown and thus can not be exploited in the prediction procedure.

5.3.1 The Combined Signal Statistics

The output signals can still be written as in Eqs. (4.1) and (4.2). Based on that, the total detected data CSNR—when assuming perfect estimation—can be written as[1]

$$\gamma_d(k;l) = \frac{\mathcal{E}_d}{N_0} \sum_{\mu=1}^{n_{\mathrm{R}}} |h_\mu(k;l)|^2. \tag{5.7}$$

[1] We use γ_d to denote the instantaneous true CSNR when estimation is assumed to be perfect. This is to distinguish it from γ where the estimation error is taken into account.

Thus, letting $r_d = \mathcal{E}_d/\mathcal{E}$, the total average CSNR is

$$\bar{\gamma}_d = \frac{\mathcal{E}_d \sum_{\mu=1}^{n_R} \mathsf{E}\left[|h_\mu(k;l)|^2\right]}{N_0} = \frac{n_R \mathcal{E}_d}{N_0} = r_d \bar{\gamma}_b n_R. \tag{5.8}$$

Furthermore, the instantaneous and the expected predicted CSNR still can be expressed as (4.10) and (4.12), respectively. The results are still based on the fact that the prediction error variance is the same on all branches.

When the subchannel gains $\beta_\mu \triangleq |h_\mu(k;l)|$ and $\hat{\beta}_\mu \triangleq |h_{p;\mu}(k;l)|$ are Rayleigh distributed and mutually uncorrelated, both γ_d and $\hat{\gamma}$ will be a sum of uncorrelated exponentially distributed RVs or, equivalently, they will follow gamma distributions. However, when there exists some correlation between the receiving branches, the output CSNRs are not gamma distributed anymore and their exact PDFs are given in [Aalo, 1995; Alouini *et al.*, 2001].[2] Note that these PDFs do not belong to standard distributions and, as a result, we can not model the correlation between γ_d and $\hat{\gamma}$ exactly. In [Mun, Kang, and Park, 1999], the PDF for the CSNR of arbitrarily correlated Nakagami-m fading channels (Rayleigh fading if $m = 1$) is shown to be well approximated by a gamma distribution, where the two first moments are equal to those of the exact PDF. With reference to that paper, we let γ_d and γ be gamma distributed with shape factor m_d and scale factors θ_1 and θ_2, respectively. Using the shorthand notation introduced in Definition 1 on page 10, $\gamma_d \sim \mathcal{G}(m_d, \theta_1)$ and $\hat{\gamma} \sim \mathcal{G}(m_d, \theta_2)$. Hence the joint distribution is still described by the bivariate gamma distribution [Holm, 2002], i.e., $(\gamma_d, \hat{\gamma}) \sim \mathcal{G}(m_d, \theta_1, \theta_2, \rho)$, where ρ is the correlation coefficient between the two variables γ_d and $\hat{\gamma}$:

$$\rho = \frac{\mathsf{Cov}(\gamma_d, \hat{\gamma})}{\sqrt{\mathsf{Var}(\gamma_d)\,\mathsf{Var}(\hat{\gamma})}} = \frac{\mathsf{Cov}(\beta^2, \hat{\beta}^2)}{\sqrt{\mathsf{Var}(\beta^2)\,\mathsf{Var}(\hat{\beta}^2)}}. \tag{5.9}$$

It can be shown that [Holter and Øien, 2006]

$$\rho = \frac{|\boldsymbol{\omega}_p^{\mathsf{H}} \mathbf{r}_p|^2}{\eta} \tag{5.10}$$

where $\boldsymbol{\omega}_p$ is the vector of predictor filter coefficients and

$$\eta = \boldsymbol{\omega}_p^{\mathsf{H}} \mathbf{R}_p \boldsymbol{\omega}_p + \frac{\|\boldsymbol{\omega}_p\|^2}{(1-\alpha)\bar{\gamma}_b L}. \tag{5.11}$$

[2] Only the exact PDF for constant correlation is given in [Aalo, 1995]. The PDF corresponding to exponential correlation is an approximation.

TABLE 5.1: Parameters of the gamma PDFs of γ_d and $\hat{\gamma}$ where ρ_s is the power correlation between the two variables

Correlation model	m_d	θ
Constant	$\frac{n_R}{1+\rho_s(n_R-1)}$	$\bar{\gamma}_b(1+\rho_s(n_R-1))$
Exponential	$\frac{n_R}{1+\frac{2\rho_s}{1-\rho_s}\left(1-\frac{1-\rho_s^{n_R}}{n_R(1-\rho_s)}\right)}$	$\bar{\gamma}_b\left(1+\frac{2\rho_s}{1-\rho_s}\left(1-\frac{1-\rho_s^{n_R}}{n_R(1-\rho_s)}\right)\right)$

In (5.9) and (5.10), $\mathbf{R}_p$ and $\mathbf{r}_p$ denotes the covariance matrix and covariance vector of the channel corresponding to the prediction case, respectively. It is noted that (5.10) is identical to the result obtained for the spatially uncorrelated case ($\rho_s = 0$) [Holm, 2002, Eq. (3.21)]. This is expected since prediction is performed independently on each branch as though the branches were uncorrelated.

Note that θ_1 and θ_2 can respectively be written as $\theta_1 = r_d\theta$ and $\theta_2 = r\theta$, and that $\bar{\gamma}_d = m_d\theta_1$ and $\bar{\hat{\gamma}} = m_d\theta_2$. Both m_d and θ are given in Tab. 5.1 and are derived from the amount of fading [Holter and Øien, 2005]. It has been shown in [Holter and Øien, 2006] that system performance with the approximated PDFs is identical or very close to the performance obtained with the exact PDFs, except at very high CSNRs.

Using the gamma approximation, the actual data power of Eq. (2.13) is written as

$$\mathcal{E}_d = \frac{\bar{\mathcal{E}}_d}{\overline{\Gamma}\left(m_d, \frac{\hat{\gamma}_1}{\theta_2}\right)}.$$

5.3.2 System Performance Analysis

BER Analysis

For a given choice of code, the system is required to operate with an instantaneous BER[3] below a predefined value BER_0. Thus, we need an expression for $\mathrm{BER}(M_n|\hat{\gamma})$, which can be found as

$$\mathrm{BER}(M_n|\hat{\gamma}) = \int_0^\infty \mathrm{BER}(M_n|\gamma) f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma}) d\gamma \tag{5.12}$$

where $\mathrm{BER}(M_n|\gamma)$ is as (2.8). The instantaneous CSNR is given by (5.7). Using Bayes' rule, $f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma}) = f_{\gamma\hat{\gamma}}(\gamma,\hat{\gamma})/f_{\hat{\gamma}}(\hat{\gamma})$, and integrating (5.12) with

[3] Instantaneous with respect to the predicted CSNR

the variable γ replaced by γ_d, the result of $\mathrm{BER}(M_n|\hat{\gamma})$ is written as

$$\mathrm{BER}(M_n|\hat{\gamma}) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \left(\frac{M_n}{\Phi}\right)^{m_d} \exp\left(-\hat{\gamma}\frac{\rho\theta_1 b_n(\ell)}{\theta_1\Phi}\right) \tag{5.13}$$

where $\Phi = M_n + \theta_1 b_n(\ell)(1-\rho)$. By once again solving $\mathrm{BER}(M_n|\hat{\gamma}) = \mathrm{BER}_0$ we obtain the switching thresholds $\{\hat{\gamma}_n\}_{n=1}^{N}$.

The average BER is calculated as (2.6) where

$$\begin{aligned}\mathrm{BER}(M_n) &= \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathrm{BER}(M_n|\hat{\gamma}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \\ &= \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \left(\frac{M_n}{M_n + b_n(\ell)\theta_1}\right)^{m_d} \\ &\quad \times \left[\overline{\Gamma}\left(m_d, \hat{\gamma}_n \frac{M_n + b_n(\ell)\theta_1}{\theta_2\Phi}\right) - \overline{\Gamma}\left(m_d, \hat{\gamma}_{n+1}\frac{M_n + b_n(\ell)\theta_1}{\theta_2\Phi}\right)\right]\end{aligned} \tag{5.14}$$

and

$$P_n = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} = \overline{\Gamma}\left(m_d, \frac{\hat{\gamma}_n}{\theta_2}\right) - \overline{\Gamma}\left(m_d, \frac{\hat{\gamma}_{n+1}}{\theta_2}\right). \tag{5.15}$$

ASE Analysis

Given R_n and P_n as (2.2) and (5.15), respectively, the ASE is expressible as

$$\begin{aligned}\mathrm{ASE} &= \sum_{n=1}^{N} R_n P_n \\ &= \frac{L-1}{L} \sum_{n=1}^{N} \left(\log_2(M_n) - \frac{1}{2}\right) \left[\overline{\Gamma}\left(m_d, \frac{\hat{\gamma}_n}{\theta_2}\right) - \overline{\Gamma}\left(m_d, \frac{\hat{\gamma}_{n+1}}{\theta_2}\right)\right].\end{aligned} \tag{5.16}$$

The optimization algorithm in Algorithm 1 on page 42 is invoked to find the optimal solutions.

5.3.3 Numerical Calculation

The same set of parameters as in Section 2.7 are used. In addition, we let the spatial correlation between the branches be in the set $\{0.2, 0.7\}$. We also consider perfect estimation, so no estimator is involved. Furthermore, since the branches are assumed to be identical the subchannels' average CSNR is also assumed to be the same on all the branches.

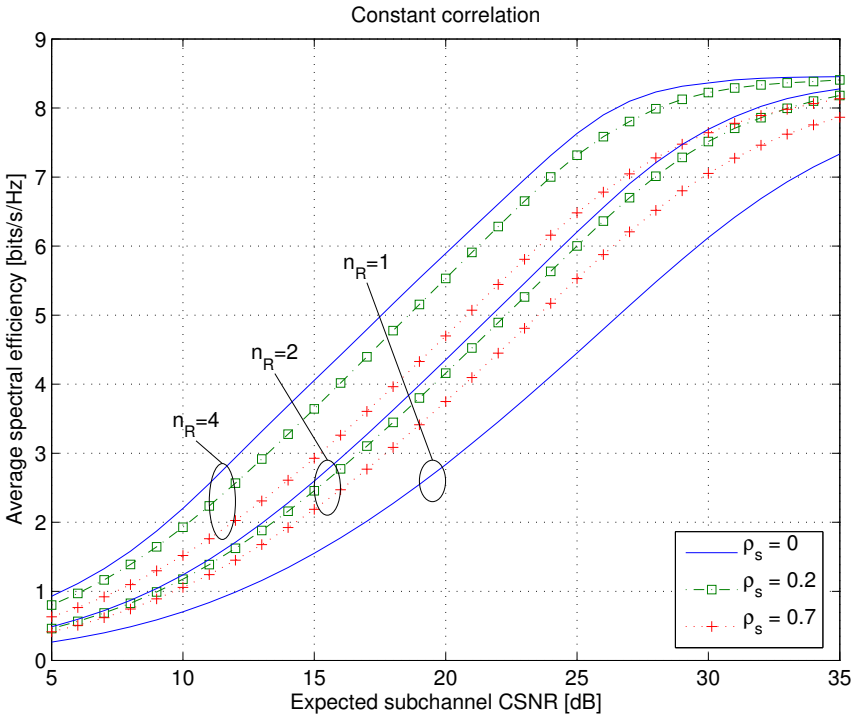


FIGURE 5.2: ASE as a function of subchannel CSNR for optimal L and optimal power allocation. The correlation between every subchannels is the same (constant correlation model) and is equal to $\rho_s = 0$, 0.2, and 0.7, respectively.

The plots in Figures 5.2, 5.4, and 5.5 correspond to the constant correlation between the branches. The results for exponential correlation are given in Figures 5.3, 5.6, and 5.7.

We see from Figures 5.2 and 5.3 that ASE is reduced with increased spatial correlation between the antennas. This is in principle expected, since the advantages of having antenna diversity become smaller with larger spatial correlation. Moreover, constant correlation causes more performance degradation than exponential correlation. This is due to the fact that, in the latter case, the correlation between one antenna with the other antennas decreases exponentially with distance between them, while the correlation is the same between any two antenna branches in the former case. Note that for $n_R = 2$ the two correlation models give identical results.

When all branches are fully correlated, the diversity gain will disappear and the effective channel approaches a channel with no spatial diversity, but with a higher average CSNR due to the array gain. This implies that

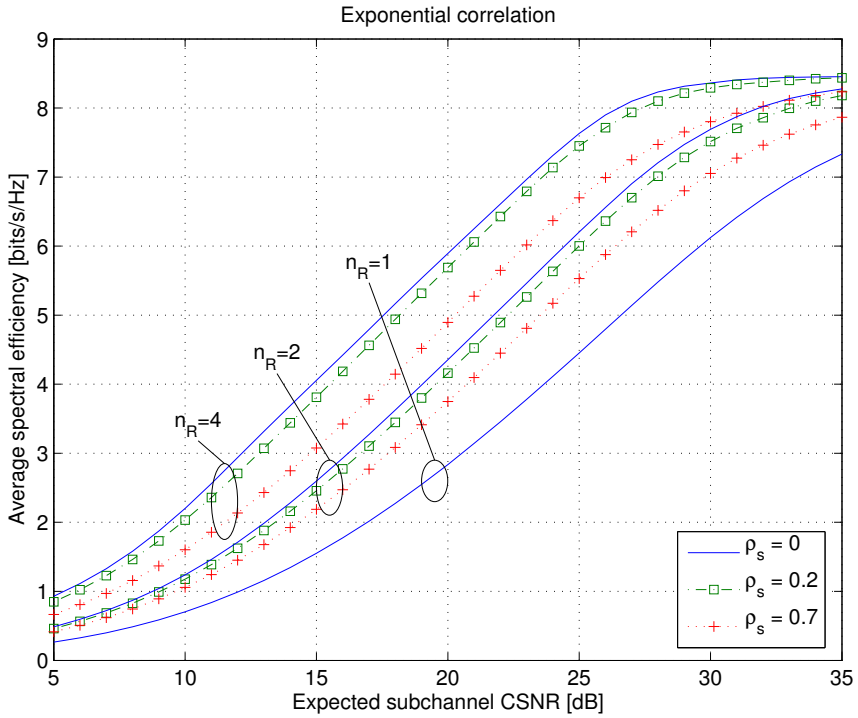


FIGURE 5.3: ASE as a function of subchannel CSNR for optimal L and optimal power allocation. The correlation between the subchannels is decreasing with the distance between them (exponential correlation model) and is equal to $\rho_s = 0$, 0.2, and 0.7, respectively.

increased spatial correlation will affect the pilot symbol period in such a way that it behaves more like in a single receive antenna system ($n_R = 1$). This is clearly shown in Figures 5.4 and 5.6.

Evidently, the pilot spacing L moves towards the curve corresponding to $n_R = 1$ when spatial correlation goes from 0 to 0.7. This happens for both constant correlation and exponential correlation. Furthermore, L decreases faster (more frequent pilot insertion) with constant correlation compared to with exponential correlation, and this also explains why ASE drops more in this case.

Having the pilot spacing L as in the one-branch case, it should then not be necessary to have high pilot power due to the array gain. Studying Figures 5.5 and 5.7 it is observed that this however is not the case, and that the pilot power in fact is even higher than for the one-branch system. As shown in the next part, this behavior is due to the suboptimality in the prediction procedure used in this section, and it can be remedied by taking

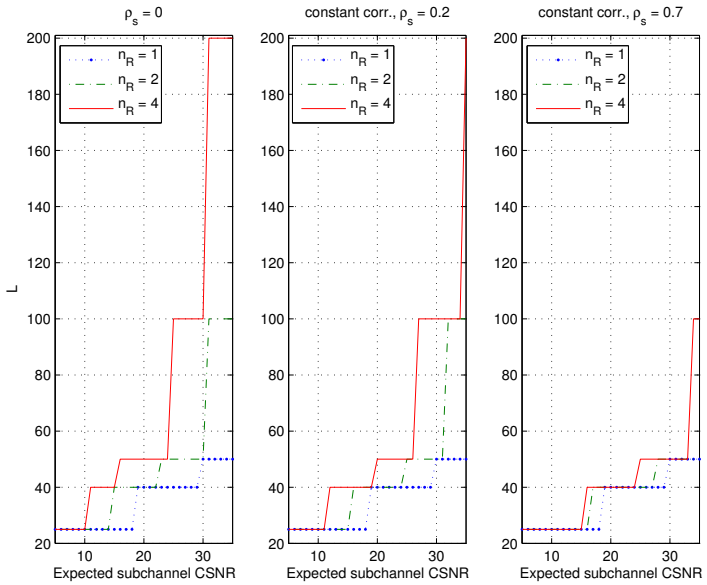


FIGURE 5.4: Optimal pilot spacing as a function of subchannel CSNR with equal correlation for $\rho_s = 0$ (left panel), $\rho_s = 0.2$ (middle panel), and $\rho_s = 0.7$ (right panel), respectively.

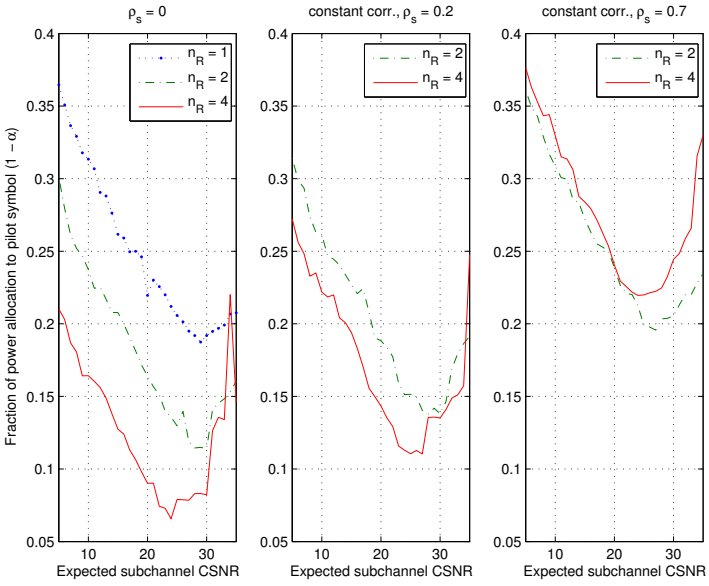


FIGURE 5.5: Optimal fraction of power allocated to pilot symbol as a function of subchannel CSNR with equal correlation for $\rho_s = 0$ (left panel), $\rho_s = 0.2$ (middle panel), and $\rho_s = 0.7$ (right panel), respectively.

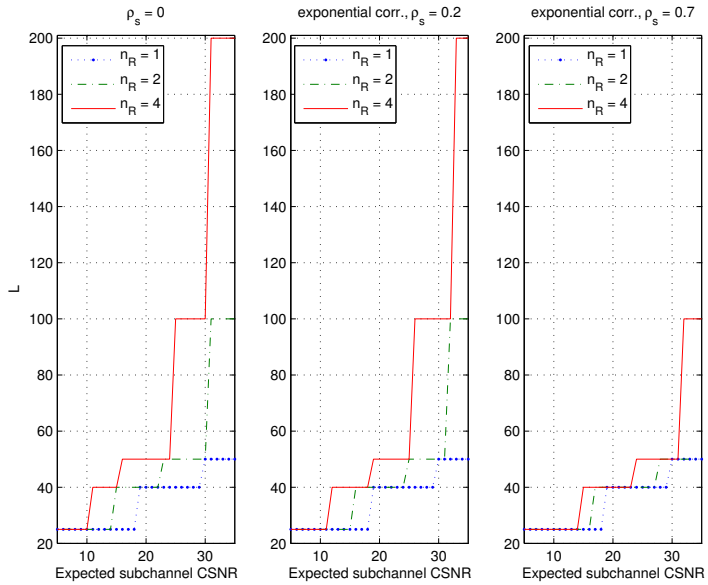


FIGURE 5.6: Optimal pilot spacing as a function of subchannel CSNR with exponential correlation for $\rho_s = 0$ (left panel), $\rho_s = 0.2$ (middle panel), and $\rho_s = 0.7$ (right panel), respectively.

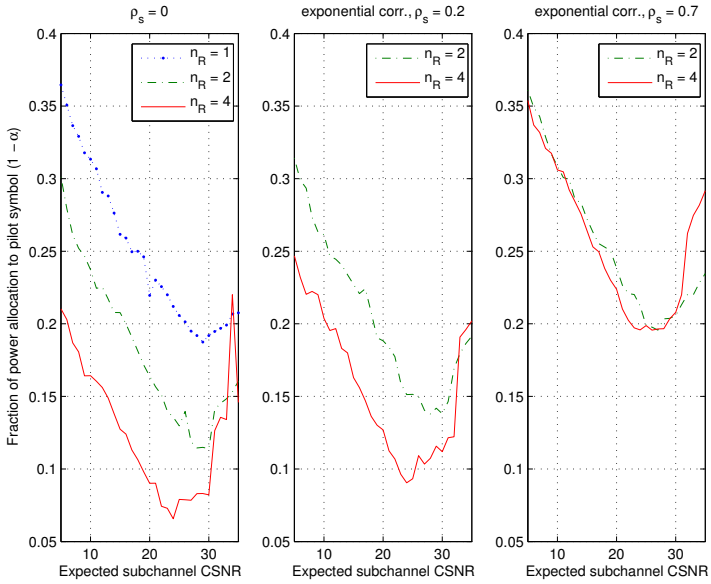


FIGURE 5.7: Optimal fraction of power allocated to pilot symbol as a function of subchannel CSNR with exponential correlation for $\rho_s = 0$ (left panel), $\rho_s = 0.2$ (middle panel), and $\rho_s = 0.7$ (right panel), respectively.

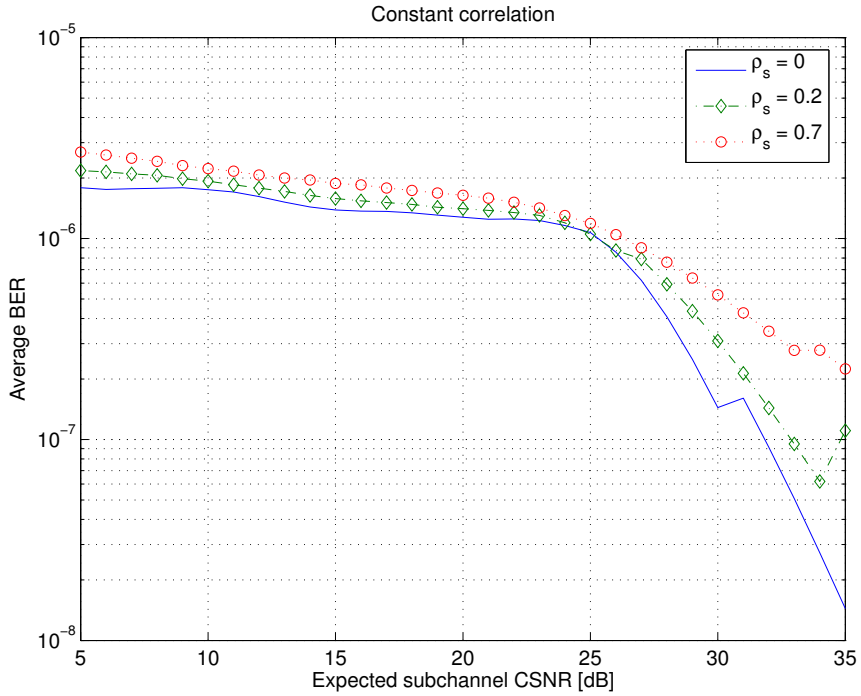


FIGURE 5.8: Average BER plotted as a function of subchannel CSNR for constant correlation model, optimal pilot spacing, and optimal power allocation. It is plotted for different spatial correlation values: $\rho_s = 0, 0, 2, 0.7$ and for 4 receive antennas.

the spatial correlation into account.

In Figure 5.8, the theoretical average BER performance with constant correlation model and different values of correlation ρ_s is depicted. It is plotted for $n_R = 4$ when pilot spacing and power allocation are optimally selected. The conclusion here is that the average BER is always below $BER_0 = 10^{-5}$ as required. We do not include the curves for exponential correlation here since they are similar and they will always satisfy the target BER_0. Note however that the slopes of the curves for $\rho_s \neq 0$ at high average CSNR is not correct. That behavior is due to the use of approximated PDFs for true and predicted CSNR, and further explanation will be given in Section 5.4.1.

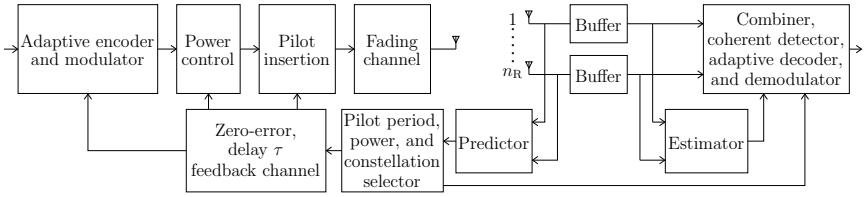


FIGURE 5.9: ACM system based on adaptive PSAM combined with multi reception of received signals where the subchannels are jointly predicted and estimated.

5.4 Link Adaptation in Spatially Correlated Antenna Diversity: Jointly Predicted Branches

As discussed above, the analysis and results in the previous section are clearly suboptimal since the spatial correlation is not taken into account in the prediction process. In the present section, the subchannels will be jointly predicted, i.e. the spatial correlation is assumed known so that it can be taken into account. Thus a "space-time predictor" is needed and must be derived. This predictor also uses the PSAM scheme to track the variations of the SIMO fading channel.

The system under study is shown in Figure 5.9, where joint prediction of the antenna branches is considered. As in the previous part, perfect channel estimation is assumed.

We buffer the pilot symbols received at all the antennas and organize them in a matrix as

$$\begin{aligned}\boldsymbol{\mathcal{Y}} &= \begin{bmatrix} y_1(k-D;0) & \cdots & y_{n_R}(k-D;0) \\ \vdots & \ddots & \vdots \\ y_1(k-D-K_p+1;0) & \cdots & y_{n_R}(k-D-K_p+1;0) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{y}_{\mathrm{pl};1} & \mathbf{y}_{\mathrm{pl};2} & \cdots & \mathbf{y}_{\mathrm{pl};n_R} \end{bmatrix}. \end{aligned} \tag{5.17}$$

The corresponding channel matrix $\boldsymbol{\mathcal{H}}$ and noise matrix $\boldsymbol{\mathcal{N}}$ are organized in a similar manner. Assuming the same pilot symbol is transmitted, the ML estimate of the received pilot vector is $\underline{\tilde{\mathbf{y}}} = \mathrm{vec}(\boldsymbol{\mathcal{Y}}) / (s\sqrt{\mathcal{E}_{\mathrm{pl}}})$, where $\mathrm{vec}(\mathbf{A})$ is the vectorized operator that converts the matrix $\mathbf{A}$ to a column vector by stacking the columns of $\mathbf{A}$ [Moon and Stirling, 2000]. The corresponding stacked vectors of channel gains and noise are denoted by $\underline{\mathbf{h}}_{\mathrm{pl}}$ and $\underline{\mathbf{n}}_{\mathrm{pl}}$, respectively.

Let the vector of the predicted channels be $\mathbf{h}_p = \mathbf{G}^{\mathsf{H}}\underline{\tilde{\mathbf{y}}}$. Then the optimal predictor $\mathbf{G}$ in the MAP sense—taking spatial correlation into account—can be found as (see Appendix A)

$$\mathbf{G} = \mathbf{R}_{\underline{\mathbf{y}}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} = \left(\mathbf{R}_{\underline{\mathbf{h}}} + \frac{1}{(1-\alpha)\bar{\gamma}_b L}\mathbf{I}_{\underline{\mathbf{n}}}\right)^{-1}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} \tag{5.18}$$

where $\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} = \mathsf{E}\left[\underline{\tilde{\mathbf{y}}}\mathbf{h}^{\mathsf{H}}\right]$ and $\mathbf{R}_{\underline{\mathbf{y}}} = \mathsf{E}\left[\underline{\tilde{\mathbf{y}}}\underline{\tilde{\mathbf{y}}}^{\mathsf{H}}\right] = \mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\underline{\mathbf{h}}_{\mathrm{pl}}^{\mathsf{H}}\right] + \mathsf{E}\left[\underline{\mathbf{n}}_{\mathrm{pl}}\underline{\mathbf{n}}_{\mathrm{pl}}^{\mathsf{H}}\right] / (|s|^2\mathcal{E}_{\mathrm{pl}})$. Note that $\underline{\mathbf{h}}_{\mathrm{pl}} = [\mathbf{h}_{\mathrm{pl};1}^{\mathsf{T}}, \mathbf{h}_{\mathrm{pl};2}^{\mathsf{T}}, \cdots, \mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}^{\mathsf{T}}]^{\mathsf{T}}$, where each element $\mathbf{h}_{\mathrm{pl};\mu} = [h_\mu(k-D;0), \cdots, h_\mu(k-D-K_p+1;0)]^{\mathsf{T}}$ is the vector of channel gains corresponding to the pilot instants received at the μth antenna and is different from $\mathbf{h} = [h_1(k;l), \cdots, h_{n_{\mathrm{R}}}(k;l)]^{\mathsf{T}}$ which is the vector of n_{R} channel gains at the time $(k;l)$. In that way, when assuming the noise is white both in space and time, the correlation between any branches of any lags is completely described in $\mathbf{R}_{\underline{\mathbf{h}}}$. Using the assumption in (5.2) we can write

$$\mathbf{R}_{\underline{\mathbf{h}}} = \mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\underline{\mathbf{h}}_{\mathrm{pl}}^{\mathsf{H}}\right] = \begin{bmatrix} \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}\mathbf{h}_{\mathrm{pl};1}^{\mathsf{H}}\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}^{\mathsf{H}}\right] \\ \vdots & \ddots & \vdots \\ \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}\mathbf{h}_{\mathrm{pl};1}^{\mathsf{H}}\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}^{\mathsf{H}}\right] \end{bmatrix} = \mathbf{R}^{\mathbf{s}} \otimes \mathbf{R}^{\mathbf{t}} \tag{5.19}$$

and

$$\begin{aligned} \mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} &= \mathsf{E}\left[\underline{\tilde{\mathbf{y}}}\mathbf{h}^{\mathsf{H}}\right] = \mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\mathbf{h}^{\mathsf{H}}\right] \\ &= \begin{bmatrix} \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}h_1^*(k;l)\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}h_{n_{\mathrm{R}}}^*(k;l)\right] \\ \vdots & \ddots & \vdots \\ \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}h_1^*(k;l)\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}h_{n_{\mathrm{R}}}^*(k;l)\right] \end{bmatrix} = \mathbf{R}^{\mathbf{s}} \otimes \mathbf{r}. \end{aligned} \tag{5.20}$$

Here, we use the following notations $\mathbf{R}^{\mathbf{s}} = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};\mu}\mathbf{h}_{\mathrm{pl};\nu}^{\mathsf{H}}\right]$, $\mathbf{R}^{\mathbf{t}} = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};\mu}\mathbf{h}_{\mathrm{pl};\mu}^{\mathsf{H}}\right]$, and $\mathbf{r} = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};\mu}h_\mu^*(k;l)\right]$ to denote the spatial correlation matrix, temporal correlation matrix, and covariance vector, respectively. Moreover, $\otimes$ is the Kronecker product [Moon and Stirling, 2000].

When using the optimal predictor given by (5.18) the total prediction error variance $\sigma_p^2 = \mathsf{E}\left[\|\mathbf{h} - \mathbf{h}_p\|^2\right]$ can be shown to be (see Appendix A)

$$\sigma_p^2 = n_{\mathrm{R}} - \mathrm{tr}\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{\mathsf{H}}\mathbf{G}\right\}. \tag{5.21}$$

In this part, we will only consider linear array antenna elements where the correlation between the elements decreases exponentially. Thus, the

power correlation between the antenna branches μ and ν is described by (5.6). As earlier, the spatial correlation of the fading gain between any two adjacent branches is $\rho_{h,s}(1) = c - \jmath d$. For simplicity, we assume that the real and imaginary parts are equally correlated. Hence, from (5.4), we see that $c = d = \sqrt{\rho_s/2}$.

Using MRC, the instantaneous true and predicted CSNR are written as

$$\gamma_d = \frac{\mathcal{E}_d \|\mathbf{h}(k;l)\|_{\mathsf{F}}^2}{N_0} \quad \text{and} \quad \hat{\gamma} = \frac{\bar{\mathcal{E}}_d \|\mathbf{h}_p(k;l)\|_{\mathsf{F}}^2}{N_0}, \tag{5.22}$$

respectively. Thus the average CSNRs can be calculated as

$$\bar{\gamma}_d = \frac{\mathcal{E}_d \mathsf{E}\left[\|\mathbf{h}\|_{\mathsf{F}}^2\right]}{N_0} = r_d \bar{\gamma}_b n_{\mathrm{R}} \quad \text{and} \quad \bar{\hat{\gamma}} = \frac{\bar{\mathcal{E}}_d \mathsf{E}\left[\|\mathbf{h}_p\|_{\mathsf{F}}^2\right]}{N_0} = r \bar{\gamma}_b n_{\mathrm{R}}, \tag{5.23}$$

respectively, where $r = \bar{\mathcal{E}}_d \mathsf{tr}\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{\mathsf{H}} \mathbf{G}\right\} / (n_{\mathrm{R}} \mathcal{E})$ (different from r for the uncorrelated case used in (4.12)). In this way, we can use the same notation as in the previous part and both the BER and ASE expressions can be reused here.

Again, the joint PDF of γ_d and $\hat{\gamma}$ is $(\gamma_d, \hat{\gamma}) \sim \mathcal{G}(m_d, \theta_1, \theta_2, \rho)$ and ρ can be calculated by means of (5.9). When the spatial correlation is incorporated, we get (look to the Appendix D for detailed derivations)

$$\rho = \frac{m_d \left(\Psi - n_{\mathrm{R}} \mathsf{tr}\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{\mathsf{H}} \mathbf{G}\right\}\right)}{n_{\mathrm{R}} \mathsf{tr}\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{\mathsf{H}} \mathbf{G}\right\}} \tag{5.24}$$

where

$$\Psi = \sum_{\mu=1}^{n_{\mathrm{R}}} \sum_{\nu=1}^{n_{\mathrm{R}}} \left\{ \underline{\mathbf{g}}_\mu^{\mathsf{H}} \left(\mathbf{R}_{\underline{\mathbf{h}}} + \Re\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\} \Re\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\}^{\mathsf{T}} + \Im\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\} \Im\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\}^{\mathsf{T}} \right.\right.$$
$$\left.\left. - \jmath \Re\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\} \Im\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\}^{\mathsf{T}} + \jmath \Im\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\} \Re\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\right\}^{\mathsf{T}} \right) \underline{\mathbf{g}}_\mu + \frac{\|\underline{\mathbf{g}}_\mu\|^2}{(1-\alpha)\bar{\gamma}_b L} \right\}.$$

In the above expression, $\underline{\mathbf{g}}_\mu$ is the μth column of the predictor $\mathbf{G}$ and $\mathbf{R}^{(\nu)}$ is the νth column of the matrix $\mathbf{R}$. Also, $\Re\{z\}$ and $\Im\{z\}$ denotes the real and imaginary parts of z, respectively. It can be shown that when the same predictor is used on all branches and the branches are predicted independent of each other—i.e., the effect of spatial correlation is not considered in the prediction process—(5.24) reduces to (5.10). It is marked that ρ in previous

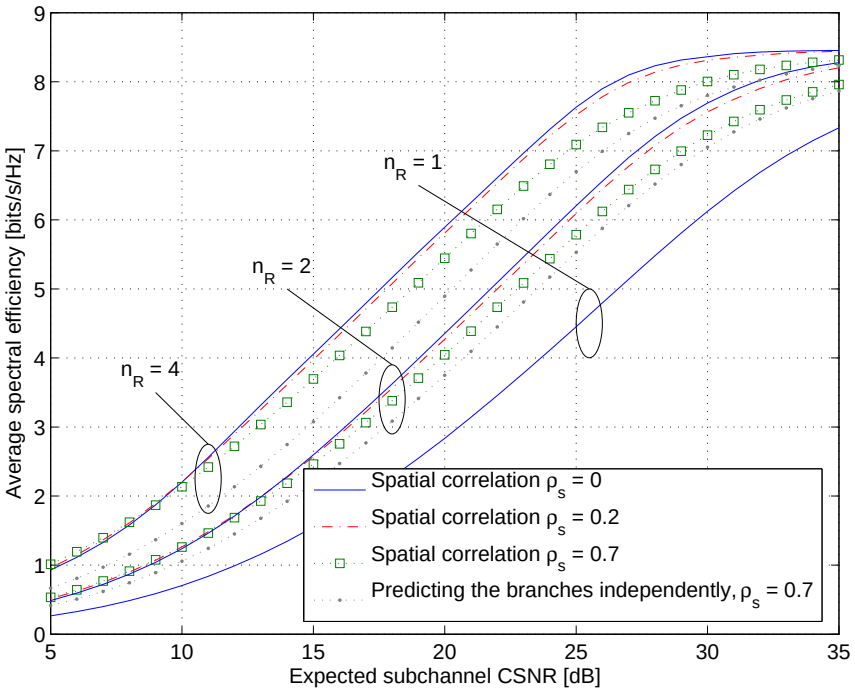


FIGURE 5.10: Average spectral efficiency as a function of expected CSNR on one branch for different combinations of number of antennas and spatial correlations. It is also plotted for when the subchannels are predicted independently for $\rho_s = 0.7$.

section and in [Holter and Øien, 2006] is independent of the spatial correlation ρ_s since the channel is predicted independently on each subchannel. In contrast to that, ρ in this section also contains ρ_s via the term m_d and the matrices $\mathbf{R}_{\underline{\mathbf{h}}}$ and $\mathbf{R}_{\underline{\mathbf{yh}}}$.

5.4.1 System Performance

To this end, since we have consequently introduced the same notations here as earlier sections, the expressions related to BER and ASE in Section 5.3 are still applicable here; together with the just-derived parameters. Hence, we are not going to reproduce them at this stage, but we will go further with the numerical results. The prediction of the channel vector in this part is based on 250 pilot symbols from each of the subchannels.

The results for ASE when both pilot spacing and power allocation are optimal are plotted in Figure 5.10. It is clear that the loss in ASE due to spatial correlation is reduced when all subchannels are *jointly* predicted. The

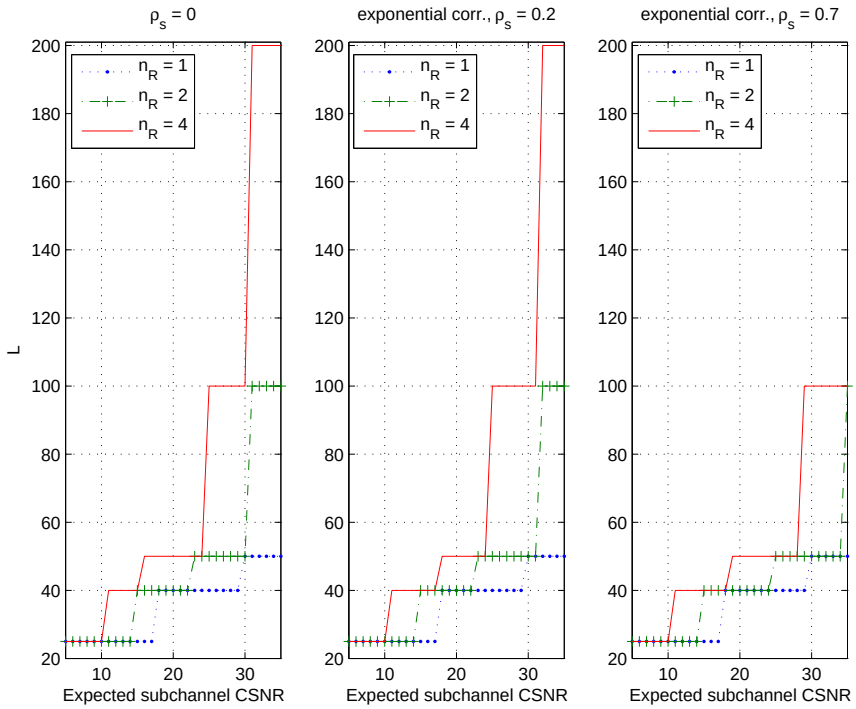


FIGURE 5.11: Optimal pilot spacing L as function of average subchannel CSNR. The subfigures are plotted for different spatial correlation ρ_s.

gain is larger when there are many antennas available to combine. However, we can see that the curves corresponding to joint prediction of the correlated branches are higher compared to the uncorrelated case at low average CSNR. This effect is an artifact stemming from the use of approximated PDFs for both true and predicted CSNRs. The same behavior was observed in [Holter and Øien, 2006] under perfect conditions, i.e. perfect channel estimation and zero feedback delay. It is shown in Figure 5.15 that the probability of choosing the dominant constellation is smaller and the probability of choosing the other constellations is greater when using the approximated PDFs in spatially correlated branches. At the lower end and the upper end of the considered average CSNR, they respectively give rise to higher and lower ASE compared to when exact PDFs are used in the uncorrelated branches systems.

Intuitively, it is expected that the diversity gain disappears when the branches become more correlated. Thus the effective (combined) channel will vary more like the one-branch case and, therefore, we need the same amount of pilot overhead to predict the channel for a given accuracy. This

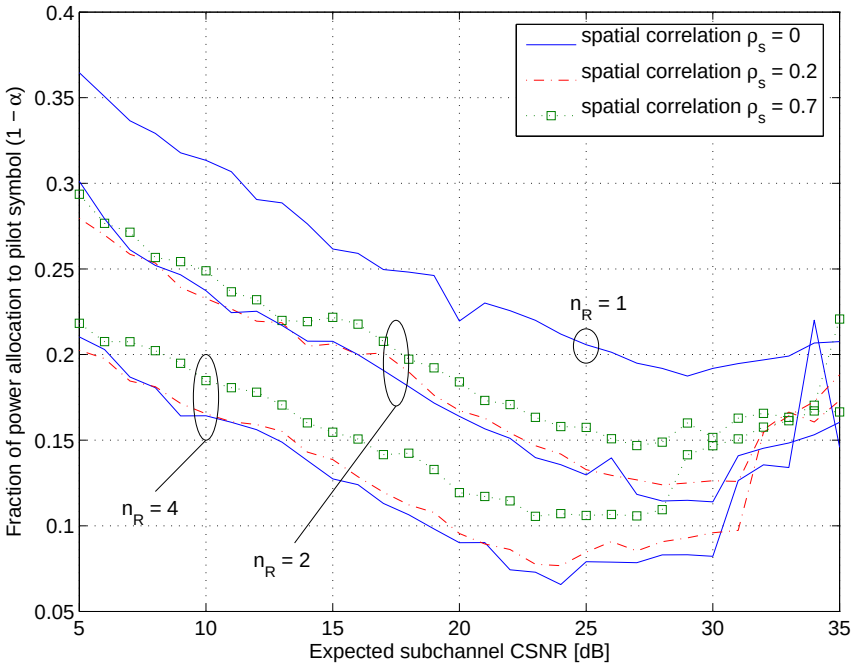


FIGURE 5.12: The amount of power allocated to pilot symbols is plotted against average subchannel CSNR for different spatial correlation values and different number of receive antennas.

effect is confirmed in Figure 5.11 where it is clear that, with increasing spatial correlation (from left to right panel), the curves for L when $n_R > 1$ approach the one when $n_R = 1$. However, due to the array gain the combined channel will approach a no-diversity scenario with higher average CSNR. Because of this array gain, the power allocated to pilot symbols does not vary much from the uncorrelated case, which is illustrated in Figure 5.12. The variations of the curves at very high average CSNRs are due to the numerical instability of the optimization process because the ASE curves saturate.

Looking at Figures 5.11 and 5.12 together we see that it might be sufficient to decrease the pilot spacing only. The statement is confirmed when we plot and compare the ASE curves in Figure 5.10 when $\rho_s = 0.7$ with the ASE performance for the same spatial correlation; still using the optimal values of L but with the power scheme for the *uncorrelated* case. As shown in Figure 5.13, the difference in performance is hardly noticeable.

Comparing these results here with the results in the previous section where both the pilot spacing is reduced and the pilot power is unneces-

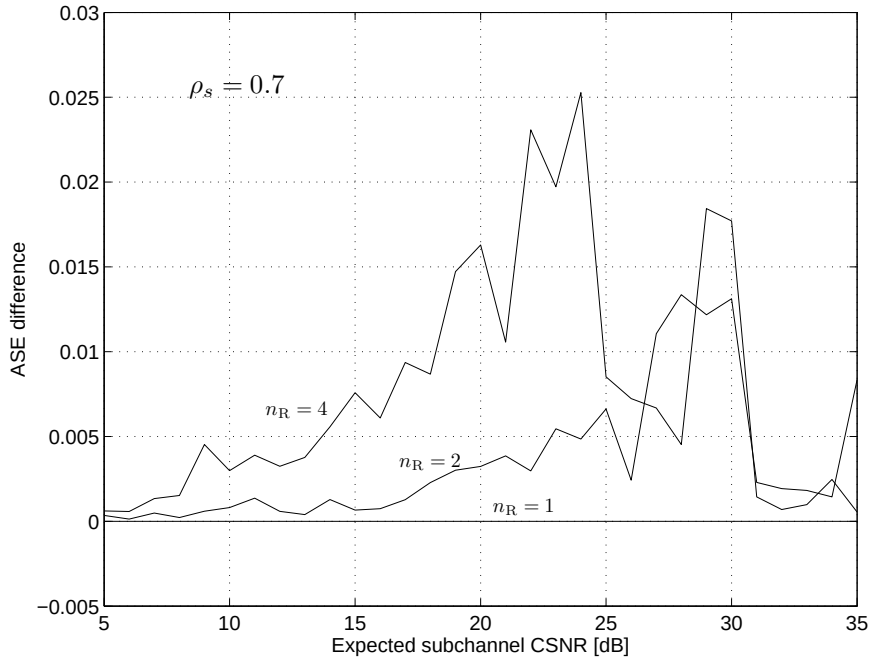


FIGURE 5.13: ASE difference between the optimal L, optimal α case and optimal L with α as in the uncorrelated case. It is plotted for $\rho_s = 0.7$.

sary increased simultaneously, it can be inferred that, in presence of spatial correlation, independent prediction of the different branches should be avoided as it has a strong negative influence on ASE.

The average BER performance is plotted in Figure 5.14 where it can be seen that the requirement of $\text{BER}_0 = 10^{-5}$ is again always satisfied. In presence of spatial correlation and using the approximate PDFs, the slope indicating the diversity order at high average CSNR is not correct. This effect was also observed in Figure 5.8. As documented in [Wang and Giannakis, 2003], the error rate performance at high average CSNR is strongly dependent on the behavior of the *lower tail* of a PDF, and this is exactly where the approximate PDF deviates from the exact one. However, this does not put any limitations on our work since we are more interested in the low and medium average CSNR regions. For high average CSNR values, the ACM system behaves like a fixed-rate system anyway, since only the largest constellation will be employed. In that case, the effects of spatially correlated channels are well documented in the literature [Simon and Alouini, 2005, Chap. 9.7].

In Figure 5.15, code selection probability as a function of average CSNR

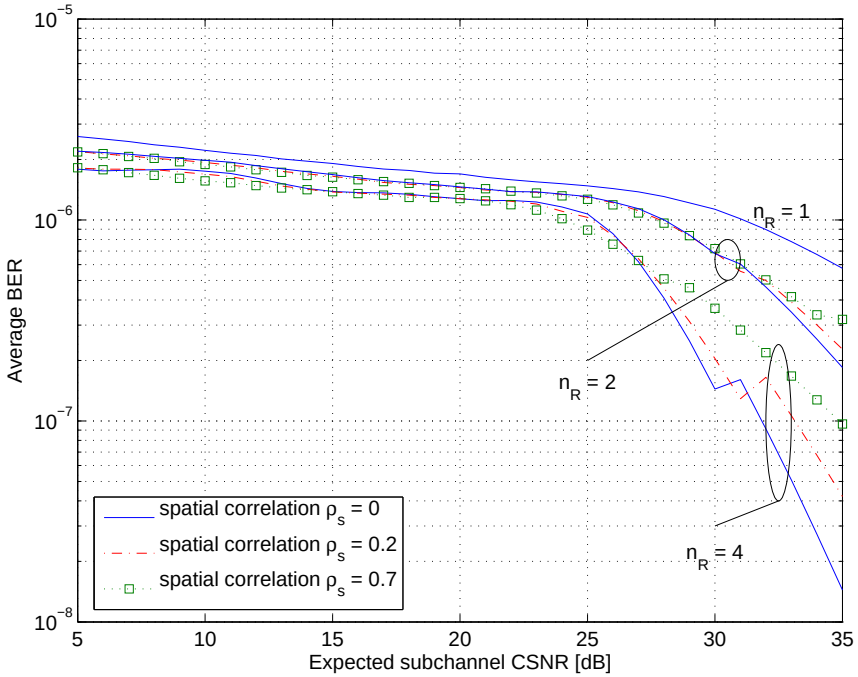


FIGURE 5.14: Average BER as a function of expected CSNR on one branch for different combinations of number of antennas and spatial correlations.

per branch is depicted. Note that the curve for $M = 0$ corresponds to the outage probability (probability of no transmission). When the average branch CSNR increases, the next higher order modulation becomes the dominant modulation scheme. In presence of spatial correlation (right panel), the probability curves are wider and lower than for the uncorrelated branches (left panel); which implies that the probability of selecting the optimal code (the dominant modulation) has been lowered, while the probability of selecting all the other codes has been increased. In addition, at the lower end of the average CSNR range, only constellations $M = 4$ and $M = 8$ are selected with considerably high probability when branches are independent. In addition to that, the constellation $M = 16$ can also be selected with high probability in presence of spatial correlation. That is why the ASE curves for correlated scenario are somewhat higher than the curves corresponding to the uncorrelated scenario.

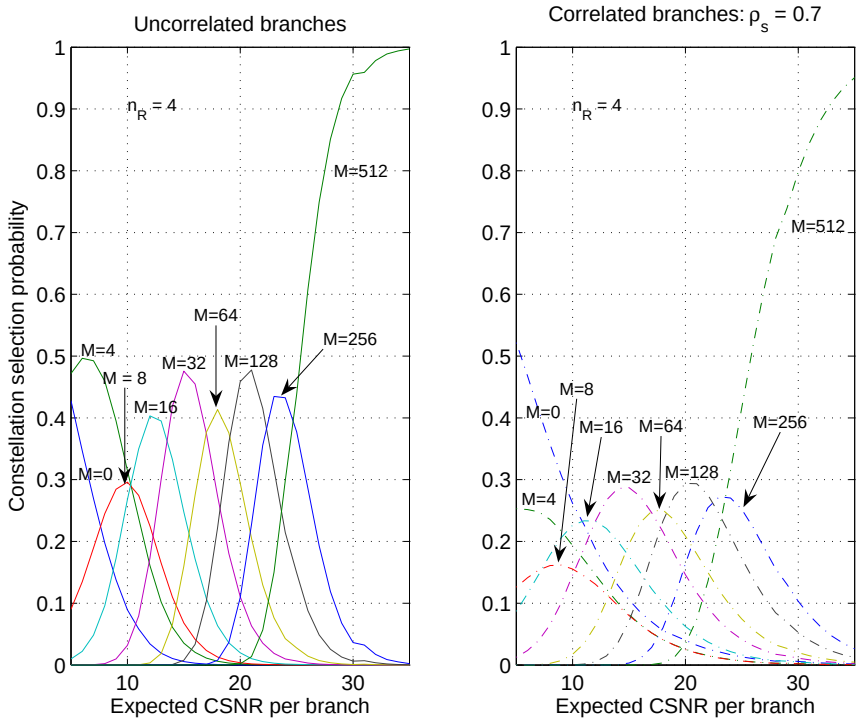


FIGURE 5.15: Code selection probability as a function of average CSNR per branch. $M = 0$ corresponds to outage. It is plotted for $n_R = 4$ and for uncorrelated branches (left panel) and correlated branches with $\rho_s = 0.7$ (right panel).

5.5 Conclusions

We have investigated an ACM system operating on a SIMO diversity Rayleigh fading channel in the presence of spatial correlation between the receive antenna branches. The analysis is done based on separability of the space-time correlation function, i.e. representing it as a product of spatial and temporal correlation functions. Furthermore, the combined instantaneous true and predicted CSNRs are approximated as gamma distributed RVs with the two first moments equal to the exact PDF.

First, we consider a suboptimal prediction procedure where the fact that the branches are correlated is not taken into account. In other words, it is the case where we use the system originally designed for spatially uncorrelated branches in a correlated one. The results for constant correlation and exponential correlation were presented and they both degrade systems performance. Moreover, that degradation was larger with constant corre-

lation compared to exponential correlation. Furthermore, the results also indicated the suboptimality of the scheme.

We then develop a MAP-optimal "space-time predictor" to incorporate the spatial correlation. In general, the throughput in terms of ASE is reduced due to the reduced diversity gain when spatial correlation increases. After applying the new space-time predictor it is observed that ASE is still lower than in the uncorrelated case, but the negative impact of spatial correlation on the ASE is substantially reduced. When the subchannels become more correlated, the channel approaches a SISO system with higher average CSNR. As a result, the optimization gives shorter pilot period where the power can be distributed as if the branches were uncorrelated.

Due to the usage of approximate PDFs, the slope of BER performance at high average CSNR is not correctly captured, but this does not put any limitations on our work. In that region, the largest constellation will be used with very high probability most of the time and the ACM system is then said to be reduced to a single-rate system. For this case, the impact of spatial correlation is well documented in the literature.

Chapter 6

Orthogonal Space-Time Block Coded Rate-Adaptive Systems With Imperfect CSI

This chapter is based on a paper presented at the conference *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* [Duong, Holter, and Øien, 2005] and a journal letter which is accepted for publication in *IEEE Transactions on Wireless Communications* [Duong and Øien, 2006].

6.1 Introduction

Extending the work in Chapters 3 and 4, we now consider a full MIMO system. In general, MIMO systems may be divided into two categories: rate maximization schemes and diversity maximization schemes, also denoted *spatial multiplexing* (SM) systems and *MIMO diversity* systems, respectively. SM offers a linear increase in the transmission rate (or capacity) at no extra bandwidth or power expenditure. This is obtained by transmitting independent data streams from each transmit antenna, or demultiplexing a single data stream in to n_{T} substreams which subsequently are transmitted from separate transmit antennas. In a rich scattering environment, the fading gains become uncorrelated. In this case, when knowledge of the channel is available at the receiver, the composite receive signal can be separated by solving linearly independent equations. Thus, the receiver can detect the different data streams, or combine the substreams into the original stream [Paulraj *et al.*, 2003].

A MIMO diversity system uses the multiple antennas to maximize the diversity order of the system. This can be achieved by using a technique called STBC at the transmitter [Alamouti, 1998; Tarokh *et al.*, 1999]. Given a number of receive antennas, STBC is used to exploit the available spatial diversity of the channel which is equal to the product of the number of transmit and receive antennas if the channels between different transmit-receive antenna pairs fade independently. In particular, orthogonal space-time block codes are able to provide full diversity with a simple decoding algorithm and without requiring any feedback from the receiver to the transmitter. Over i.i.d. Rayleigh fading channels with multiple receive antennas, however, these orthogonal space-time block codes incur a loss in capacity because they convert the channel matrix into a scalar AWGN channel whose capacity is smaller than the true channel capacity [Sandhu and Paulraj, 2000].

As in the previous chapters, we will optimize the pilot spacing and power in such a MIMO diversity system. The optimization is done assuming that imperfect channel knowledge is available at both transmitter and receiver. Note that there is an inherent trade-off between the use of diversity and the need for adaptivity. The higher the diversity order, the more stable the channel is and the less need there is for link adaptation. We have therefore restricted ourselves to moderate diversity orders to demonstrate the power of link adaptation in this context.

Adapting optimal transmitter eigen-beamforming coupled with STBC is analyzed by Zhou and Giannakis [2002], where imperfect CSI is available at the transmitter. This technique is utilized in [Zhou and Giannakis, 2004a] and [Zhou and Giannakis, 2004b]. Note however that, in these systems, the *whole* channel gain matrix must be fed back to the sender. To facilitate a low feedback rate, in our system we only send the overall predicted CSNR back to the transmitter. Our feedback is thus clearly less complex than feeding back the whole channel matrix.

The chapter is organized as follows: the system under consideration is described in Section 6.2, where the employed orthogonal designs are introduced together with the frame structure after STBC. The channel estimation and prediction procedures are re-captured in Section 6.3. An analysis of BER performance and optimization of ASE is presented in Sections 6.4 and 6.5, respectively. Numerical examples are given in Section 6.6 and conclusions are drawn in Section 6.7.

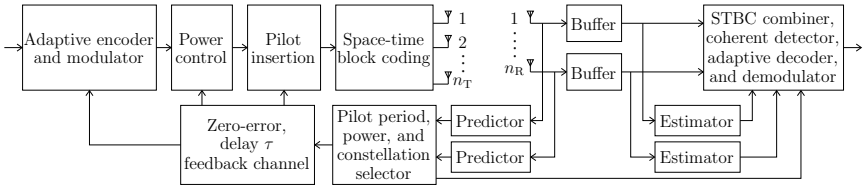


FIGURE 6.1: ACM system with adaptive PSAM-based channel prediction and estimation. The predicted channels are used for system adaptation and the estimated channels are used for coherent detection. The system is operating on a MIMO diversity channel.

6.2 System Model

The system under consideration is shown in Figure 6.1. In addition to n_R receive antennas as in the previous two chapters, the system now also has n_T transmit antennas. All subchannels between any transmitter-receiver pair are assumed to be mutually independent and Rayleigh distributed, with Jakes correlation profile used in the numerical examples.

The orthogonal space-time block codes used in this paper are listed in Tab. 6.1. These orthogonal designs and some other orthogonal designs (for both real and complex signals) can be found in [Tarokh *et al.*, 1999]. The space-time encoder maps S symbols into n_T orthogonal sequences of length T (given as $T = (S/R_s)T_s$) where T_s is still the channel symbol interval and R_s is the rate of the employed space-time block code. Note that the pilot symbols are not space-time block coded. Thus, the pilot is transmitted once from each antenna (time-multiplexing). While a pilot is transmitted from one antenna, the other antennas are silent such that each receiver branch can track the channel between itself and the transmitting antenna. The same pilot scheme is utilized in [Holter *et al.*, 2003]. Note that making the pilot symbol orthogonal by spreading it with each antenna's signature code of length n_T does not improve the system performance since the channel predictor is found to be independent of this factor, and the mean and variance of the noise remains unchanged [Zhou and Giannakis, 2004b]. Also, the channel is still the same after de-spreading at the receiver.

Another possibility is to transmit sequences of orthogonal pilot symbols from different antennas (not spreading one symbol). In this case, we will need another predictor/estimator and which, we believe, is more complex than the Wiener filter in our system. This is due to the correlation properties which will be more involved since we also have to consider the intersymbol correlation within one sequence in addition to the correlation between different sequences. Moreover, the system is using the same amount of time

TABLE 6.1: Orthogonal designs for STBC used in this chapter. Orthogonal design $\mathcal{O}_1$ and $\mathcal{O}_2$ corresponds to the regular data stream with no STBC and the Alamouti scheme, respectively. $\mathcal{O}_4$ is the orthogonal design given by matrix (40) in [Tarokh *et al.*, 1999].

Orthogonal design	n_T	R_s	S	T
$\mathcal{O}_1$	1	1	1	1
$\mathcal{O}_2$	2	1	2	2
$\mathcal{O}_4$	4	3/4	3	4

slots also in this case [Alamouti, 1998].

Let $\mathbf{z}_d \in \mathbb{C}^{n_\mathrm{R}\times 1}$ be the received, noisy, and faded data symbol vector in complex baseband. It can be written as

$$\mathbf{z}_d(k;l) = \sqrt{\frac{\mathcal{E}_d}{n_\mathrm{T}}}\mathbf{H}(k;l)\mathbf{s}(k;l) + \mathbf{n}(k;l), \quad l \in [n_\mathrm{T}, \cdots, L_b - 1], \tag{6.1}$$

where

$$L_b = \frac{mS}{R_s} + n_\mathrm{T} = \frac{L-1}{R_s} + n_\mathrm{T} \tag{6.2}$$

is the pilot symbol spacing *on a single antenna branch* after STBC [Holter *et al.*, 2003] and m is a non-zero positive integer. This is illustrated in Figure 6.2 for $m = 1$. Furthermore, let $\mathbf{z}_\mathrm{pl} \in \mathbb{C}^{n_\mathrm{R}\times 1}$ be the received pilot symbol vector:

$$\mathbf{z}_\mathrm{pl}(k;l) = \sqrt{\frac{\mathcal{E}_\mathrm{pl}}{n_\mathrm{T}}}\mathbf{H}(k;l)\mathbf{s}(k;l) + \mathbf{n}(k;l), \quad l \in [0, \cdots, n_\mathrm{T} - 1]. \tag{6.3}$$

Antenna no. 1	P	$\circ$	$\circ$	$\circ$	D_1	$-D_2^*$	$\frac{D_3^*}{\sqrt{2}}$	$\frac{D_3^*}{\sqrt{2}}$
Antenna no. 2	$\circ$	P	$\circ$	$\circ$	D_2	D_1^*	$\frac{D_3^*}{\sqrt{2}}$	$-\frac{D_3^*}{\sqrt{2}}$
Antenna no. 3	$\circ$	$\circ$	P	$\circ$	$\frac{D_3}{\sqrt{2}}$	$\frac{D_3}{\sqrt{2}}$	$\cdots$	$\cdots$
Antenna no. 4	$\circ$	$\circ$	$\circ$	P	$\frac{D_3}{\sqrt{2}}$	$-\frac{D_3}{\sqrt{2}}$	$\cdots$	$\cdots$

FIGURE 6.2: Example of a frame structure after STBC where $n_\mathrm{T} = 4$, and the orthogonal design $\mathcal{O}_4$ in Tab. 6.1 is used. Here, P stands for pilot and $\circ$ denotes that the system does not send anything, while Ds are data symbols. To reduce the size of the figure we avoid to write out the four last data symbols, and demonstrate only the smallest frame size.

In both equations above, the notation $x(k;l)$ is the compact way of writing $x(kL_bT_s + lT_s)$, $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ is the channel gain matrix, $\mathbf{s} \in \mathbb{C}^{n_T \times 1}$ is the vector of transmit symbols, and $\mathbf{n} \in \mathbb{C}^{n_R \times 1}$ is the channel AWGN. Without knowing the whole channel gain matrix (as in the beamforming case) we consider when power is equally allocated to different transmit antennas. As such, $\mathcal{E}_{pl}/n_T$ and $\mathcal{E}_d/n_T$ is the power per pilot and per data symbol transmitted from one antenna, respectively. Moreover, we assume that $\mathsf{E}\left[|s_\mu(k;l)|^2\right] = 1$; $\mu \in [1, \cdots, n_T]$; $l \in [n_T, \cdots, L_b - 1]$ (data symbols), and that the pilots have unit magnitude (i.e., $|s(k;l)| = 1$ for $l \in [0, \cdots, n_T - 1]$). Furthermore, the noise is assumed white in both space and time with zero mean and variance equal to N_0. The elements of the channel gain matrix are assumed to come from a stationary complex Gaussian RP with zero mean and unit variance.

Similarly to [Cai and Giannakis, 2005], $\bar{\mathcal{E}}_d$ can be calculated as

$$\bar{\mathcal{E}}_d = \frac{\alpha\mathcal{E}[(L_b - n_T)R_s + 1]}{(L_b - n_T)R_s} = \frac{\alpha\mathcal{E}L}{L-1} \tag{6.4}$$

which is unchanged from all the previous chapters. Thus, the pilot power $\mathcal{E}_{pl}$ must be equal to $\mathcal{E}_{pl} = (1-\alpha)\mathcal{E}L$. Again, the actual data power $\mathcal{E}_d$ is given by (2.13). Later on, we will show that the total predicted CSNR is gamma distributed, i.e., $\hat{\gamma} \sim \mathcal{G}\left(n_T n_R, r\bar{\gamma}_b/n_T\right)$. As a result, $\mathcal{E}_d$ is

$$\mathcal{E}_d = \frac{\bar{\mathcal{E}}_d}{\overline{\Gamma}\left(n_T n_R, \dfrac{\hat{\gamma}_1 n_R}{r\bar{\gamma}_b}\right)}, \tag{6.5}$$

where $\bar{\gamma}_b = \mathcal{E}/N_0$ is the expected CSNR on *one* receive branch,[1] and, similarly to the uncorrelated SIMO case, $r = \bar{\mathcal{E}}_d(1-\sigma_p^2)/\mathcal{E}$. To obtain this, we have again used the assumption that the subchannels are independent and the fact that the prediction error variance is the same on all branches. Hence, $\sigma^2_{p;\mu\nu} = \sigma_p^2 \; \forall \; \mu, \nu$ where μ and ν is the antenna index at the transmitter and receiver, respectively.

6.3 Channel Estimation and Prediction

The transmit frame structure is illustrated in Figure 6.2. Clearly, it is somewhat different from the structure in Figure 2.3. The expressions regarding channel estimation and prediction are similar to those obtained earlier, but

[1] Which is the sum of all the subchannels' average CSNR received in one antenna. That is, $\bar{\gamma}_b = \sum_{a=1}^{n_T} \bar{\gamma}_{ab} = \sum_{a=1}^{n_T} \mathcal{E}/(n_T N_0) = \mathcal{E}/N_0$.

we will reproduce them here since some care of the time indices must be taken due to the new frame structure.

Both the estimator and predictor are linear and made optimal in the MAP sense. Moreover, the branches are uncorrelated by assumption and we thus can perform estimation and prediction independently on each of the receive branch. The temporal correlation is also assumed known.

6.3.1 Estimation

Based on the non-causal vector of K_e received pilot symbols $\mathbf{y}_{\mathrm{pl};\mu\nu}(k;\mu-1)$, corresponding to the pilot instances of the μ-ν antenna pair and the channel gains $\mathbf{h}_{\mathrm{pl};\mu\nu} = [h_{\mu\nu}(k-\lfloor K_e/2\rfloor;\mu-1),\cdots,h_{\mu\nu}(k+\lfloor (K_e-1)/2\rfloor;\mu-1)]^{\mathsf{T}}$, a *non-causal* MAP estimator estimates the fading channel gain in the set $\{h_{\mu\nu}(k;l)\}_{l=n_{\mathrm{T}}}^{L_b-1}$ as $h_{e;\mu\nu}(k;l) = \mathbf{w}_e^{\mathsf{H}}\mathbf{y}_{\mathrm{pl};\mu\nu}(k;\mu-1)$, $\mu\in[1,\cdots,n_{\mathrm{T}}]$, and $\nu\in[1,\cdots,n_{\mathrm{R}}]$ where

$$\mathbf{w}_e = \sqrt{\frac{\mathcal{E}_{\mathrm{pl}}}{n_{\mathrm{T}}}}\left(\frac{\mathcal{E}_{\mathrm{pl}}}{n_{\mathrm{T}}}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}_e\boldsymbol{\mathcal{D}}^*(\mathbf{s})+N_0\mathbf{I}_{K_e}\right)^{-1}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_e \tag{6.6}$$

is the MAP-optimal estimator.[2] Assuming the same pilot is transmitted from all of the transmit antennas and defining the estimation error as $\epsilon_{e;\mu\nu}(k;l) = h_{\mu\nu}(k;l)-h_{p;\mu\nu}(k;l)$, the MMSE of any subchannel is given as

$$\sigma^2_{e;\mu\nu}(l) = \mathsf{E}\left[|\epsilon_{e;\mu\nu}|^2\right] = 1-\sum_{\kappa=1}^{K_e}\frac{|\mathbf{u}_\kappa^{\mathsf{H}}\mathbf{r}_e|^2(1-\alpha)L\bar{\gamma}_b}{(1-\alpha)L\bar{\gamma}_b\lambda_\kappa+n_{\mathrm{T}}} \tag{6.7}$$

which differs from the MMSEs in the previous two chapters by the term n_{T}. In (6.7), $\{\mathbf{u}_\kappa\}$ denotes the eigenvectors of the covariance matrix $\mathbf{R}_e = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};\mu\nu}\mathbf{h}_{\mathrm{pl};\mu\nu}^{\mathsf{H}}\right]$, $\{\lambda_\kappa\}$ are the corresponding eigenvalues, and $\mathbf{r}_e$ is the covariance vector; $\mathbf{r}_e = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};\mu\nu}h^*_{\mu\nu}(k;l)\right]$.

6.3.2 Prediction

We assume that the transmitter adaptation occurs only once per transmission frame of L_b symbols. Thus, the *causal* predictor uses K_p pilot symbols from the past to predict one sample in the set $\{h_{\mu\nu}(k;l)\}_{l=n_{\mathrm{T}}}^{L_b-1}$ of the kth frame, which is $\tau = DL_bT_s$ seconds ahead in time. Here, D is the distance measured in the number of frames.

[2] There are misprints in [Duong *et al.*, 2005] where both the estimator and predictor should read $\mathbf{w}_t = \sqrt{\mathcal{E}_{\mathrm{pl}}/n_{\mathrm{T}}}((\mathcal{E}_{\mathrm{pl}}/n_{\mathrm{T}})\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}_t\boldsymbol{\mathcal{D}}^*(\mathbf{s})+N_0\mathbf{I}_{K_t})^{-1}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_t$. The subscript $t=e$ or p indicates estimator or predictor, respectively.

Let the channel gain vector of one subchannel (corresponding to the pilot instants vector $\mathbf{y}_{\mathrm{pl};\mu\nu}(k;\mu-1)$) used in the prediction be $\mathbf{h}_{\mathrm{pl};\mu\nu} = [h_{\mu\nu}(k-D;\mu-1),\cdots,h_{\mu\nu}(k-D-K_p+1;\mu-1)]^{\mathsf{T}}$, the covariance matrix and the covariance vector is now $\mathbf{R}_p = \mathsf{E}\big[\mathbf{h}_{\mathrm{pl};\mu\nu}\mathbf{h}_{\mathrm{pl};\mu\nu}^{\mathsf{H}}\big]$ and $\mathbf{r}_p = \mathsf{E}\big[\mathbf{h}_{\mathrm{pl};\mu\nu}h_{\mu\nu}^{*}(k;l)\big]$, respectively. Similar to the estimation case, the predicted channel is a linear combination of the received pilot symbols; denoted by $h_{p;\mu\nu}(k;l) = \mathbf{w}_p^{\mathsf{H}}\mathbf{y}_{\mathrm{pl};\mu\nu}(k;\mu-1)$. Let $\epsilon_{p;\mu\nu}(k;l) = h_{\mu\nu}(k;l) - h_{p;\mu\nu}(k;l)$ and given the predictor $\mathbf{w}_p$—which is similar to (6.6) but with $\mathbf{R}_e$ and $\mathbf{r}_e$ respectively replaced by $\mathbf{R}_p$ and $\mathbf{r}_p$, and with a correct dimension on the identity matrix—the MMSE of the prediction error is

$$\sigma_{p;\mu\nu}^2(l) = \mathsf{E}\big[|\epsilon_{p;\mu\nu}|^2\big] = 1 - \sum_{\kappa=1}^{K_p}\frac{|\mathbf{u}_\kappa^{\mathsf{H}}\mathbf{r}_e|^2(1-\alpha)L\bar{\gamma}_b}{(1-\alpha)L\bar{\gamma}_b\lambda_\kappa + n_{\mathrm{T}}}, \tag{6.8}$$

where $\{\mathbf{u}_\kappa\}$ and $\{\lambda_\kappa\}$ are now the sets of eigenvectors and eigenvalues of $\mathbf{R}_p$, respectively.

Due to the assumption of independent branches, both estimation MMSE and prediction MMSE are the same on all branches. That is,

$$\sigma_{e;\mu\nu}^2(l) = \sigma_e^2(l), \tag{6.9a}$$

$$\sigma_{p;\mu\nu}^2(l) = \sigma_p^2(l) \tag{6.9b}$$

$\forall\ \mu \in [1,\cdots,n_{\mathrm{T}}]$ and $\nu \in [1,\cdots,n_{\mathrm{R}}]$.

6.4 Evaluation of BER Performance

Based on the estimated channel gains, the signal is space-time combined and decoded at the receiver [Alamouti, 1998].[3] Hence, the total received CSNR can be written as

$$\gamma(k;l) = \frac{\mathcal{E}_d\|\mathbf{H}_e\|_{\mathsf{F}}^2}{n_{\mathrm{T}}(N_0 + g\mathcal{E}_d\sigma_e^2(l))}, \tag{6.10}$$

where $\mathbf{H}_e = \mathbf{H} - \boldsymbol{\Xi}_e$ is the matrix containing the estimated channel and $\boldsymbol{\Xi}_e$ is the matrix of the corresponding estimation errors. The elements $[\boldsymbol{\Xi}_e]_{\nu\mu} = \epsilon_{e;\mu\nu}$.

[3] Although the result of the space-time combining scheme is different from the traditional MRC, the resulting effective CSNR is still the same. The difference of these two combining techniques is a complex conjugation of the noise which appears in the space-time combining scheme. This conjugation does not play any role when we take the absolute value; which is exactly what we do when we calculate the CSNR. Thus, the total CSNR still can be found by using the MRC approach.

Similarly to previous chapters, a set of 4-D trellis codes are used as component codes in the numerical examples in Section 6.6. Thus, the BER approximation in (2.8) is employed. Having estimated the channels, the BER now becomes

$$\mathrm{BER}\left(M_n \middle| \left\{h_{e;\mu\nu}\right\}\right) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \exp\left(-A'_n(\ell)\mathcal{E}_d \|\mathbf{H}_e\|_{\mathsf{F}}^2\right), \tag{6.11}$$

where $A'_n(\ell) = A_n(\ell)/n_{\mathrm{T}}$ and $A_n(\ell)$ is given by (4.5a).

With the assumption that the subchannels are independent and recalling the results for the uncorrelated SIMO case (discussed in Chapter 4), we further infer that the overall PDF of $\left\{|h_{e;\mu\nu}|\right\} \mid \left\{h_{p;\mu\nu}\right\}$ is

$$f_{h_e|h_p}\left(\left\{|h_{e;\mu\nu}|\right\} \mid \left\{h_{p;\mu\nu}\right\}\right) = \prod_{\mu=1}^{n_{\mathrm{T}}} \prod_{\nu=1}^{n_{\mathrm{R}}} f_{h_{e;\mu}|h_{p;\mu}}\left(|h_{e;\mu\nu}| \middle| h_{p;\mu\nu}\right).$$

Each subchannel is still Rician distributed with the Rice factor $K = |(1-\varrho)h_{p;\mu\nu}(k;l)|^2/\sigma^2_{h_{e;\mu\nu}|h_{p;\mu\nu}}$. Also, here, we set ϱ—the normalized correlation between the estimation error $\epsilon_{e;\mu\nu}$ and the predicted channel $h_{p;\mu\nu}$—to be zero. Moreover, the estimation error $\epsilon_{e;\mu\nu}(k;l)$ is correlated with the prediction error $\epsilon_{p;\mu\nu}(k;l)$. Invoking the assumption in (6.9), the Rician factor of each subchannel simplifies to $K = |h_{p;\mu\nu}(k;l)|^2/\sigma^2_{h_e|h_p}$ where, by assumption, $\sigma^2_{h_{e;\mu\nu}|h_{p;\mu\nu}} = \sigma_p^2 - \sigma_e^2 = \sigma^2_{h_e|h_p}\ \forall \mu,\nu$.

The BER conditioned on the set of predicted channels is obtained by averaging (6.11) over the product of the Rician PDFs:

$$\begin{aligned}\mathrm{BER}\left(M_n \middle| \left\{h_{p;\mu\nu}\right\}\right) &= \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n_{\mathrm{T}}n_{\mathrm{R}}-\text{fold}} \mathrm{BER}\left(M_n \middle| \left\{|h_{e;\mu\nu}|\right\}\right) \\ &\quad \times f_{h_e|h_p}\left(\left\{|h_{e;\mu\nu}|\right\} \mid \left\{h_{p;\mu\nu}\right\}\right) d|h_{e;1,1}| \cdots d|h_{e;n_{\mathrm{T}},n_{\mathrm{R}}}| \\ &= \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) d'_n(\ell)^{n_{\mathrm{T}}n_{\mathrm{R}}} \exp\left(-A'_n(\ell)d'_n(\ell)\mathcal{E}_d\|\mathbf{H}_p\|_{\mathsf{F}}^2\right),\end{aligned} \tag{6.12}$$

after some straightforward integrations (with the help of [Gradshteyn and Ryzhik, 2000, Eq. (6.633-4)]) where $d'_n(\ell) = 1/(A'_n(\ell)\mathcal{E}_d\sigma^2_{h_e|h_p} + 1)$. It can be seen that when $n_{\mathrm{T}} = 1$ both $A'_n(\ell)$ and $d'_n(\ell)$ coincide with $A_n(\ell)$ and $d_n(\ell)$, and (6.12) reduces to (4.9).

Let $\mathbf{H}_p = \mathbf{H} - \mathbf{\Xi}_p$ be the matrix containing the predicted channels and let $\mathbf{\Xi}_p$ be the matrix of the corresponding prediction errors with the elements $[\mathbf{\Xi}_p]_{\nu\mu} = \epsilon_{p;\mu\nu}$. The total predicted CSNR per symbol can be written

as

$$\hat{\gamma} = \frac{\bar{\mathcal{E}}_d \|\mathbf{H}_p\|_{\mathsf{F}}^2}{n_{\mathrm{T}} N_0} = \frac{\bar{\gamma}_b \bar{\mathcal{E}}_d \|\mathbf{H}_p\|_{\mathsf{F}}^2}{n_{\mathrm{T}} \mathcal{E}}. \tag{6.13}$$

Solving it with respect to $\|\mathbf{H}_p\|_{\mathsf{F}}^2$ and inserting the solution into (6.12) gives

$$\mathrm{BER}(M_n|\hat{\gamma}) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) d_n'(\ell)^{n_{\mathrm{T}} n_{\mathrm{R}}} \exp\left(-\frac{\hat{\gamma} A_n'(\ell) d_n'(\ell) n_{\mathrm{T}} \mathcal{E} \mathcal{E}_d}{\bar{\gamma}_b \bar{\mathcal{E}}_d} \right), \tag{6.14}$$

and the optimal switching thresholds $\{\hat{\gamma}\}_{n=1}^{N}$ are subsequently found by solving $\mathrm{BER}(M_n|\hat{\gamma}) = \mathrm{BER}_0$.

From (6.13) the average predicted CSNR is

$$\bar{\hat{\gamma}} = \frac{\bar{\mathcal{E}}_d \mathsf{E}\left[\|\mathbf{H}_p\|_{\mathsf{F}}^2\right]}{n_{\mathrm{T}} N_0} = \frac{r \bar{\gamma}_b n_{\mathrm{T}} n_{\mathrm{R}}}{n_{\mathrm{T}}} \tag{6.15}$$

where $r = \bar{\mathcal{E}}_d (1 - \sigma_p^2)/\mathcal{E}$ as earlier, and the total predicted CSNR follows the gamma distribution with the mean given in (6.15) [Holter *et al.*, 2003; Ko and Tepedelenlioglu, 2006]. That is: $\hat{\gamma} \sim \mathcal{G}\left(n_{\mathrm{T}} n_{\mathrm{R}}, r\bar{\gamma}_b / n_{\mathrm{T}}\right)$.

Similar to (2.6), using STBC, the average BER is still defined as the ratio between the average number of bits received in error, and the number of bits transmitted in total:

$$\overline{\mathrm{BER}} = \frac{\sum_{n=1}^{N} \mathrm{BER}(M_n) \cdot R_n^{\mathrm{STBC}}}{\sum_{n=1}^{N} P_n \cdot R_n^{\mathrm{STBC}}}, \tag{6.16}$$

where R_n^{STBC} is the SE of the nth constellation after STBC (to be derived in Section 6.5),

$$\begin{aligned}
\mathrm{BER}(M_n) &= \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathrm{BER}(M_n|\hat{\gamma}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \\
&= \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \left(\frac{d_n'(\ell) \bar{\mathcal{E}}_d}{r d_n'(\ell) A_n'(\ell) \mathcal{E} \mathcal{E}_d + \bar{\mathcal{E}}_d} \right)^{n_{\mathrm{T}} n_{\mathrm{R}}} \\
&\quad \times \left\{ \overline{\Gamma}\left(n_{\mathrm{T}} n_{\mathrm{R}}, n_{\mathrm{T}} \hat{\gamma}_n \frac{r d_n'(\ell) A_n'(\ell) \mathcal{E} \mathcal{E}_d + \bar{\mathcal{E}}_d}{r \bar{\gamma}_b \bar{\mathcal{E}}_d} \right) \right. \\
&\quad \left. - \overline{\Gamma}\left(n_{\mathrm{T}} n_{\mathrm{R}}, n_{\mathrm{T}} \hat{\gamma}_{n+1} \frac{r d_n'(\ell) A_n'(\ell) \mathcal{E} \mathcal{E}_d + \bar{\mathcal{E}}_d}{r \bar{\gamma}_b \bar{\mathcal{E}}_d} \right) \right\},
\end{aligned} \tag{6.17}$$

and P_n is the probability that $\hat{\gamma} \in [\hat{\gamma}_n, \hat{\gamma}_{n+1}\rangle$:

$$P_n = \overline{\Gamma}\left(n_{\mathrm{T}} n_{\mathrm{R}}, \frac{n_{\mathrm{T}} \hat{\gamma}_n}{r \bar{\gamma}_b} \right) - \overline{\Gamma}\left(n_{\mathrm{T}} n_{\mathrm{R}}, \frac{n_{\mathrm{T}} \hat{\gamma}_{n+1}}{r \bar{\gamma}_b} \right). \tag{6.18}$$

6.5 Optimization of ASE

The SE of the nth constellation used by a 4-D trellis code with a PSAM scheme is $R_n = (1 - 1/L)(\log_2(M_n) - 1/2)$. After STBC using the orthogonal designs in Tab. 6.1, the effective SE becomes

$$\begin{aligned} R_n^{\text{STBC}} &\overset{(a)}{=} \left(\log_2(M_n) - \frac{1}{2}\right) \frac{L_b - n_{\text{T}}}{L_b} R_s \\ &= \left(\log_2(M_n) - \frac{1}{2}\right) \frac{(L-1)R_s}{L - 1 + n_{\text{T}} R_s}. \end{aligned} \tag{6.19}$$

The term outside of the parenthesis in the equality marked with (a) corresponds to the fact that if the frame length L_b is equal to n_{T}, then no data information is transmitted. Thus, the SE must be zero. Using orthogonal STBC, there is a rate penalty for complex signals when $n_{\text{T}} > 2$. That explains the rate R_s in the above expressions. The second equality is obtained by using $L_b = (L-1)/R_s + n_{\text{T}}$ (introduced in Eq. (6.2)).

Hence, the overall ASE is given by

$$\text{ASE} = \sum_{n=1}^{N} R_n^{\text{STBC}} \cdot P_n. \tag{6.20}$$

Before invoking Algorithm 1 on page 42 to do the maximization, the following choices are made. It is obvious that the variance of the *prediction* error is largest when predicting the last symbol in a frame ($l = L_b - 1$). As a result, prediction of the symbol located at the end of the frame based on the pilots transmitted from, e.g., the fourth antenna is slightly more accurate than it would be if based on pilots from the first antenna (cf. Figure 6.2). On the other hand, the variance of the *estimation* error is almost the same for all l, if the order of the estimator is $K_e \geq 20$ [Cai and Giannakis, 2005]. Thus, we use the *estimation* error variance $\sigma_e^2 = \sigma_{e;\mu v}^2(L_b - 1)$, and the conservative choice of the *prediction* error variance $\sigma_p^2 = \sigma_{p;1v}^2(L_b - 1)$—note the subscript index—when finding the optimal switching thresholds $\{\hat{\gamma}_n\}_{n=1}^N$, as well as in the further optimization process.

It is noted that when both $n_{\text{T}} = 1$ and $n_{\text{R}} = 1$, the analysis in this chapter reduces to the SISO case in Chapter 3. The uncorrelated SIMO system in Chapter 4 is obtained if $n_{\text{T}} = 1$ and $n_{\text{R}} > 1$. Furthermore, when $n_{\text{R}} = 1$ and $n_{\text{T}} > 1$ we have a MISO system.

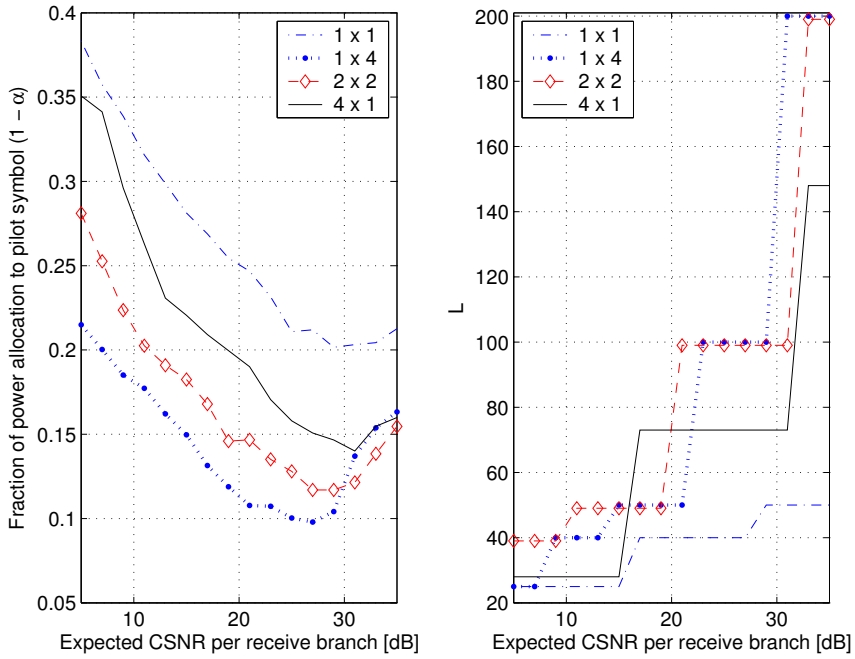


FIGURE 6.3: Left panel: Optimal fraction of power allocated to pilot symbols (i.e. $1-\alpha$) when the pilot period L is optimal. Right panel: Optimum pilot spacing L when the power is optimally allocated between pilot and data symbols. Both figures are generated for different combinations of transmit and receive antennas.

6.6 Numerical Example and Discussion

The same set of parameters as in Section 2.7 are used when giving numerical examples. However, since STBC is used when the number of transmit antennas is greater than 2, some modifications must be done. In particular, similarly to Section 2.7, we have $\tau = DL_bT_s = 1$ ms and, at the same time, $L_b > n_{\mathrm{T}}$. Applying (6.2) the following results are obtained:

- $n_{\mathrm{T}} = 1$: $L = L_b \in \{2, 4, 5, 8, 10, 20, 25, 40, 50, 100, 200\}$.
- $n_{\mathrm{T}} = 2$: $L_b \in \{4, 8, 10, 20, 40, 50, 100, 200\}$.
  Thus, $L \in \{3, 7, 9, 19, 39, 49, 99, 199\}$.
- $n_{\mathrm{T}} = 4$: $L_b \in \{8, 20, 40, 100, 200\}$. Thus, $L \in \{4, 13, 28, 73, 148\}$.

How power is optimally allocated to pilot symbols and how the optimal pilot symbol spacing L is distributed with the average CSNR can be read from Figure 6.3. The amount of power allocated to data symbols is

also easily read from that figure.[4] Clearly, the pilot period increases with increasing CSNR and it increases faster with higher diversity order. Moreover, it is apparent that, in most of the CSNR region, necessary pilot power decreases with increasing average CSNR—i.e. more power should be put on data symbols when the average CSNR is increased. The more antennas there are—either on the transmitter side or on the receiver side, or on both sides—the less power is allocated to pilot symbols. This is also to be expected since the antennas in this particular system are used to increase the diversity order, i.e. to stabilize the channel. For the same diversity order (the product $n_{\mathrm{T}} \times n_{\mathrm{R}}$), most of the power is left for data symbols in the SIMO case. Having higher transmit diversity, the pilot power *on each antenna* is reduced when the average total transmit power $\mathcal{E}$ is fixed. Thus, if the pilots are to "survive" after the transmission over the individual channels, more *total* power must be assigned on them. That is the case for $n_{\mathrm{T}} = 4$ in the left panel of Figure 6.3. Seeing this together with the ASE performance in Figure 6.4, from an ASE point of view the SIMO solution is clearly preferred to other combinations of transmit and receive antennas yielding the *same diversity order*.

In [Holter *et al.*, 2003], the ASE was always reduced when having two transmit antennas compared to when only one transmit antenna is employed. In contrast to this, Figure 6.4 shows that the ASE is increased by going from 1 transmit antenna to 2 transmit antennas, as long as the pilot spacing and the power distribution are optimal. In general, when using STBC, the channel capacity is reduced, except when either the rate of the employed space-time block code is one, or the channel is rank one [Sandhu and Paulraj, 2000]. In our example, only the orthogonal design $\mathcal{O}_2$ for $n_{\mathrm{T}} = 2$ has rate one (no STBC is necessary for $n_{\mathrm{T}} = 1$). Optimization of the system with 2 transmit antennas also gives a larger pilot spacing and a lower pilot power. As a result, in this case, the ASE becomes higher compared to the system with only one transmit antenna. This result agrees well with [Ko and Tepedelenlioglu, 2003].

For comparison purposes, we include two plots of ASE corresponding to 1) optimal power and optimal pilot period, and 2) equal power allocation and optimal pilot period. They are depicted on the left and right panel of Figure 6.5, respectively. Also, here, the gain by having optimal L and α (optimal power distribution) is larger. The gain is up to approximately 0.5 bits/s/Hz. In the latter case, both pilot spacing L and fraction of power

[4] It is noted that although the analysis in [Duong *et al.*, 2005] is valid, unfortunately, the results are subject to an error when implementing [Duong *et al.*, 2005, Eqs. (4) and (5)] in MATLAB (the term n_{T} is missing). Thus, the results given here are the correct ones.

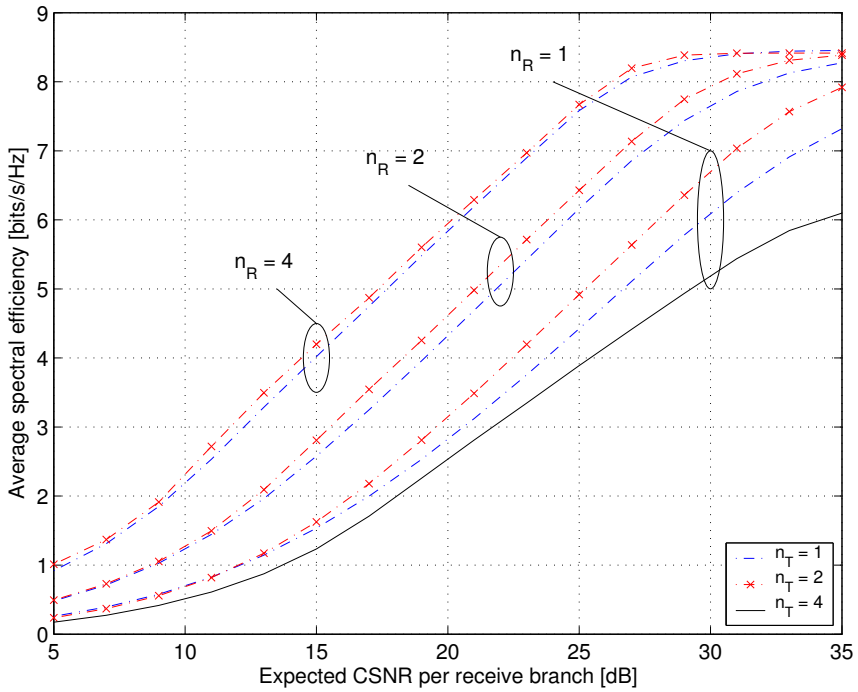


FIGURE 6.4: Average spectral efficiency for different combinations of transmit and receive antennas ($n_T \times n_R$): (1×1), (1×2), (1×4), (2×1), (2×2), (2×4), and (4×1), respectively. The curves are generated when both power distribution between pilot and data symbol and pilot spacing are optimal.

allocation to the pilots are also smaller (not shown here).

When we have 4 transmit antennas, the employed space-time block code only has rate 3/4; hence, some symbols must be transmitted several times, and the throughput is significantly reduced due to that loss. When performing channel estimation and prediction using PSAM in a MIMO diversity system with n_T transmit antennas, the overall number of pilot symbols to be transmitted is n_T times the number of pilot symbols that are needed in the non-MIMO case [Alamouti, 1998] (either by using orthogonal pilot symbols, or the pilots must be transmitted one at a time from each antenna). Moreover, using the pilot transmission scheme in this paper (or spreading one pilot symbol as in [Zhou and Giannakis, 2004b]) the system is losing $n_T(n_T - 1)$ symbol intervals where data symbols could be transmitted such that the system performance could be increased.

The $\overline{\text{BER}}$ performance is shown in Figure 6.6. As observed, our requirement of $\text{BER}_0 = 10^{-5}$ is always satisfied since we require the *instantaneous*

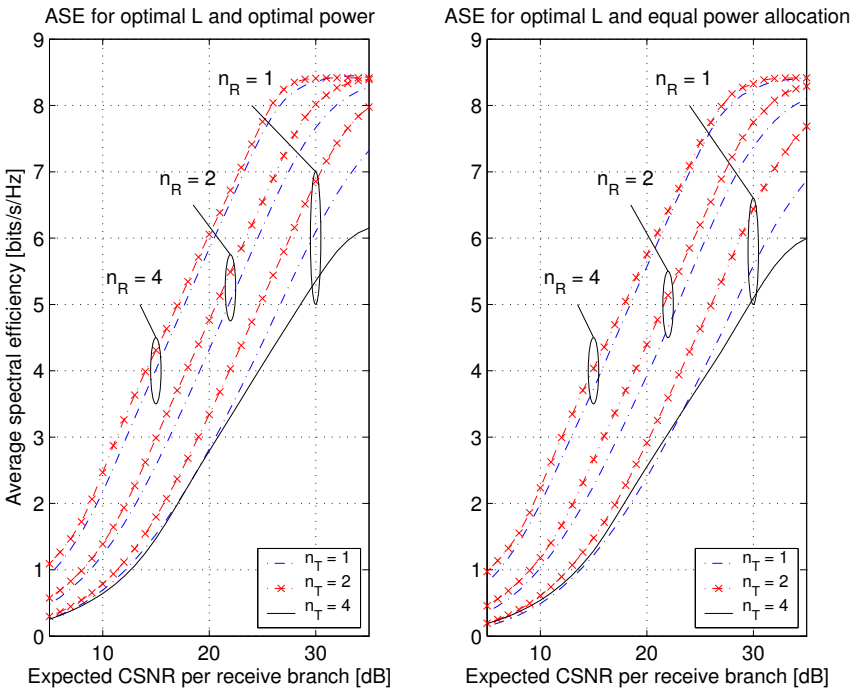


FIGURE 6.5: Comparison of ASE when both power allocation and pilot spacing are optimal (left panel) with optimal pilot spacing and equal power allocation (right panel).

BER of the system to be always below BER_0. On the other hand, the curves are unnecessarily far below the requirement. To reduce that gap, we can modify the constraint in such a way that the *average* BER (instead of the instantaneous BER) must be lower than BER_0. In that way, the throughput increases and the system becomes less sensitive to the time delay, as [Ko and Tepedelenlioglu, 2004] concluded. Optimization of switching thresholds with respect to average BER constraint is also analyzed in [Torrance and Hanzo, 1996; Hanzo, Münster, Choi, and Keller, 2003, Chap. 12]. Figure 6.6 also confirms the fact that orthogonal STBC gives full spatial diversity order (the product $n_\text{T} \times n_\text{R}$) by looking at the slope of the $\overline{\text{BER}}$ curves at high CSNR.

Shown in Figure 6.7 are the curves for outage probability for different $n_\text{T} \times n_\text{R}$ combinations. It illustrates both the SISO system of Chapter 3, uncorrelated SIMO system of Chapter 4, the correlated SIMO system of the second part of Chapter 5 and MIMO diversity system in this chapter.

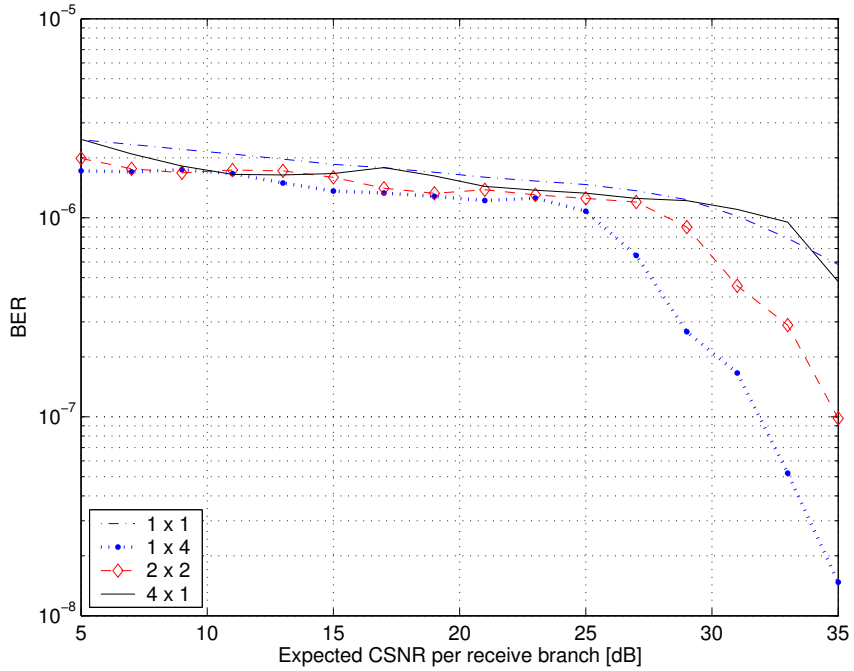


FIGURE 6.6: Average BER when the power and the pilot spacing are optimal for different combinations of transmit and receive antennas.

For a fixed radiated power constraint—which is the case for the analysis in this thesis—clearly, the SIMO system gives lower outage probability. Thus, it is preferable to the MISO system with the same diversity order $n_T \times n_R$. This conclusion was also drawn from an ASE point of view in Figure 6.4, since the ASE is considerably larger in a SIMO system compared to a MISO system for the same product $n_T \times n_R$. In the uncorrelated branches system and with the same diversity order, probability of outage suffers a 3-dB loss for each doubling of the number of transmit antenna. This performance penalty will disappear if we double the transmit power with the doubling of the number of transmit antennas [Alamouti, 1998].

Moreover, spatial correlation gives rise to degradation of diversity gain. As shown in Figure 6.7 the curves for correlated receive antennas in a 1×4 system reduces gradually to the 1×1 system with increasing ρ_s. Note that, as with the ASE performance, it will never coincide with the 1×1 scenario due to the array gain.

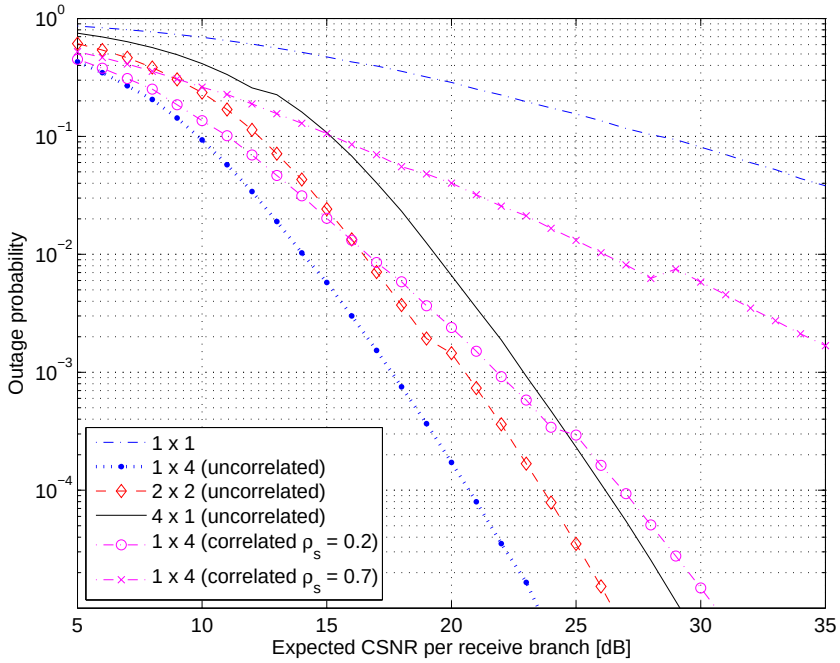


FIGURE 6.7: Outage probability depicted for optimal L and optimal power allocation. It is plotted as a function of average CSNR on each branch and different combinations of transmit-receive antennas.

6.7 Concluding Remarks

We have analyzed and optimized an ACM system operating on a MIMO diversity channel. The throughput in terms of ASE, when the transmitter is equipped with 2 antennas, outperforms the same system with only one transmit antenna. This is due to the diversity gain and to the fact that the employed orthogonal design for STBC has full rate (rate 1). This result is in contrast to what was obtained in [Holter *et al.*, 2003] which is due to the optimization of pilot period and power performed in the present work. Having more than 2 transmit antennas gives even higher diversity order, but the overall rate is reduced due to the rate loss of the employed space-time block code.

In conclusion, for a fixed product $n_T \times n_R$ (the same diversity order), the ASE is still always highest for $n_T = 1$. This indicates that when ACM is used, it is best to exploit spatial diversity by means of multiple receive antennas. In this case, even with high mobile terminal speeds, the system is, on the average, using the same constellation a reasonably long time before switching to another constellation (cf. Tab. 4.1).

Chapter 7

Effects of Simplifications and Suboptimalities

In a practical implementation, it is desired to keep the complexity of any communication system as low as possible. To this point, both pilot spacing L and power allocation between pilot and data symbols are optimized. Thus, which spacing the pilots will have for a given average CSNR is predicted at the receiver and must be fed back to the transmitter. While this information is dependent only on average CSNR and, hence, slowly varying, it still needs to be done. Especially, in a multiuser system, where all the users transmit their own CSI, the feedback load can be huge. Hence, it motivates us to reduce the feedback load. In the next section, by a numerical example, we will study the effect of fixing the pilot symbol period and only optimize the power distribution. In this way, feedback information about the pilot spacing can be omitted.

Furthermore, we will also address the robustness of having shorter predictor length when both power distribution and pilot spacing are optimal. The predictor length used in the thesis is not optimal, but is satisfactory [Øien *et al.*, 2002]. In the just-mentioned paper, where both pilots and data symbols are transmitted with the same power, and pilot spacing is fixed, it is demonstrated that the predictor order may be reduced to a feasible complexity without a too dramatic effect on the system performance. Note however that, this is valid only for systems with limited mobility. In this chapter, we will decrease the order even further to see what impact it has on the system performance.

To ease mathematical analysis, perfect estimation is commonly assumed in many works. Later on, we will demonstrate what impact this assumption has on the prediction performance for our spatially uncorrelated SIMO

system.

We are not providing an analytical analysis in this chapter, but rather some numerical examples to illustrate the potential of different simplifications. Only spatially uncorrelated subchannels is considered. The effects of fixing the pilot spacing is presented in Section 7.1, whereas the impact of a complexity-limited predictor on the system performance is given in Section 7.2. Finally, how the analytic simplification of perfect estimation would impact the system performance is discussed in Section 7.3.

7.1 Fixing the Pilot Spacing

In our systems, the feedback CSI is already reduced to only a quantized version of of the predicted CSNR. The purpose of optimization in all the previous chapters is to avoid the problem of having a short pilot period at high average CSNR and having a too large pilot spacing at low average CSNR, which would give rise to a waste of bandwidth when the channel condition is good, and bad channel estimation/prediction at bad channel conditions, respectively. Thus, the pilot spacing has also to be sent back to the transmitter when the average CSNR changes (as in the case in shadowing conditions, where the average gain is a slowly varying RV) since it is required for optimal performance. However, in what follows, we will simplify it further by using a fixed pilot spacing; meaning that no feedback information on the pilot period is needed.

Plotted in Figure 7.1 are the curves of ASE for optimal power distribution and L, and optimal power distribution with L fixed to certain values. Thus, in the latter case, the system is optimized with respect to transmit power allocation between pilot and data symbols only. It can be seen that degradations due to fixed pilot spacing are small, implying that fixing the pilot period to these values does not limit system performance. The values of L are chosen around the average value of the optimum values (cf. Figure 4.2 and Figure 6.3).

Evidently, the curves of optimal pilot power in Figure 7.2 show that the achievements of ASE in Figure 7.1 are due to the somewhat increasing pilot power, at least at low average CSNR (bad channel conditions) where the spacing between two pilots is larger compared to when L is optimal. At high average CSNR (good channel condition), the pilot period is smaller than in the optimal case. Thus, lower pilot power can be used.

It can be concluded that, for certain fixed choices of L, the system performance is very close to that of optimal L at the expense of only a little more power allocated to the pilot symbols at low average CSNR. It is a

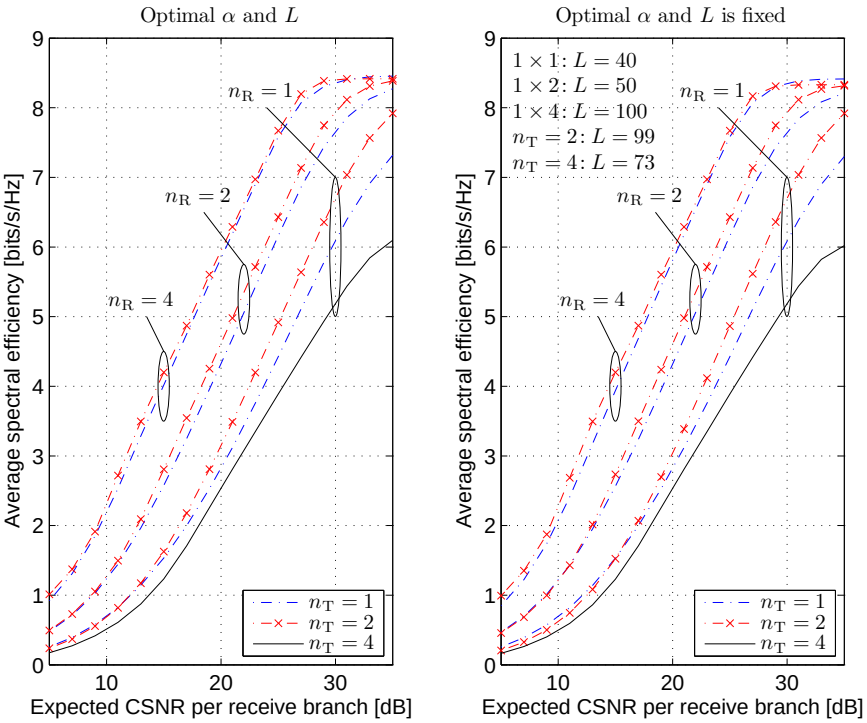


FIGURE 7.1: ASE plotted against subchannel's average CSNR. Left panel: it is plotted for optimal α and optimal L. Right panel: it is plotted for optimal α, but L is fixed to certain values depends on the combination of numbers of transmit and receive antennas. Both figures are generated for: predictor and estimator order is $K_p = 250$ and $K_e = 20$, respectively.

good trade-off in order to reduce the feedback rate and to keep the system design simple.

7.2 Effect of Complexity-Limited Predictor

In addition to reduce the system feedback rate, the complexity in terms of computation is also an important issue of handheld devices for an acceptable processing delay. In [Holm, 2002; Øien *et al.*, 2004] a predictor of length $K_p = 1000$ is used, something that is clearly too complex for most practical situations. The order can be substantially reduced as shown in [Øien *et al.*, 2002] where the pilot is transmitted with the same power as a data symbol and L is fixed. Here, we will present some plots showing that when the system parameters are optimized (which was not the case in [Øien *et al.*,

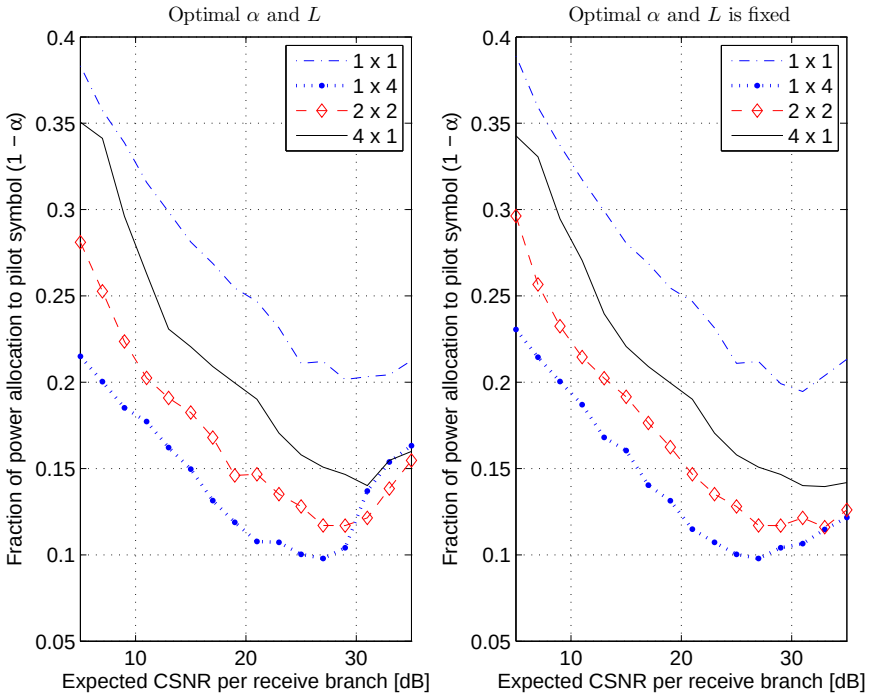


FIGURE 7.2: Optimal fraction of power allocation to pilot symbol is plotted as a function of subchannel's CSNR. Left panel: it is plotted for optimal α and optimal L. Right panel: it is plotted for optimal α, but L is fixed to certain values depends on the combination of numbers of transmit and receive antennas. Both figures are generated for: predictor and estimator order is $K_p = 250$ and $K_e = 20$, respectively. The chosen L values are given in Figure 7.1.

2002]), the predictor order can be further reduced, which only contribute to a small decrease in system performance.

Shown in Figure 7.3 are the ASE curves when the predictor order (number of filter taps) is 250 and 30, respectively. Clearly, the ASE degradation is the most severe at low average CSNR or with low diversity order, meaning that a longer predictor should be used in these cases. It can be concluded that, if the constraint is subject to computation complexity, then the predictor length can be dramatically shortened at only a small expense of reduced performance. Since the predictor is based on a small number of samples to predict the channel, the pilot symbols are transmitted with higher power than in the optimal case. Hence, the pilot spacing can be somewhat larger. These effects are seen in Figure 7.4 at low and medium average CSNRs. When channel conditions are good, the variation of L seems to be the same

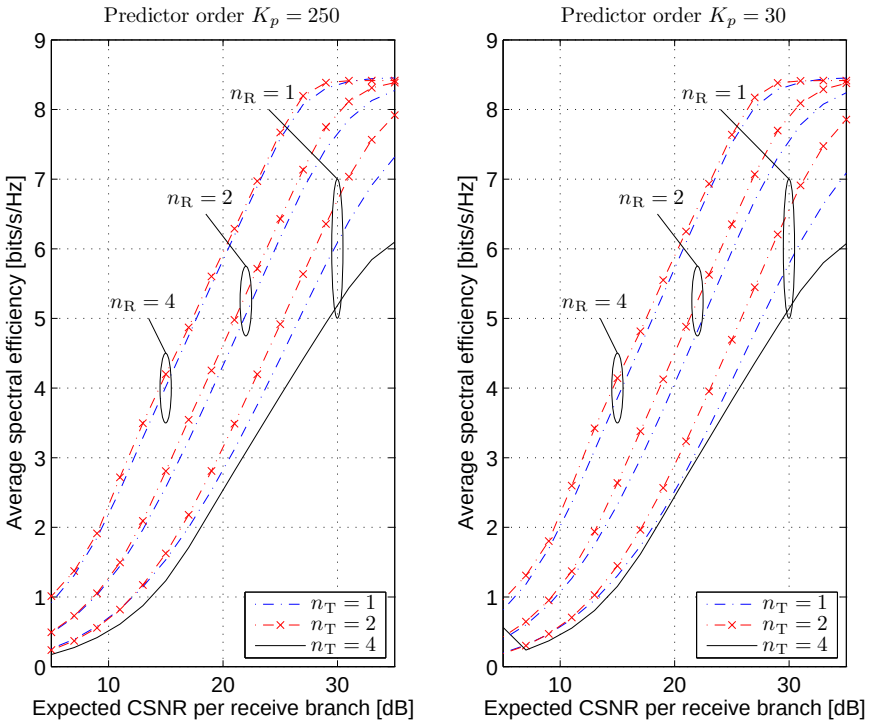


FIGURE 7.3: Comparision of ASE curves for $K_p = 250$ and $K_p = 30$, respectively. The noncausal estimator is still of 20 taps. Here, both α and L are optimized. In the right panel, the lower end of the 4×1 case is subject to numerical instability as can also be observed in Figure 7.4.

for both predictor orders. Moreover, it is observed that the pilot spacing does not vary that much in the case of complexity-limited predictor. The behavior of the plot of L in Figure 7.4 (lower right corner) motivates us further to fix the pilot spacing for low predictor orders, similar to what we did in Section 7.1.

7.3 Effect of the Assumption on Perfect Estimation

Since channel estimation in the receiver can be done with high precision, perfect estimation is usually imposed for a simpler analysis. However, the estimation process of course still needs to be performed in practical implementations. A nature question to ask is then which kinds of errors such an assumption will induce in the system performance predicted by the simplified analysis. By comparing the results of analysis with and without the simplifying assumption, we can answer this question. Here, in Fig-

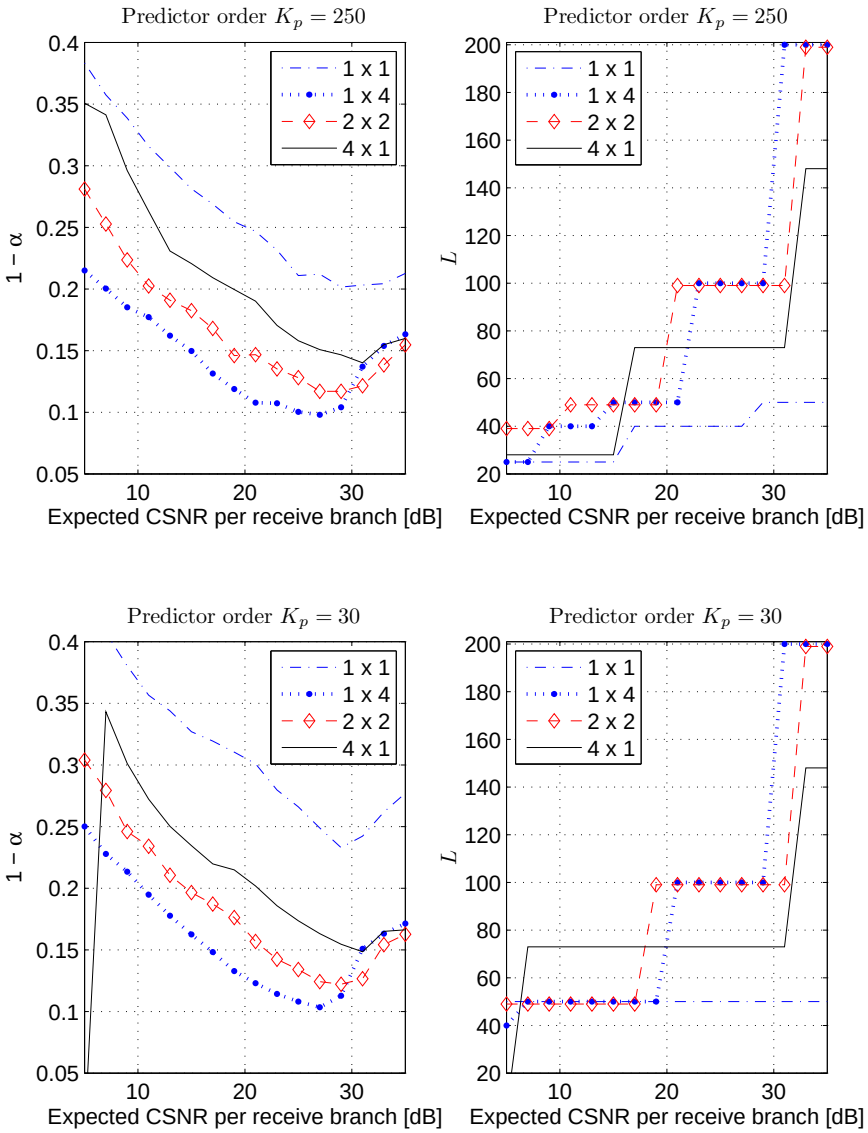


FIGURE 7.4: Comparision of power allocation on pilot symbol $(1-\alpha)$ and L curves for $K_p = 250$ and $K_p = 30$, respectively. The noncausal estimator is still of 20 taps. Here, both α and L are optimized.

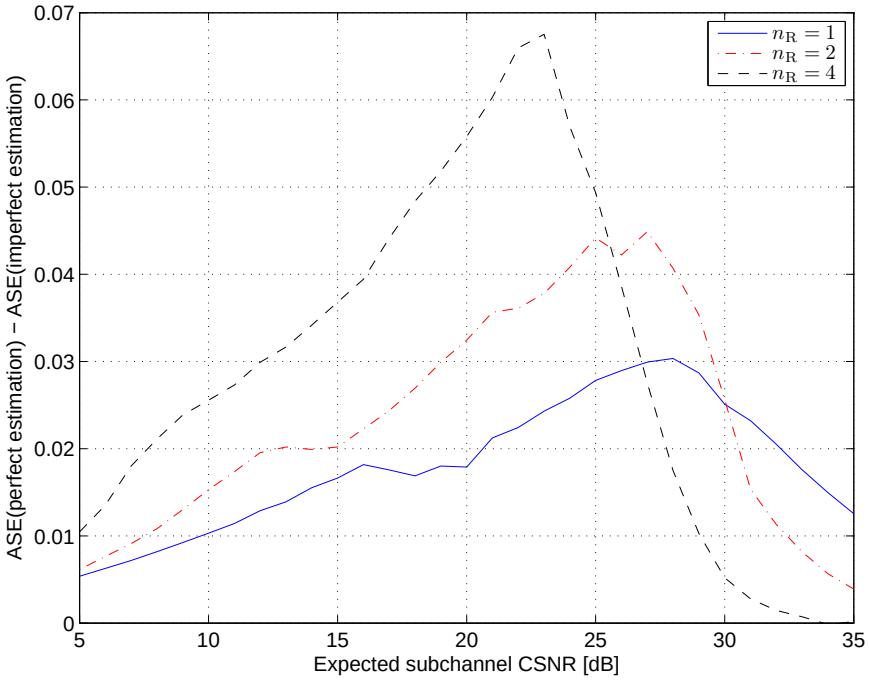


FIGURE 7.5: Difference of ASE corresponding to the perfect and imperfect estimation. Both L and power allocation constant α are optimal and $K_p = 250$.

ure 7.5, we show a plot comparing the ASE performance of the perfect and imperfect estimation cases of our coded system. As expected, perfect estimation results in a better predicted performance. This behavior can be forecasted by inspecting Eq. (4.3). In that equation, once estimation is perfect, $\sigma^2_{e;\mu} = 0$ and $h_{e;\mu}(k;l) = h_\mu(k;l)$. Because of the orthogonality principle, $|h_{e;\mu}(k;l)| \leq |h_\mu(k;l)|$. Thus, the combined CSNR is greater compared to when estimation error is presented. However, as depicted in the figure, the difference is small; which justifies the use of this assumption during analysis.

In the multiplot in Figure 7.6 we include the plots for optimal fraction of power allocation on pilot symbols and pilot spacing L for both perfect (left column) and imperfect (right column) channel estimation cases. The perfect estimation assumption causes somewhat lower pilot power, whereas the pilot period is almost the same. Because of that low pilot power, the optimal pilot spacing is shifting to a higher value at higher average CSNR level. In conclusion, optimizing a system with perfect estimation assumption will overestimate ASE and underestimate pilot power somewhat, but

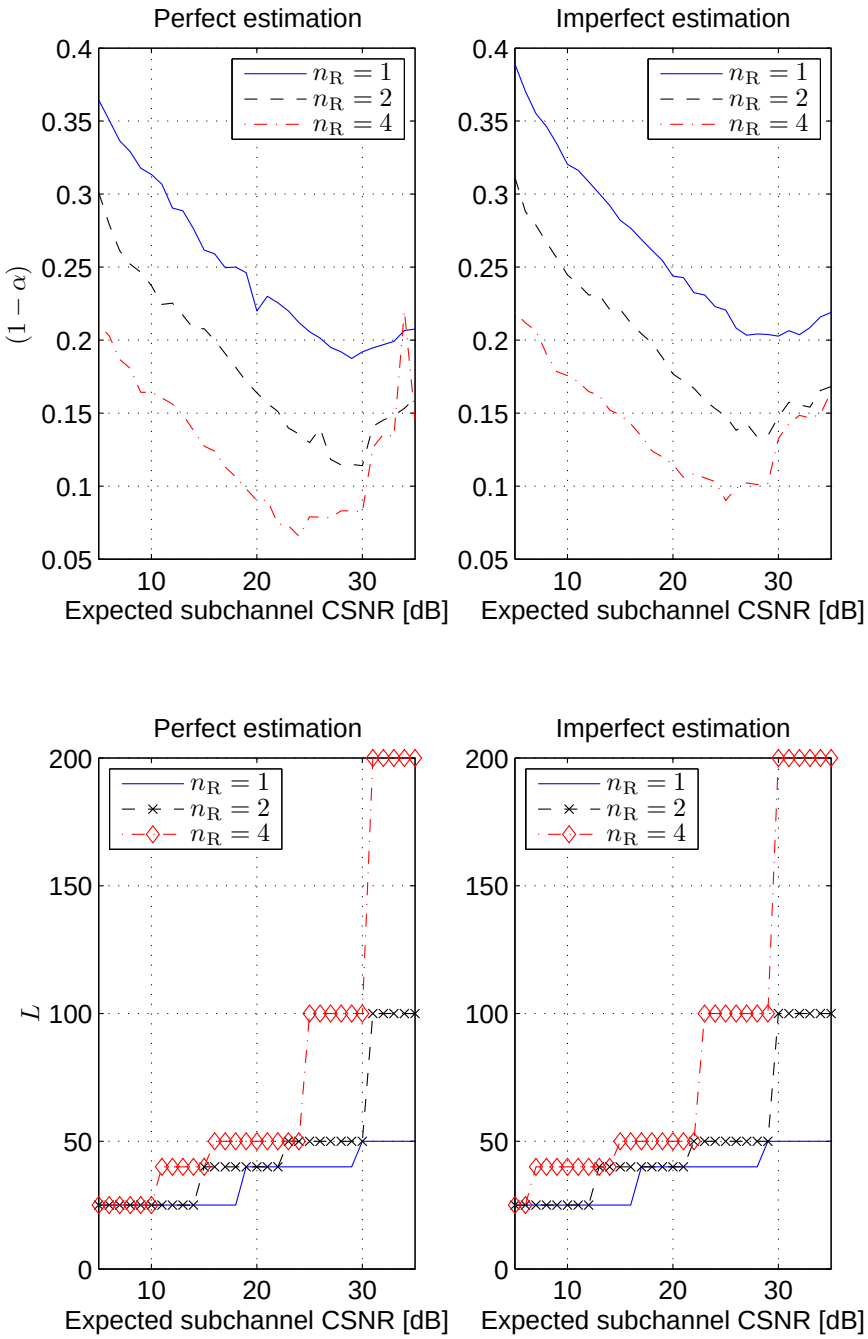


FIGURE 7.6: How optimal power allocation on pilot symbols corresponding to perfect estimation (top left plot) differs from that of imperfect estimation (top right plot) is depicted. Variations of optimal pilot spacing L when estimation is perfect (bottom left plot) and not perfect (bottom right plot). Here, the predictor order $K_p = 250$.

not very much. This indicates that we freely can make that assumption to simplify the mathematical analysis if we keep in mind that the results are a little too optimistic.

Chapter 8

Conclusions

8.1 Results and Contributions

In this dissertation, an ACM technique based on adaptive PSAM is analyzed and optimized in both SISO, spatially uncorrelated and correlated SIMO, and MIMO diversity systems to overcome the hostile fading nature of the wireless channel and to increase the spectral efficiency of wireless systems. The analysis and optimization have been done under practical conditions such as predicted and estimated CSI. The objective of the adaptive transmission is to increase the throughput and, at the same time, keep the BER below a pre-defined threshold. We have demonstrated the merit of having the pilot spacing and the power distribution between pilot and data symbols optimized with respect to average CSNR: higher throughput is achieved without sacrificing BER performance compared to non-optimized systems. Moreover, in presence of known spatial correlation in a multi-antenna system, the rate loss is substantially reduced when having all the branches jointly predicted, taking spatial correlation into account.

By numerical examples, we have also presented the effects on performance of shortening the predictor order and of fixing the pilot spacing. Both the simplifications contribute to reducing the complexity of the receiver and the system, since less computation is needed in the prediction process and no pilot adaptivity is needed (hence, less feedback rate). The results of ASE with a fixed L is very close to what is obtained by having L optimally varied with average CSNR. Letting both pilot period and power allocation on these pilots be optimal, the impact of reducing the predictor order is also small. Note that, in both cases, the pilot power is however somewhat higher than required for optimal performance.

The impact of assuming perfect estimation during system analysis is

also considered for the spatially uncorrelated SIMO case. As expected, the analysis in that case overestimates ASE somewhat, but not much. As such, this shows and justifies that this assumption can be made to ease mathematical analysis of such systems.

8.1.1 Contributions of the Thesis

A summary of the main contributions of each chapter are listed as follows.

- In Chapter 2, the principle of our ACM system is presented where different building blocks are mapped into our cases. Furthermore, the MAP-optimal linear filtering is re-develop there and visualizations of the predictor's performance are also presented.

- In Chapter 3, a generalization of the work by Cai and Giannakis [2005] were presented, where channel coding using trellis codes is introduced. The pilot spacing and power allocation between pilot and data symbols are optimized to maximize ASE without sacrificing BER performance. The optimization algorithm introduced in this chapter is simpler than the one used by [Cai and Giannakis, 2005]. As expected, higher ASE is obtained due to optimization. An improvement of 3 dB is also achieved due to coding. This chapter serves at a foundation for further analysis in the rest of the thesis since the optimization introduced in this chapter is also invoked to find optimal solutions in other chapters.

- In Chapter 4, the above scheme was extended further to include multiantenna reception and MRC of received signals. Here the subchannels are assumed to be mutually independent of each other. In contrast to [Cai and Giannakis, 2005] (and to the previous chapter), we also considered the estimation error to be *correlated* with the prediction error. This is due to the fact that both estimator and predictor are using partly the same information.

  Comparison to a similar system ([Øien *et al.*, 2004]) where pilot symbol spacing is fixed, and where both data and pilot symbols are transmitted with the same power is made. In this case, it is shown that our system is always superior to the one of [Øien *et al.*, 2004].

  Also, the average time during which the channel remains within a fading interval is provided to quantify how often system adaptation is required on the average. In other words, it is the same as saying how often the system needs to reconfigure its transmission mode.

These results can be used to determine how often feedback is required in an adaptive transmission system.

- In Chapter 5, spatial correlation at the receiver was introduced in the above system. The analysis is carried out to reflect practical systems where the antennas might be closely spaced, or the scatterings are insufficient. At first, a predicted system designed for uncorrelated antenna branches is used. This is clearly suboptimal, but it gives performance to be expected when the spatial correlation is not known or not exploited. The different branches are then jointly predicted taking spatial correlation into account. For this purpose, we derived a MAP-optimal "space-time predictor". The rate loss compared to the former case is substantially reduced.

- In Chapter 6, we transmit the same data over multiple transmit antennas. In combination with the multireception at the receiver we thus have a MIMO diversity system. This system is analyzed and optimized when all the subchannels are assumed mutually uncorrelated and orthogonal space-time block codes (OSTBCs) are use. Due to the rate loss of the employed OSTBC for $n_{\mathrm{T}} > 2$ and the silent time slots during pilot transmissions, the resulting ASE performance is reduced. For the same diversity order ($n_{\mathrm{T}} \times n_{\mathrm{R}}$) and from an ASE performance point of view, SIMO diversity is always preferred.

- In Chapter 7, by numerical examples, we demonstrate the effect of limiting the complexity of the predictor and of fixing the pilot period. It is shown that both predictor order and pilot spacing can be respectively limited and fixed at certain values without losing much of the performance, at the expense of high pilot power at low CSNRs. Thus, it enables good trade-off in practical systems where both the feedback rate and the computational complexity are limiting factors. The impact of assuming perfect estimation in spatially uncorrelated SIMO systems is also presented where the results show and justify that this assumption can be made during theoretical analysis of such systems.

8.2 Suggestions for Further Research

In the following list, we highlight some further research topics that might be of interest.

- In multiuser systems, many users are feeding back their CSI, and the total feedback rate can thus be huge. A natural extension of the

research on single-user and single-carrier systems presented in this thesis is to examine this scheme in a multiuser multi-carrier system, where the feedback information of one user can be co-optimized with that of other users. This might be used to reduce the feedback load by means of user cooperation.

- It can be interesting to extend/generalize our optimized system to the multi-carrier OFDM case, to mitigate time- and frequency-selective fading channels in a practical ACM system. Some research contributions on non-optimized PSAM-based channel estimation for multi-carrier systems are reported in [Hoeher, Kaiser, and Robertson, 1997; Li, 2000]. The optimal PSAM-based training parameters (pilot grid arrangement and power allocation between information bearing symbols and pilot symbols) to maximize the capacity lower bound are considered in [Ma, Giannakis, and Ohno, 2003; Ohno and Giannakis, 2004].

- ACM systems are designed to combat fading and to avoid unnecessarily retransmissions in wireless communications. The ASE performance in most of the contributions to the field of adaptive transmission at the physical layer, including ours, are analyzed assuming no retransmission is used. However, due to imperfect channel knowledge, there is always a certain probability for choosing other constellations which are not optimal for transmission. It then results in errors not accounted for in our analysis, and the data must be retransmitted. This issue is considered in [Liu, Zhou, and Giannakis, 2004] for a single-user SISO system. An interesting topic for further research is to combine our optimization system for the MIMO case with their cross-layer combining technique, to reflect the real world where retransmissions might be necessary.

- We have assumed that the feedback channel is error free. However, error can occur and may limit the practicality of ACM systems. Thus, the effect of errors on the feedback channel is an interesting topic for further research within the context of adaptive modulation and coding systems. Current research within this area is reported in [Ekpenyong and Huang, 2004, 2005].

- In the case of correlated antenna branches, we have to approximate the PDF of both true CSNR and predicted CSNR to be gamma PDFs. As such, the joint PDF can be described by the standard bivariate gamma PDF. The approximation is used in spite of the fact that the

exact *marginal* PDF is already reported in the literature [Aalo, 1995; Alouini *et al.*, 2001], whereas the *joint PDF* is still unknown. Therefore, it is of high interest to derive the joint PDF $f_{\gamma,\hat{\gamma}}(\gamma, \hat{\gamma})$ such that the correlation between predicted and true CSNR can be modelled exactly.

- All the results in this thesis are theoretically derived or based on simulation results from nonfading AWGN channels. In all theoretical derivations, some assumptions are made. To see the overall impact of these assumptions on the performance analysis, a comparison between theoretical and simulated results on fading channels models should be done.

Appendix A

Derivation of the MAP-Optimal Predictor and Estimator with the Corresponding MMSE

In this appendix we will derive in detail the MAP-optimal predictor and the corresponding prediction MMSE using the PSAM scheme. The predictor utilized in Chapter 5 will be calculated since it is the most general one. The predictors used in the other papers can be deduced from this.

The MAP-optimal estimator is noncausal and can be found in the same way. The noncausality implies that we also need some future information in the calculation, which can be obtained by using buffers. Hence we will omit to derive the estimator here, and only concentrate on the approach for finding the predictor.

A.1 MAP-optimal Predictor and MMSE of the Prediction Error

Let the ML estimate of the received pilot symbol from all the antennas be $\underline{\tilde{\mathbf{y}}} = \mathrm{vec}(\boldsymbol{\mathcal{Y}}) / (s\sqrt{\mathcal{E}_{\mathrm{pl}}})$ where $\mathrm{vec}(\boldsymbol{\mathcal{Y}})$ is the operator which stacks the columns in the matrix $\boldsymbol{\mathcal{Y}}$ on top of each other [Moon and Stirling, 2000] and $\boldsymbol{\mathcal{Y}}$ is as in (5.17). Let the channel gain matrix and noise matrix corresponding to the received matrix $\boldsymbol{\mathcal{Y}}$ be $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{N}}$, respectively. The correlation

matrix of $\underline{\tilde{\mathbf{y}}}$ can be found as

$$\mathbf{R}_{\underline{\mathbf{y}}} = \mathsf{E}\left[\underline{\tilde{\mathbf{y}}}\underline{\tilde{\mathbf{y}}}^{\mathsf{H}}\right] = \mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\underline{\mathbf{h}}_{\mathrm{pl}}^{\mathsf{H}}\right] + \frac{1}{|s|^2\mathcal{E}_{\mathrm{pl}}}\mathsf{E}\left[\underline{\mathbf{n}}_{\mathrm{pl}}\underline{\mathbf{n}}_{\mathrm{pl}}^{\mathsf{H}}\right] = \mathbf{R}_{\underline{\mathbf{h}}} + \frac{N_0}{|s|^2\mathcal{E}_{\mathrm{pl}}}\mathbf{I}_{\underline{\mathbf{n}}} \tag{A.1}$$

where $\underline{\mathbf{h}}_{\mathrm{pl}} = \mathsf{vec}(\mathcal{H})$ and $\underline{\mathbf{n}}_{\mathrm{pl}} = \mathsf{vec}(\mathcal{N})$. Also the covariance matrix between the pilot vector, $\underline{\tilde{\mathbf{y}}}$, and the vector of channel gains at the point we want to predict, $\mathbf{h}(k;l) = [h_1(k;l), \cdots, h_{n_{\mathrm{R}}}(k;l)]^{\mathsf{T}}$, can be written as

$$\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} = \mathsf{E}\left[\underline{\tilde{\mathbf{y}}}\mathbf{h}^{\mathsf{H}}(k;l)\right] = \mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\mathbf{h}^{\mathsf{H}}(k;l)\right]. \tag{A.2}$$

Note that the matrix $\mathbf{R}_{\underline{\mathbf{y}}}$ is block-Hermitian.

Since the channel is complex Gaussian, the optimal predicted channel is a linear combination of the observations. As a result the predicted channel can be found as

$$\mathbf{h}_p = \mathbf{G}^{\mathsf{H}}\underline{\tilde{\mathbf{y}}} \tag{A.3}$$

where $\mathbf{G}$ is the optimal predictor chosen such that the MSE is minimized. Defining the prediction error as $\boldsymbol{\epsilon}_p(k;l) = \mathbf{h}(k;l) - \mathbf{h}_p(k;l)$ the MSE variance matrix can be found as

$$\begin{aligned}
\mathbf{R}_\epsilon &= \mathsf{E}\left[\boldsymbol{\epsilon}_p(k;l)\boldsymbol{\epsilon}_p(k;l)^{\mathsf{H}}\right] \\
&= \mathsf{E}\left[\left(\mathbf{h} - \mathbf{G}^{\mathsf{H}}\underline{\tilde{\mathbf{y}}}\right)\left(\mathbf{h} - \mathbf{G}^{\mathsf{H}}\underline{\tilde{\mathbf{y}}}\right)^{\mathsf{H}}\right] \\
&= \mathsf{E}\left[\mathbf{h}\mathbf{h}^{\mathsf{H}}\right] - \mathsf{E}\left[\mathbf{h}\underline{\tilde{\mathbf{y}}}^{\mathsf{H}}\mathbf{G}\right] - \mathsf{E}\left[\mathbf{G}^{\mathsf{H}}\underline{\tilde{\mathbf{y}}}\mathbf{h}^{\mathsf{H}}\right] + \mathsf{E}\left[\mathbf{G}^{\mathsf{H}}\underline{\tilde{\mathbf{y}}}\underline{\tilde{\mathbf{y}}}^{\mathsf{H}}\mathbf{G}\right] \\
&= \mathbf{R}_{\mathbf{h}} - \mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{\mathsf{H}}\mathbf{G} - \mathbf{G}^{\mathsf{H}}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} + \mathbf{G}^{\mathsf{H}}\mathbf{R}_{\underline{\mathbf{y}}}\mathbf{G}.
\end{aligned} \tag{A.4}$$

The total MSE is the sum of the MSE on each branch and is given by the trace of $\mathbf{R}_\epsilon$. As a result, the total MSE denoted by σ_p^2 is

$$\sigma_p^2 = \mathsf{tr}\{\mathbf{R}_{\mathbf{h}}\} - \mathsf{tr}\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{\mathsf{H}}\mathbf{G}\right\} - \mathsf{tr}\left\{\mathbf{G}^{\mathsf{H}}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}\right\} + \mathsf{tr}\left\{\mathbf{G}^{\mathsf{H}}\mathbf{R}_{\underline{\mathbf{y}}}\mathbf{G}\right\}. \tag{A.5}$$

To find the optimal filter coefficients in the MMSE sense we derive the conjugate gradient where the following identities are used [Moon and Stirling, 2000, Appendix E.8]:

$$\frac{\partial \mathsf{tr}\{\mathbf{AXB}\}}{\partial \mathbf{X}^*} = \mathbf{0}, \quad \frac{\partial \mathsf{tr}\{\mathbf{AX}^{\mathsf{H}}\mathbf{B}\}}{\partial \mathbf{X}^*} = \mathbf{BA}, \quad \text{and} \quad \frac{\partial \mathsf{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{AXB}\}}{\partial \mathbf{X}^*} = \mathbf{AXB}. \tag{A.6}$$

Recognizing that (A.5) can be written as

$$\sigma_p^2 = \mathsf{tr}\{\mathbf{R}_{\mathbf{h}}\} - \mathsf{tr}\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{\mathsf{H}}\mathbf{GI}\right\} - \mathsf{tr}\left\{\mathbf{IG}^{\mathsf{H}}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}\right\} + \mathsf{tr}\left\{\mathbf{G}^{\mathsf{H}}\mathbf{R}_{\underline{\mathbf{y}}}\mathbf{GI}\right\} \tag{A.7}$$

where $\mathbf{I}$ is the square identity matrix with proper dimension.

Applying the identities (A.6) to the derivative of (A.7), equating the solution to the zero matrix, and solving for $\mathbf{G}$, gives the predictor

$$\mathbf{G} = \mathbf{R}_{\underline{\mathbf{y}}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} = \left(\mathbf{R}_{\underline{\mathbf{h}}} + \frac{N_0}{|s|^2\mathcal{E}_{\mathrm{pl}}}\mathbf{I}_{\underline{\mathbf{n}}}\right)^{-1}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}. \tag{A.8}$$

Setting (A.8) back to (A.5) gives the MMSE

$$\sigma_p^2 = \mathsf{tr}\{\mathbf{R}_\epsilon\} = n_{\mathrm{R}} - \mathsf{tr}\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{\mathsf{H}}\mathbf{G}\right\}, \tag{A.9}$$

when assuming that the subchannel is zero mean and has unity variance. Also we have assumed that the noise is white both in space and time. Replacing the actual values of $\mathcal{E}_{\mathrm{pl}}$ and N_0, (A.8) is expressed as (5.18):

$$\mathbf{G} = \mathbf{R}_{\underline{\mathbf{y}}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} = \left(\mathbf{R}_{\underline{\mathbf{h}}} + \frac{1}{(1-\alpha)\bar{\gamma}_b L}\mathbf{I}_{\underline{\mathbf{n}}}\right)^{-1}\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}. \tag{A.10}$$

A.2 Derivation of the Predictor Used in Other Papers

If the branches are mutually independent, we can use the same "reduced-predictor" on each branch and predict the branches separately. As will be shown here the reduced-predictor can easily be obtained from (A.8).

Recognize that

$$\mathbf{R}_{\underline{\mathbf{h}}} = \mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\underline{\mathbf{h}}_{\mathrm{pl}}^{\mathsf{H}}\right] = \begin{bmatrix} \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}\mathbf{h}_{\mathrm{pl};1}^{\mathsf{H}}\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}^{\mathsf{H}}\right] \\ \vdots & \ddots & \vdots \\ \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}\mathbf{h}_{\mathrm{pl};1}^{\mathsf{H}}\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}^{\mathsf{H}}\right] \end{bmatrix} \tag{A.11}$$

and

$$\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}} = \mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\mathbf{h}^{\mathsf{H}}(k;l)\right] = \begin{bmatrix} \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}h_1^*(k;l)\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}h_{n_{\mathrm{R}}}^*(k;l)\right] \\ \vdots & \ddots & \vdots \\ \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}h_1^*(k;l)\right] & \cdots & \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};n_{\mathrm{R}}}h_{n_{\mathrm{R}}}^*(k;l)\right] \end{bmatrix}. \tag{A.12}$$

When the branches are uncorrelated, all the off-diagonal blocks in both (A.11) and (A.12) are zero matrices and vectors, respectively. Moreover, the remaining sub-matrices in (A.11) are identical. Likewise for the sub-vectors in (A.12). Inspecting (A.8), we see that the predictor becomes the

matrix

$$\begin{pmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_{K_p} \end{bmatrix} & & & \\ & \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_{K_p} \end{bmatrix} & & \\ & & \ddots & \\ & & \ddots & \\ & & \ddots & \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_{K_p} \end{bmatrix} \end{pmatrix} \tag{A.13}$$

where all the vectors are the identical, and all the other elements are zeros. This implies that one can analyze one of these blocks (corresponding to one branch) and the same result will be directly applicable to other blocks (branches).

Without losing generality let us consider the first branch. Let $\mathbf{R} = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}\mathbf{h}_{\mathrm{pl};1}^{\mathsf{H}}\right]$ and let $\mathbf{r} = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}h_1^*(k;l)\right]$ then the first vector of $\mathbf{G}$ in (A.13) corresponds to the predictor coefficents and is given in the form of a vector $\boldsymbol{\omega} = [\omega_1, \cdots, \omega_{K_p}]^{\mathsf{T}}$ (dimension K_p) padded with $(n_{\mathrm{R}} - 1)K_p$ zeros where the vector $\boldsymbol{\omega}$ is

$$\boldsymbol{\omega} = \left(\mathbf{R} + \frac{N_0}{|s|^2\mathcal{E}_{\mathrm{pl}}}\mathbf{I}\right)^{-1}\mathbf{r} = \left(\mathbf{R} + \frac{1}{(1-\alpha)\bar{\gamma}_b L}\mathbf{I}\right)^{-1}\mathbf{r}. \tag{A.14}$$

Since the other elements are only zeros, the predictor $\boldsymbol{\omega}$ is used on every branch as in the cases where there is no spatial correlation between the branches (Chapter 3, Chapter 4, first part of Chapter 5, and Chapter 6).

Comparing to the MAP-optimal filter in Øien *et al.* [2004] we see that our filter is slightly different since the factor in front of the identity matrix involves the frame size and the variable α which determines how power should be allocated between pilot and data symbols. This is obvious since the same pilot and data power are assumed in Øien *et al.* [2004], and in our papers these two parameters will be optimized to achieve maximum throughput.

MMSE of One Branch

It can be deduced from (A.9) that the MMSE on one branch is given by

$$\sigma^2_{p;\mu} = 1 - \mathbf{r}^{\mathsf{H}}\boldsymbol{\omega} = 1 - \mathbf{r}^{\mathsf{H}}\left(\mathbf{R} + \frac{1}{(1-\alpha)L\bar{\gamma}_b}\mathbf{I}\right)^{-1}\mathbf{r}. \tag{A.15}$$

Moreover, the covariance matrix is Hermitian symmetric and it can be diagonalized as $\mathbf{R} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^{\mathsf{H}}$ with $\mathbf{E}$ (unitary matrix) and $\boldsymbol{\Lambda}$ (diagonal matrix) containing respectively eigenvectors $\{\mathbf{u}_\kappa\}_{\kappa=1}^{K_p}$ and eigenvalues $\{\lambda_\kappa\}_{\kappa=1}^{K_p}$ of $\mathbf{R}$, where K_p is the order of the filter. Hence

$$\begin{aligned}\sigma^2_{p;\mu} &= 1 - \mathbf{r}^{\mathsf{H}}\mathbf{E}\left(\boldsymbol{\Lambda} + \frac{1}{(1-\alpha)L\bar{\gamma}_b}\mathbf{I}\right)^{-1}\mathbf{E}^{\mathsf{H}}\mathbf{r} \\ &= 1 - \sum_{\kappa=1}^{K_p}\frac{|\mathbf{u}_\kappa^{\mathsf{H}}\mathbf{r}|^2(1-\alpha)L\bar{\gamma}_b}{(1-\alpha)L\bar{\gamma}_b\lambda_\kappa + 1}.\end{aligned} \tag{A.16}$$

An Alternative Expression for the MMSE

The MMSE of the prediction error can also be written as

$$\sigma^2_{p;\mu} = \sigma^2_{h_\mu} - \sigma^2_{h_{p;\mu}}, \tag{A.17}$$

where $\sigma^2_{h_{p;\mu}} = \boldsymbol{\omega}^{\mathsf{H}}\boldsymbol{\mathcal{R}}\boldsymbol{\omega}$ is variance of the prediction error [Cavers, 1991], and $\boldsymbol{\mathcal{R}} = \mathsf{E}\left[\mathbf{h}_{\mathrm{pl};1}\mathbf{h}^{\mathsf{H}}_{\mathrm{pl};1}\right] + \frac{1}{|s|^2\mathcal{E}_{\mathrm{pl}}}\mathsf{E}\left[\mathbf{n}_{\mathrm{pl};1}\mathbf{n}^{\mathsf{H}}_{\mathrm{pl};1}\right]$. Again, by assuming that the channel has unity variance we get

$$\begin{aligned}\sigma^2_{p;\mu} &= 1 - \boldsymbol{\omega}^{\mathsf{H}}\left(\mathbf{R} + \frac{1}{(1-\alpha)L\bar{\gamma}_b}\mathbf{I}\right)\boldsymbol{\omega} \\ &= 1 - \boldsymbol{\omega}^{\mathsf{H}}\mathbf{R}\boldsymbol{\omega} - \frac{\|\boldsymbol{\omega}\|^2}{(1-\alpha)L\bar{\gamma}_b}.\end{aligned} \tag{A.18}$$

Appendix B

Derivation of BER(M_n) in Chapter 3

B.1 Useful Integration Rules and Identities

Definition 2 (The normalized incomplete gamma function)
According to Temme [1996], the incomplete gamma function and the normalized incomplete gamma function is respectively defined as

$$\Gamma(\eta,\zeta) = \int_{\zeta}^{\infty} t^{\eta-1}e^{-t}dt \tag{B.1}$$

$$\overline{\Gamma}(\eta,\zeta) = \Gamma(\eta,\zeta)/\Gamma(\eta), \tag{B.2}$$

where $\Gamma(\eta)$ is the gamma function (also known as Euler's integral of the second kind) [Gradshteyn and Ryzhik, 2000, Eq. 8.310-1].

Definition 3 (Integration rule involving the generalized Marcum Q-function)
Based on the non-central chi-square distribution defined in [Temme, 1996, Sec. 11.4] and the normalized incomplete gamma function defined above, Holm provided the following formulas [Holm, 2002, Lemma 3]:

$$\int_0^y u^{\frac{H-1}{2}} e^{-u\alpha} I_{H-1}(\sqrt{u}2\beta)du = \frac{\beta^{H-1}}{\alpha^H}\exp\left(\frac{\beta^2}{\alpha}\right)\left(1 - Q_H\left(\frac{\beta^2}{\alpha}, y\alpha\right)\right), \tag{B.3}$$

where $Q_H(\cdot,\cdot)$ is the generalized Marcum Q-function given by [Temme, 1996, Eq. (11.63)]. Analogous to [Temme, 1996, Eq. (11.61)] the following serie expansion was also given

$$1 - Q_\mu(x,y) = e^{-x}\sum_{m=0}^{\infty}\frac{x^m}{m!}(1 - \overline{\Gamma}(\mu+m,y)). \tag{B.4}$$

Definition 4 (Integration rule involving Bessel functions)
The following integration rule is given by [Gradshteyn and Ryzhik, 2000, Eq. (6.633-4), p. 699], but reproduce here for the sake of completeness

$$\int_0^\infty u e^{-\alpha u^2} I_\nu(\beta u) J_\nu(\lambda u) du = \frac{1}{2\alpha} \exp\left(\frac{\beta^2 - \lambda^2}{4\alpha}\right) J_\nu\left(\frac{\beta\lambda}{2\alpha}\right), \quad \Re\{\alpha\} > 0, \ \Re\{\nu\} > -1. \tag{B.5}$$

Definition 5 (Integration rule involving incomplete gamma function)
Making use of [Gradshteyn and Ryzhik, 2000, Eq. (3.381-3)] in combination with the definition of the normalized incomplete gamma function given by (B.2) the following integration rule is obtained:

$$\int_x^y t^{\eta-1} e^{-\mu t} dt = \frac{\Gamma(\eta)}{\mu^\eta} \left[\overline{\Gamma}(\eta, x\mu) - \overline{\Gamma}(\eta, y\mu)\right]. \tag{B.6}$$

B.2 Derivations

In order to find the switching thresholds we need to solve $\text{BER}(M_n|\hat{\gamma}) = \text{BER}_0$. Thus, we need first to calculate the following integral:

$$\text{BER}(M_n|h_p) = \int_0^\infty \text{BER}\left(M_n\middle||h_e|\right) f_{h_e|h_p}\left(|h_e|\middle|h_p\right) d|h_e| \tag{B.7}$$

where the following BER is used

$$\text{BER}\left(M_n\middle||h_e|\right) = \begin{cases} a_n \exp\left(-\dfrac{b_n \mathcal{E}_d |h_e|^2}{M_n(N_0 + g\mathcal{E}_d\sigma_e^2(l))}\right) & \text{when } |h_e| \geq |h_e|_{n,T} \\ \dfrac{1}{2} & \text{when } |h_e| < |h_e|_{n,T} \end{cases} \tag{B.8}$$

for $|h_e|_{n,T} = \sqrt{\ln(2a_n) M_n(N_0 + g\mathcal{E}_d\sigma_e^2)/(b_n\mathcal{E}_d)}$. Inserting (B.8) in (B.7) the integral can be splitted into three integrals. We now define these integrals as

$$\mathcal{T}(0,\infty) = \int_0^\infty a_n \exp\left(-\frac{b_n\mathcal{E}_d|h_e|^2}{M_n(N_0 + g\mathcal{E}_d\sigma_e^2)}\right) f_{h_e|h_p}\left(|h_e|\middle|h_p\right) d|h_e| \tag{B.9}$$

$$\mathcal{T}(0,|h_e|_{n,T}) = \int_0^{|h_e|_{n,T}} a_n \exp\left(-\frac{b_n\mathcal{E}_d|h_e|^2}{M_n(N_0 + g\mathcal{E}_d\sigma_e^2)}\right) f_{h_e|h_p}\left(|h_e|\middle|h_p\right) d|h_e| \tag{B.10}$$

$$\mathcal{F}(0,|he|_{n,T}) = \int_0^{|h_e|_{n,T}} \frac{1}{2} f_{h_e|h_p}\left(|h_e|\middle|h_p\right) d|h_e|. \tag{B.11}$$

Then

$$\text{BER}(M_n|h_p) = \mathcal{T}(0,\infty) - \mathcal{T}(0,|he|_{n,T}) + \mathcal{F}(0,|he|_{n,T}). \tag{B.12}$$

In the following we will proceed to solve the integrals (B.9)–(B.11).

B.2.1 Calculation of $\mathcal{T}(0,\infty)$

Recall now the fact that $|h_e|$ given h_p is Rician distributed and the PDF is given by (1.2). Applying this,

$$\mathcal{T}(0,\infty) = \frac{2a_n}{\sigma^2_{h_e|h_p}} \exp\left(-\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right) \times \int_0^\infty |h_e| \exp\left(-|h_e|^2\left(\frac{b_n\mathcal{E}_d}{M_n(N_0+g\mathcal{E}_d\sigma_e^2)} + \frac{1}{\sigma^2_{h_e|h_p}}\right)\right) I_0\left(\frac{|h_e|2|h_p|}{\sigma^2_{h_e|h_p}}\right) d|h_e|. \tag{B.13}$$

Using the definitions of A_n and d_n in (3.20) and recognizing that $J_0(0) = 1$, the integration rule in (B.5) is utilized to obtain

$$\mathcal{T}(0,\infty) = a_n d_n \exp\left(-|h_p|^2 A_n d_n \mathcal{E}_d\right). \tag{B.14}$$

B.2.2 Calculation of $\mathcal{T}(0,|h_e|_{n,T})$

The integral is expressible as

$$\mathcal{T}(0,|h_e|_{n,T}) = \frac{2a_n}{\sigma^2_{h_e|h_p}} \exp\left(-\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right) \times \int_0^{|h_e|_{n,T}} |h_e| \exp\left(-|h_e|^2\left(\frac{b_n\mathcal{E}_d}{M_n(N_0+g\mathcal{E}_d\sigma_e^2)} + \frac{1}{\sigma^2_{h_e|h_p}}\right)\right) I_0\left(\frac{|h_e|2|h_p|}{\sigma^2_{h_e|h_p}}\right) d|h_e|. \tag{B.15}$$

We let $u = |h_e|^2$, then $du = 2|h_e|d|h_e|$. The above equation is now

$$\begin{aligned}\mathcal{T}(0,|h_e|_{n,T}) =& \frac{a_n}{\sigma^2_{h_e|h_p}} \exp\left(-\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right) \\ &\times \int_0^{|h_e|^2_{n,T}} \exp\left(-u\left(A_n\mathcal{E}_d + \frac{1}{\sigma^2_{h_e|h_p}}\right)\right) I_0\left(\frac{2|h_p|}{\sigma^2_{h_e|h_p}}\sqrt{u}\right) du \\ =& a_n d_n \exp\left(-|h_p|^2 A_n d_n \mathcal{E}_d\right)\left(1 - Q_1\left(\frac{|h_p|^2 d_n}{\sigma^2_{h_e|h_p}}, \frac{|h_e|^2_{n,T}}{d_n\sigma^2_{h_e|h_p}}\right)\right).\end{aligned} \tag{B.16}$$

Here the last equality is obtained by invoking (B.3). Furthermore, using (B.4), the above expression is

$$\mathcal{T}(0,|h_e|_{n,T}) = a_n d_n \exp\left(-\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right) \times \sum_{m=0}^{\infty} \frac{1}{m!}\left(\frac{|h_p|^2 d_n}{\sigma^2_{h_e|h_p}}\right)^m \left(1-\overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{d_n\sigma^2_{h_e|h_p}}\right)\right). \quad \text{(B.17)}$$

Clearly, it can be seen that (B.14) is a special case of (B.16) where $|h_e|^2_{n,T} \to \infty$. That case, $\overline{\Gamma}(\mu+m,\infty) = 0$ (cf. Definition 2 on page 125), and it is well known that the summation of the right hand side of (B.4) is equal to e^x. Thus, $Q_1(x,\infty) = 0$ and the two equations (B.14) and (B.16) coincides.

B.2.3 Calculation of $\mathcal{F}(0,|h_e|_{n,T})$

This integral is half of the CDF of the conditional Rician PDF evaluated at $|h_e|_{n,T}$. That is

$$\begin{aligned}
\mathcal{F}(0,|h_e|_{n,T}) &= \frac{1}{\sigma^2_{h_e|h_p}} \exp\left(-\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right) \times \int_0^{|h_e|_{n,T}} |h_e| \exp\left(-\frac{|h_e|^2}{\sigma^2_{h_e|h_p}}\right) I_0\left(\frac{|h_e|2|h_p|}{\sigma^2_{h_e|h_p}}\right) d|h_e| \\
&\overset{(i)}{=} \frac{1}{\sigma^2_{h_e|h_p}} \exp\left(-\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right) \times \frac{1}{2}\int_0^{|h_e|^2_{n,T}} \exp\left(-\frac{u}{\sigma^2_{h_e|h_p}}\right) I_0\left(\frac{2|h_p|}{\sigma^2_{h_e|h_p}}\sqrt{u}\right) du \\
&\overset{(ii)}{=} \frac{1}{2}\left(1-Q_1\left(\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}, \frac{|h_e|^2_{n,T}}{\sigma^2_{h_e|h_p}}\right)\right) \\
&\overset{(iii)}{=} \frac{1}{2}\exp\left(-\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right) \times \sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{|h_p|^2}{\sigma^2_{h_e|h_p}}\right)^m \left(1-\overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{\sigma^2_{h_e|h_p}}\right)\right). \quad \text{(B.18)}
\end{aligned}$$

Here the change of variable $u = |h_e|^2$ has been made in the equality marked with (i). The equality marked with (ii) uses (B.3) and the one marked with (iii) exploits (B.4).

B.3 BER$(M_n|\hat{\gamma})$

Combining (1.5) with (3.17) and solving it for $|h_p|^2$ gives $|h_p|^2 = \hat{\gamma}\mathcal{E}/(\bar{\gamma}\bar{\mathcal{E}}_d)$. Inserting this into Eqs. (B.14), (B.17), and (B.18) and combining them similar to (B.12) would result in (3.19). Thus the switching thresholds are found by solving $\mathrm{BER}(M_n|\hat{\gamma}) = \mathrm{BER}_0$ with respect to $\hat{\gamma}$ for different M_n.

B.4 BER(M_n)

Once the switching thresholds are found, the average BER for the M_n constellation is calculated as

$$\mathrm{BER}(M_n) = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathrm{BER}(M_n|\hat{\gamma}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \tag{B.19}$$

$$= \mathcal{B}1(M_n) - \mathcal{B}21(M_n) + \mathcal{B}22(M_n) \tag{B.20}$$

where

$$\mathcal{B}1(M_n) = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathcal{T}(0,\infty) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \tag{B.21}$$

$$\mathcal{B}21(M_n) = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathcal{T}(0,|h_e|_{n,T}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \tag{B.22}$$

$$\mathcal{B}22(M_n) = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \mathcal{F}(0,|h_e|_{n,T}) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma}. \tag{B.23}$$

We will now solve these integrals.

B.4.1 Calculation of $\mathcal{B}1(M_n)$

Knowing that the predicted CSNR is exponentially distributed—i.e., a special case of gamma distribution given as $\mathcal{G}(1, r\bar{\gamma})$—and inserting the definition of $f_{\hat{\gamma}}(\hat{\gamma})$ gives

$$\begin{aligned}\mathcal{B}1(M_n) =& \frac{a_n d_n}{r\bar{\gamma}} \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \exp\left(-\hat{\gamma}\left(\frac{A_n d_n \mathcal{E}\mathcal{E}_d}{\bar{\gamma}\bar{\mathcal{E}}_d} + \frac{1}{r\bar{\gamma}}\right)\right) d\hat{\gamma} \\ =& \frac{a_n d_n \bar{\mathcal{E}}_d}{r A_n d_n \mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d} \\ & \times \left[\exp\left(-\hat{\gamma}_n \frac{r A_n d_n \mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}\bar{\mathcal{E}}_d}\right) - \exp\left(-\hat{\gamma}_{n+1} \frac{r A_n d_n \mathcal{E}\mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}\bar{\mathcal{E}}_d}\right)\right].\end{aligned} \tag{B.24}$$

B.4.2 Calculation of $\mathcal{B}21(M_n)$

Similarly, $\mathcal{B}21(M_n)$ can be written as

$$\mathcal{B}21(M_n) = \frac{a_n d_n}{r\bar{\gamma}} \sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{d_n \mathcal{E}}{\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right)^m \left(1 - \overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{d_n \sigma^2_{h_e|h_p}} \right) \right) \times \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \hat{\gamma}^m \exp\left(-\hat{\gamma} \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) d\hat{\gamma}. \tag{B.25}$$

Invoking (B.6) the following result is obtained:

$$\mathcal{B}21(M_n) = \frac{a_n \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\mathcal{E}} \sum_{m=0}^{\infty} \left(\frac{r d_n \mathcal{E}}{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right)^{m+1} \left(1 - \overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{d_n \sigma^2_{h_e|h_p}} \right) \right) \times \left[\overline{\Gamma}\left(1+m, \hat{\gamma}_n \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) - \overline{\Gamma}\left(1+m, \hat{\gamma}_{n+1} \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) \right]. \tag{B.26}$$

B.4.3 Calculation of $\mathcal{B}22(M_n)$

Needless to say, the integral $\mathcal{B}22(M_n)$ is expressible as

$$\mathcal{B}22(M_n) = \frac{1}{2r\bar{\gamma}} \sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{\mathcal{E}}{\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right)^m \left(1 - \overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{\sigma^2_{h_e|h_p}} \right) \right) \times \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \hat{\gamma}^m \exp\left(-\hat{\gamma} \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) d\hat{\gamma}. \tag{B.27}$$

Recognize that this integral can be solved as in previous subsection. That is:

$$\mathcal{B}22(M_n) = \frac{\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{2r\mathcal{E}} \sum_{m=0}^{\infty} \left(\frac{r\mathcal{E}}{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right)^{m+1} \left(1 - \overline{\Gamma}\left(1+m, \frac{|h_e|^2_{n,T}}{\sigma^2_{h_e|h_p}} \right) \right) \times \left[\overline{\Gamma}\left(1+m, \hat{\gamma}_n \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) - \overline{\Gamma}\left(1+m, \hat{\gamma}_{n+1} \frac{r\mathcal{E} + \bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}}{r\bar{\gamma}\bar{\mathcal{E}}_d \sigma^2_{h_e|h_p}} \right) \right]. \tag{B.28}$$

Now the Eqs. (B.24), (B.26), and (B.28) are combined as (B.19) to yield BER(M_n).

Appendix C

An Alternative Proof of the Expression for $\mathsf{Var}\left(h_{e;\mu}|h_{p;\mu}\right)$

In this appendix, we will give an alternative proof of the expression for the quantity $\mathsf{Var}\left(h_{e;\mu}|h_{p;\mu}\right)$. To do so, we need the following property of the expectation operator.

Theorem 1 (Property of the expectation)
The following property

$$\mathsf{E}[XY] = \mathsf{E}[Y \cdot \mathsf{E}[X|Y]] \tag{C.1}$$

is true for all jointly distributed RVs X and Y.

Proof: Let $Y = y$, then we have

$$\begin{aligned}
\mathsf{E}[Y \cdot \mathsf{E}[X|Y]] &= \int_{-\infty}^{\infty} y\mathsf{E}[X|Y=y]\, f_Y(y)dy \\
&= \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx \right) f_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \underbrace{f_{X|Y}(x|y)f_Y(y)}_{f_{XY}(x,y)} dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)dxdy \\
&= \mathsf{E}[XY]\,.
\end{aligned} \tag{C.2}$$

□

Omitting the time indices, we express the estimated channel as

$$h_{e;\mu} = h_{p;\mu} + \epsilon_{p;\mu} - \epsilon_{e;\mu},$$

where all the RVs are Gaussian distributed with zero mean.

Let $h_{e;\mu}$ and $h_{p;\mu}$ be jointly Gaussian distributed with the correlation coefficient ζ. Then, invoking (C.1), the covariance between the two is

$$\mathsf{Cov}\left(h_{e;\mu}, h_{p;\mu}\right) = \mathsf{E}\left[h_{e;\mu}h_{p;\mu}\right] = \mathsf{E}\left[h_{p;\mu} \cdot \mathsf{E}\left[h_{e;\mu}|h_{p;\mu}\right]\right].$$

As a matter of fact, using mean square estimation of Gaussian RVs, it is easy to show that $\mathsf{E}\left[h_{e;\mu}|h_{p;\mu}\right] = (1-\varrho)h_{p;\mu}$ where ϱ is the normalized correlation between $\epsilon_{e;\mu}$ and $h_{p;\mu}$ (see e.g. [Papoulis and Pillai, 2002, Sec. 17-3]). As a result, the above equation is

$$\mathsf{Cov}\left(h_{e;\mu}, h_{p;\mu}\right) = \mathsf{E}\left[(1-\varrho)\, h_{p;\mu}^2\right] = (1-\varrho)\, \sigma_{h_{p;\mu}}^2,$$

and the correlation coefficient is

$$\zeta = \frac{\mathsf{Cov}\left(h_{e;\mu}, h_{p;\mu}\right)}{\sqrt{\mathsf{Var}\left(h_{e;\mu}\right)\mathsf{Var}\left(h_{p;\mu}\right)}} = \frac{\sigma_{h_{p;\mu}}}{\sigma_{h_{e;\mu}}}\,(1-\varrho)\,.$$

Moreover, in general, it is well known that if $(h_{e;\mu}, h_{p;\mu})$ is jointly Gaussian denoted by $(h_{e;\mu}, h_{p;\mu}) \sim \mathcal{N}(m_{h_{e;\mu}}, m_{h_{p;\mu}}, \sigma_{h_{e;\mu}}^2, \sigma_{h_{p;\mu}}^2, \zeta)$, then $h_{e;\mu}|h_{p;\mu} \sim \mathcal{N}(m_{h_{e;\mu}} + \zeta\frac{\sigma_{h_{e;\mu}}}{\sigma_{h_{p;\mu}}}(h_{p;\mu} - m_{h_{p;\mu}}), \sigma_{h_{e;\mu}}^2(1-\zeta^2))$. Therefore, in our case

$$\sigma_{h_{e;\mu}|h_{p;\mu}}^2 = \mathsf{Var}\left(h_{e;\mu}|h_{p;\mu}\right) = \sigma_{h_{e;\mu}}^2(1-\zeta^2) = \sigma_{h_{e;\mu}}^2 - \sigma_{h_{p;\mu}}^2 \cdot (1-\varrho)^2\,.$$

Upon assuming $\varrho = 0$ it can be expressed as

$$\sigma_{h_{e;\mu}|h_{p;\mu}}^2 = \sigma_{h_{e;\mu}}^2 - \sigma_{h_{p;\mu}}^2 = \sigma_{p;\mu}^2 - \sigma_{e;\mu}^2$$

and that completes the proof.

Appendix D

Calculation of the Correlation Coefficient ρ in Equation (5.24)

With a simple transformation we obtained $\beta^2 \sim \mathcal{G}\left(m_d, n_R/m_d\right)$ and $\hat{\beta}^2 \sim \mathcal{G}\left(m_d, \mathrm{tr}\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{\mathsf{H}}\mathbf{G}\right\}/m_d\right)$. Thus the correlation coefficient in (5.24) is

$$\begin{aligned}\rho &= \frac{\mathrm{Cov}(\hat{\gamma},\gamma_d)}{\sqrt{\mathrm{Var}(\hat{\gamma})\,\mathrm{Var}(\gamma_d)}} = \frac{\mathrm{Cov}\left(\hat{\beta}^2,\beta^2\right)}{\sqrt{\mathrm{Var}\left(\hat{\beta}^2\right)\mathrm{Var}\left(\beta^2\right)}} \\ &= \frac{\mathsf{E}\left[\hat{\beta}^2\beta^2\right] - n_R\mathrm{tr}\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{\mathsf{H}}\mathbf{G}\right\}}{n_R\mathrm{tr}\left\{\mathbf{R}_{\underline{\mathbf{yh}}}^{\mathsf{H}}\mathbf{G}\right\}} m_d \end{aligned} \tag{D.1}$$

where

$$\mathsf{E}\left[\hat{\beta}^2\beta^2\right] = \mathsf{E}\left[\sum_{\mu=1}^{n_R}\hat{\beta}_\mu^2\sum_{\nu=1}^{n_R}\beta_\nu^2\right] = \sum_{\mu=1}^{n_R}\sum_{\nu=1}^{n_R}\mathsf{E}\left[\hat{\beta}_\mu^2\beta_\nu^2\right]. \tag{D.2}$$

Using (A.3) we can write

$$\begin{aligned}\mathsf{E}\left[\hat{\beta}_\mu^2\beta_\nu^2\right] &= \mathsf{E}\left[|\underline{\mathbf{g}}_\mu^{\mathsf{H}}\underline{\tilde{\mathbf{y}}}|^2\cdot|h_\nu(k;l)|^2\right] \\ &= \underline{\mathbf{g}}_\mu^{\mathsf{H}}\mathsf{E}\left[\left(\underline{\mathbf{h}}_{\mathrm{pl}}+\frac{1}{s\sqrt{\mathcal{E}_{\mathrm{pl}}}}\underline{\mathbf{n}}_{\mathrm{pl}}\right)\left(\underline{\mathbf{h}}_{\mathrm{pl}}+\frac{1}{s\sqrt{\mathcal{E}_{\mathrm{pl}}}}\underline{\mathbf{n}}_{\mathrm{pl}}\right)^{\mathsf{H}}|h_\nu(k;l)|^2\right]\underline{\mathbf{g}}_\mu \\ &= \underline{\mathbf{g}}_\mu^{\mathsf{H}}\mathsf{E}\left[\underline{\mathbf{h}}_{\mathrm{pl}}\underline{\mathbf{h}}_{\mathrm{pl}}^{\mathsf{H}}\cdot|h_\nu(k;l)|^2\right]\underline{\mathbf{g}}_\mu + \frac{\|\underline{\mathbf{g}}_\mu\|^2}{(1-\alpha)\bar{\gamma}_b L}\end{aligned} \tag{D.3}$$

where $\underline{\mathbf{g}}_\mu$ is the μth column vector of the predictor matrix $\mathbf{G}$.

We now drop the subscript "pl" for notational brevity and use the notation $h_{b\Re}$ (or $h_{b\Im}$) to denote the real part (or imaginary part) of the channel gain h on the bth branch. Likewise, $\underline{\mathbf{h}}_{\Re}$ and $\underline{\mathbf{h}}_{\Im}$ is used to denote the real and imaginary part of vector $\underline{\mathbf{h}}$, respectively. The expectation term in the above equation is

$$\begin{aligned}
\mathsf{E}\left[\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathsf{H}} \cdot |h_\nu(k;l)|^2\right] =& \mathsf{E}\left[(\underline{\mathbf{h}}_{\Re} + \jmath\underline{\mathbf{h}}_{\Im})\left(\underline{\mathbf{h}}_{\Re}^{\mathsf{T}} - \jmath\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right)\left(h_{\nu\Re}^2(k;l) + h_{\nu\Im}^2(k;l)\right)\right] \\
=& \mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}} h_{\nu\Re} h_{\nu\Re}\right] + \mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}} h_{\nu\Im} h_{\nu\Im}\right] \\
&+ \mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}} h_{\nu\Re} h_{\nu\Re}\right] + \mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}} h_{\nu\Im} h_{\nu\Im}\right] \\
&- \jmath\mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}} h_{\nu\Re} h_{\nu\Re}\right] - \jmath\mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}} h_{\nu\Im} h_{\nu\Im}\right] \\
&+ \jmath\mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}} h_{\nu\Re} h_{\nu\Re}\right] + \jmath\mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}} h_{\nu\Im} h_{\nu\Im}\right].
\end{aligned} \tag{D.4}$$

Here we have also omitted the time indices for notational brevity. Expanding this equation using Gaussian fourth-order moments [Holm, 2002, Lemma 1] gives

$$\begin{aligned}
\mathsf{E}\left[\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathsf{H}} \cdot |h_\nu(k;l)|^2\right] =& \mathsf{E}[h_{\nu\Re}h_{\nu\Re}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right] + 2\mathsf{E}[h_{\nu\Re}\underline{\mathbf{h}}_{\Re}]\,\mathsf{E}\left[h_{\nu\Re}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right] \\
&+ \mathsf{E}[h_{\nu\Im}h_{\nu\Im}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right] + 2\mathsf{E}[h_{\nu\Im}\underline{\mathbf{h}}_{\Re}]\,\mathsf{E}\left[h_{\nu\Im}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right] \\
&+ \mathsf{E}[h_{\nu\Re}h_{\nu\Re}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] + 2\mathsf{E}[h_{\nu\Re}\underline{\mathbf{h}}_{\Im}]\,\mathsf{E}\left[h_{\nu\Re}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] \\
&+ \mathsf{E}[h_{\nu\Im}h_{\nu\Im}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] + 2\mathsf{E}[h_{\nu\Im}\underline{\mathbf{h}}_{\Im}]\,\mathsf{E}\left[h_{\nu\Im}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] \\
&- \jmath\mathsf{E}[h_{\nu\Re}h_{\nu\Re}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] - \jmath 2\mathsf{E}[h_{\nu\Re}\underline{\mathbf{h}}_{\Re}]\,\mathsf{E}\left[h_{\nu\Re}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] \\
&- \jmath\mathsf{E}[h_{\nu\Im}h_{\nu\Im}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Re}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] - \jmath 2\mathsf{E}[h_{\nu\Im}\underline{\mathbf{h}}_{\Re}]\,\mathsf{E}\left[h_{\nu\Im}\underline{\mathbf{h}}_{\Im}^{\mathsf{T}}\right] \\
&+ \jmath\mathsf{E}[h_{\nu\Re}h_{\nu\Re}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right] + \jmath 2\mathsf{E}[h_{\nu\Re}\underline{\mathbf{h}}_{\Im}]\,\mathsf{E}\left[h_{\nu\Re}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right] \\
&+ \jmath\mathsf{E}[h_{\nu\Im}h_{\nu\Im}]\,\mathsf{E}\left[\underline{\mathbf{h}}_{\Im}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right] + \jmath 2\mathsf{E}[h_{\nu\Im}\underline{\mathbf{h}}_{\Im}]\,\mathsf{E}\left[h_{\nu\Im}\underline{\mathbf{h}}_{\Re}^{\mathsf{T}}\right].
\end{aligned} \tag{D.5}$$

The space-time correlation between any two branches and at any lag

can be written as [Therrien, 1992]

$$\begin{aligned}\mathsf{E}\left[h_\mu(n)h_\nu^*(n+m)\right] =&\mathsf{E}\left[\left\{h_{\mu\Re}(n)+\jmath h_{\mu\Im}(n)\right\}\left\{h_{\nu\Re}(n+m)-\jmath h_{\nu\Im}(n+m)\right\}\right]\\ =&\mathsf{E}\left[h_{\mu\Re}(n)h_{\nu\Re}(n+m)\right]-\jmath\mathsf{E}\left[h_{\mu\Re}(n)h_{\nu\Im}(n+m)\right]\\ &+\jmath\mathsf{E}\left[h_{\mu\Im}(n)h_{\nu\Re}(n+m)\right]+\mathsf{E}\left[h_{\mu\Im}(n)h_{\nu\Im}(n+m)\right]\\ =&2\mathsf{E}\left[h_{\mu\Re}(n)h_{\nu\Re}(n+m)\right]-\jmath 2\mathsf{E}\left[h_{\mu\Re}(n)h_{\nu\Im}(n+m)\right].\end{aligned} \tag{D.6}$$

Applying the separability of the space-time correlation defined in (5.2) the above equation can be written in terms of its real and imaginary parts as [Hasna *et al.*, 2001]

$$\begin{aligned}\mathsf{E}\left[h_\mu(n)h_\nu^*(n+m)\right] =&(c-\jmath d)^{|\mu-\nu|}\rho_{h,t}(m)\\ =&\left(\sqrt{\rho_s/2}-\jmath\sqrt{\rho_s/2}\right)^{|\mu-\nu|}\rho_{h,t}(m).\end{aligned} \tag{D.7}$$

Combining (D.6) with (D.7) we have

$$\mathsf{E}\left[\underline{\mathbf{h}}_\Re\underline{\mathbf{h}}_\Re^\mathsf{T}\right]=\mathsf{E}\left[\underline{\mathbf{h}}_\Im\underline{\mathbf{h}}_\Im^\mathsf{T}\right]=\frac{1}{2}\Re\{\mathbf{R}_{\underline{\mathbf{h}}}\}$$

$$\mathsf{E}\left[\underline{\mathbf{h}}_\Im\underline{\mathbf{h}}_\Re^\mathsf{T}\right]=-\mathsf{E}\left[\underline{\mathbf{h}}_\Re\underline{\mathbf{h}}_\Im^\mathsf{T}\right]=\frac{1}{2}\Im\{\mathbf{R}_{\underline{\mathbf{h}}}\}$$

and

$$\mathsf{E}[h_{\nu\Re}\underline{\mathbf{h}}_\Re]=\mathsf{E}[h_{\nu\Im}\underline{\mathbf{h}}_\Im]=\frac{1}{2}\Re\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}$$

$$\mathsf{E}[h_{\nu\Re}\underline{\mathbf{h}}_\Im]=-\mathsf{E}[h_{\nu\Im}\underline{\mathbf{h}}_\Re]=\frac{1}{2}\Im\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}$$

where $\mathbf{R}_{\underline{\mathbf{h}}}$ is the matrix in (5.19), $\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}$ is the νth column of the correlation matrix $\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}$ in (5.20), and $\Re\{A\}$ (or $\Im\{A\}$) is the real (or imaginary) part of A. We also have $\mathsf{E}[h_{\nu\Re}h_{\nu\Re}]=\mathsf{E}[h_{\nu\Im}h_{\nu\Im}]=1/2$. Thus,

$$\begin{aligned}\mathsf{E}\left[\underline{\mathbf{h}}\underline{\mathbf{h}}^\mathsf{H}\cdot|h_\nu(k;l)|^2\right]=&\mathbf{R}_{\underline{\mathbf{h}}}+\Re\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}\Re\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}^\mathsf{T}+\Im\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}\Im\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}^\mathsf{T}\\ &-\jmath\Re\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}\Im\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}^\mathsf{T}+\jmath\Im\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}\Re\left\{\mathbf{R}_{\underline{\mathbf{y}}\mathbf{h}}^{(\nu)}\right\}^\mathsf{T}.\end{aligned}$$

Then (D.1) can be calculated as

$$\rho = \frac{\begin{aligned} m_d \sum_{\mu=1}^{n_R} \sum_{\nu=1}^{n_R} \Bigg\{ \underline{\mathbf{g}}_\mu^H \Bigg(\mathbf{R}_{\underline{\mathbf{h}}} + \Re\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\} \Re\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\}^T + \Im\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\} \Im\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\}^T \\ -\jmath\Re\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\} \Im\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\}^T + \jmath\Im\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\} \Re\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^{(\nu)}\Big\}^T \Bigg) \underline{\mathbf{g}}_\mu + \frac{\|\underline{\mathbf{g}}_\mu\|^2}{(1-\alpha)\bar{\gamma}_b L} \Bigg\} \end{aligned}}{n_R \mathrm{tr}\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^H \mathbf{G}\Big\}} \\ - m_d n_R \mathrm{tr}\Big\{\mathbf{R}_{\underline{\mathbf{yh}}}^H \mathbf{G}\Big\}}.$$

(D.8)

The correlation coefficient used in the first part of Chapter 5 can be obtained in the same way (not shown here). Instead, we refer to [Holter and Øien, 2006] for a detailed derivation.

Bibliography

Aalo, V. A. (1995, August).
Performance of maximal-ratio diversity systems in a correlated Nakagami-fading environment.
IEEE Transactions on Communications 43(8), 2360–2369.

Abdi, A. and M. Kaveh (2000, August).
A versatile spatio-temporal correlation function for mobile fading channels with non-isotropic scattering.
In *IEEE Workshop on Statistical Signal and Array Processing*, pp. 58–62.

Alamouti, S. M. (1998, October).
A simple transmit diversity technique for wireless communications.
IEEE Journal on Selected Areas in Communications 16(8), 1451–1458.

Alouini, M.-S., A. Abdi, and M. Kaveh (2001, November).
Sum of gamma variates and performance of wireless communication systems over Nakagami-fading channels.
IEEE Transactions on Vehicular Technology 50(6), 1471–1480.

Alouini, M.-S. and A. Goldsmith (1997, May).
Capacity of Nakagami multipath fading channels.
In *Proc. IEEE Vehicular Technology Conference (VTC-Spring)*, Phoenix, AZ.

Alouini, M.-S. and A. J. Goldsmith (1999, July).
Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques.
IEEE Transactions on Vehicular Technology 48(4), 1165–1181.

Alouini, M.-S. and A. J. Goldsmith (2000, May).
Adaptive modulation over Nakagami fading channels.
Kluwer J. Wireless Communications 13, 119–143.

Alouini, M.-S., X. Tang, and A. J. Goldsmith (1999, May).

An adaptive modulation scheme for simultaneous voice and data transmission over fading channels.
IEEE Journal on Selected Areas in Communications 17(5), 837–850.

Beaulieu, N. C. and X. Dong (2003, May).
Level crossing rate and average fade duration of MRC and EGC diversity in ricean fading.
IEEE Transactions on Communications 51(5), 722–726.

Brennan, D. G. (2003, February).
Linear diversity combining techniques.
Proceedings of the IEEE 91(2), 331–356.

Byers, G. J. and F. Takawira (2004, May).
Spatially and temporally correlated MIMO channels: modelling and capacity analysis.
IEEE Transactions on Vehicular Technology 53(3), 634–643.

Cai, X. and G. B. Giannakis (2005, January).
Adaptive PSAM accounting for channel estimation and prediction errors.
IEEE Transactions on Wireless Communications 4(1), 246–256.

Cavers, J. K. (1972, February).
Variable-rate transmission for Rayleigh fading channels.
IEEE Transactions on Communications 20(1), 15–22.

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

Chen, T.-A., M. P. Fitz, S. Li, and M. D. Zoltowski (2004, June).
Two-dimensional space-time pilot-symbol assisted demodulation for frequency-nonselective Rayleigh fading channels.
IEEE Transactions on Communications 52(6), 953–963.

Chung, S. T. and A. J. Goldsmith (2001, September).
Degrees of freedom in adaptive modulation: a unified view.
IEEE Transactions on Communications 49(9), 1561–1571.

Dong, X. and N. C. Beaulieu (2002, January).
Optimal maximal ratio combining with correlated diversity branches.
IEEE Communication Letters 6(1), 22–24.

Duel-Hallen, A., S. Hu, and H. Hallen (2000, May).
Long-range prediction of fading signals: Enabling adaptive transmission for mobile radio channels.
IEEE Signal Processing Magazine 17(3), 62–75.

Duong, D. V. (2002).
Channel prediction for adaptive coded modulation with limited predictor complexity.
Master's thesis, Norwegian University of Science and Technology.
Available at: www.iet.ntnu.no/projects/beats/theses.htm (in Norwegian).

Duong, D. V., B. Holter, and G. E. Øien (2005, June).
Optimal pilot spacing and power in rate-adaptive MIMO diversity systems with imperfect transmitter CSI.
In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, New York City, USA.

Duong, D. V., B. Holter, and G. E. Øien (2006a).
Analysis and optimization of adaptive coded modulation systems in spatially correlated SIMO Rayleigh fading channels.
submitted to *IEEE Transactions on Vehicular Technology*. Available at www.iet.ntnu.no/~duong/publications/duong-VT06.pdf.

Duong, D. V., B. Holter, and G. E. Øien (2006b, September).
Analysis and optimization of SIMO systems with adaptive coded modulation in spatially correlated Rayleigh fading.
In *Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, Montreal, Quebec, Canada.
(accepted for publication).

Duong, D. V. and G. E. Øien (2004, August).
Adaptive trellis-coded modulation with imperfect channel state information at the receiver and transmitter.
In *Proc. Nordic Radio Symposium*, Oulu, Finland.

Duong, D. V. and G. E. Øien (2005, January).
Adaptive trellis-coded modulation with receive antenna diversity and imperfect channel knowledge at both receiver and transmitter.
In *12th Management Committee Meeting, COST Action 273: Towards Mobile Multimedia Broadband Networks*, Bologna, Italy.
Temorary document TD(05)-009.

Duong, D. V. and G. E. Øien (2006).
Optimal pilot spacing and power in rate-adaptive MIMO diversity system with imperfect CSI.
Accepted for publication in *IEEE Transactions on Wireless Communications*. Available at
www.iet.ntnu.no/~duong/publications/twcomm.pdf.

Duong, D. V., G. E. Øien, and K. J. Hole (2005, September).
Adaptive coded modulation with receive antenna diversity and imperfect channel knowledge at receiver and transmitter.
In *Proc. European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey.

Duong, D. V., G. E. Øien, and K. J. Hole (2006, March).
Adaptive coded modulation with receive antenna diversity and imperfect channel knowledge at receiver and transmitter.
IEEE Transactions on Vehicular Technology 55(2), 458–465.

Duong, D. V., M. Å. Wingar, and G. E. Øien (2005, September).
Link adaptation in correlated antenna diversity environments.
In *Proc. Norwegian Signal Processing Symposium (NORSIG)*, Stavanger, Norway.

Ekpenyong, A. E. and Y.-F. Huang (2004, November).
Markov channel-based feedback schemes for adaptive modulation systems.
In *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dallas, Texas, USA.

Ekpenyong, A. E. and Y.-F. Huang (2005, March).
Bayesian feedback detection for adaptive transmission systems.
In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, USA.

Falahati, S., A. Svensson, T. Ekman, and M. Sternad (2004, February).
Adaptive modulation systems for predicted wireless channels.
IEEE Transactions on Communications 52(2), 307–316.

Falahati, S., A. Svensson, M. Sternad, and H. Mei (2003, October).
Adaptive trellis-coded modulation over predicted flat fading channels.
In *Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, Orlando, Florida.

Gjendemsjø, A., G. E. Øien, and H. Holm (2005, November).
Optimal power control for discrete-rate link adaptation schemes with capacity-approaching coding.
In *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, St. Louis, Missouri, USA.

Gjendemsjø, A., G. E. Øien, and P. Orten (2006, June).
Optimal discrete-level power control for adaptive coded modulation schemes with capacity-approaching component codes.
In *Proc. IEEE International Conference on Communications (ICC)*, Istanbul, Turkey.

Godara, L. C. (1997a, August).
Application of antenna arrays to mobile communications, part II: beam-forming and direction-of-arrival considerations.
Proc. of the IEEE 85(8), 1195–1245.

Godara, L. C. (1997b, July).
Applications of antenna arrays to mobile communications, part I: performance improvement, feasibility, and system considerations.
Proc. of the IEEE 85(7), 1031–1060.

Goeckel, D. L. (1999, June).
Adaptive coding for time-varying channels using outdated fading estimates.
IEEE Transactions on Communications 47(6), 844–855.

Goldsmith, A. (2005).
Wireless Communications.
Cambridge University Press.

Goldsmith, A. J. and S.-G. Chua (1997, October).
Variable-rate variable-power MQAM for fading channels.
IEEE Transactions on Communications 45(10), 1218–1230.

Goldsmith, A. J. and S.-G. Chua (1998, May).
Adaptive coded modulation for fading channels.
IEEE Transactions on Communications 46(5), 595–602.

Goldsmith, A. J. and P. P. Varaiya (1997, November).
Capacity of fading channels with channel side information.
IEEE Transactions on Information Theory 43(6), 1986–1992.

Gradshteyn, I. S. and I. M. Ryzhik (2000).
Table of Integrals, Series and Products (6 ed.).
Academic Press.

Hanzo, L., M. Münster, B. J. Choi, and T. Keller (2003).
OFDM and MC-CDMA for Broadcasting Multi-User Communications, WLANs and Broadcasting.
John Wiley & Sons.

Hanzo, L., C. H. Wong, and M. S. Yee (2002).
Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalized and Space-Time Coded TDMA, CDMA and OFDM Systems.
John Wiley & Sons.

Hasna, M. O., M.-S. Alouini, and M. K. Simon (2001, October).
Effect of fading correlation on the outage probability of cellular mobile radio systems.
In *Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, Atlantic City, New Jersey, USA.

Hayes, J. F. (1968, February).
Adaptive feedback communications.
IEEE Transactions on Communications 16(1), 29–34.

Ho Van Khuong and H.-Y. Kong (2006, March).
General expression for pdf of a sum of indedendent exponential random variables.
IEEE Communication Letters 10(3), 159–161.

Hoeher, P., S. Kaiser, and P. Robertson (1997, April).
Two-dimensional pilot-symol-aided channel estimation by Wiener filtering.
In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Munich, Germany.

Hole, K. J., H. Holm, and G. E. Øien (2000, July).
Adaptive multi-dimensional coded modulation over flat fading channels.
IEEE Journal on Selected Areas in Communications 18(7), 1153–1158.

Hole, K. J., H. Holm, and G. E. Øien (2001, September).
Analysis of adaptive coded modulation with antenna diversity and feedback delay.
In *Proc. IST Mobile Communications Summit*.

Hole, K. J. and G. E. Øien (2001, January).
Spectral efficiency of adaptive coded modulation in urban microcellular networks.
IEEE Transactions on Vehicular Technology 50(1), 205–222.

Holm, H. (2002).
Adaptive Coded Modulation Performance and Channel Estimation Tools for Flat Fading Channels.
Ph. D. thesis, Norwegian University of Science and Technology.
Available at: www.iet.ntnu.no/projects/beats/theses.htm.

Holm, H. and M.-S. Alouini (2004, August).
Sum and difference of two squared correlated Nakagami variates in connection with the McKay distribution.
IEEE Transactions on Communications 52(8), 1367–1376.

Holter, B. (2005).
Adaptive Coded Modulation in Spatial and Multiuser Diversity Systems.
Ph. D. thesis, Norwegian University of Science and Technology.
Available at: www.iet.ntnu.no/projects/beats/theses.htm.

Holter, B. and G. E. Øien (2005, September).
On the amount of fading in MIMO diversity systems.
IEEE Transactions on Wireless Communications 4(5), 2498–2507.

Holter, B. and G. E. Øien (2006).
Impact of spatial correlation on adaptive coded modulation performance in Rayleigh fading.
to appear in *IEEE Transactions on Vehicular Technology*. Available at www.iet.ntnu.no/~duong/publications/holter-vt-2005.pdf.

Holter, B., G. E. Øien, K. J. Hole, and H. Holm (2003, April).
Limitations in spectral efficiency of a rate adaptive MIMO system utilizing pilot-aided channel prediction.
In *Proc. IEEE Vehicular Technology Conference (VTC-Spring)*, Jeju, Korea.

Hu, S. and A. Duel-Hallen (2001, November).
Combined adaptive modulation and transmitter diversity using longrange prediction for flat fading mobile radio channels.
In *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, San Antonio, Texas, USA, pp. 24–29.

Hu, S., A. Duel-Hallen, and H. Hallen (2000, March).

Long range predictionmakes adaptive modulation feasible for realistic mobile readio channels.
In *Proc. Annual Conference on Information Sciences and Systems*, Princeton, NJ, USA, pp. WP4–7–WP4–13.

Jafar, S. A. and A. Goldsmith (2001, June).
On optimality of beamforming for multiple antenna systems with imperfect feedback.
In *Proc. IEEE International Symposium on Information Theory*, Washington DC, USA, pp. 321.

Jafar, S. A., S. Vishwanath, and A. Goldsmith (2001, June).
Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback.
In *Proc. IEEE International Conference on Communications (ICC)*, Helsinki, Finland, pp. 2266–2270.

Jakes, W. C. (1994).
Microwave Mobile Communications.
An IEEE Press Classic Reissue, New Jersey: IEEE Press.

Ko, Y. and C. Tepedelenlioglu (2003, November).
Space-time block coded rate-adaptive modulation with uncertain SNR feedback.
In *Proc. Asilomar Conference On Systems and Computers*.

Ko, Y. and C. Tepedelenlioglu (2004, May).
Optimal switching thresholds for space-time block coded rate-adaptive M-QAM.
In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada.

Ko, Y. and C. Tepedelenlioglu (2006, February).
Orthogonal space-time block coded rate-adaptive modulation with outdated feedback.
IEEE Transactions on Wireless Communications 5(2), 290–295.

Lau, V. K. N. and M. D. Macleod (2001, September).
Variable-rate adaptive trellis coded QAM for flat-fading channels.
IEEE Transactions on Communications 49(9), 1550–1560.

Lee, E. A. and D. G. Messerschmitt (2000).
Digital Communication (2 ed.).
Kluwer Academic Publishers.

Lee, W. C. Y. (1970, August).
Level crossing rates of an equal-gain predetection diversity combiner.
IEEE Transactions on Communications Technology COM-18(4), 417–426.

Li, Y. (2000, July).
Pilot-symbol-aided channel estimation for OFDM in wireless systems.
IEEE Transactions on Vehicular Technology 49(4), 1207–1215.

Liu, Q., S. Zhou, and G. B. Giannakis (2004, September).
Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless link.
IEEE Transactions on Wireless Communications 3(5), 1746–1755.

Loskot, P. and N. C. Beaulieu (2004, March).
Maximum ratio combining with arbitrary correlated generalized Ricean branches.
In *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Atlanta, Georgia, USA, pp. 333–338.

Ma, X., G. B. Giannakis, and S. Ohno (2003, May).
Optimal training for block transmission over doubly selective wireless fading channels.
IEEE Transactions on Signal Processing 51(5), 1351–1366.

Meyr, H., M. Moeneclaey, and S. A. Fechtel (1998).
Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing.
John Wiley & Sons.

Moon, T. K. and W. C. Stirling (2000).
Mathematical Methods and Algorithms for Signal Processing.
Prentice Hall, Inc.

Mun, C., C.-H. Kang, and H.-K. Park (1999, February).
Approximation of SNR statistics for MRC diversity systems in arbitrarily correlated Nakagami fading channels.
Electronics Letters 35(4), 266–267.

Nakagami, M. (1960).
The *m*-distribution—a general formula of intensity distribution of rapid fading.
Statistical Methods in Radio Wave Propagation, Oxford, U.K.: Pergamon Press, 3–36.

Ohno, S. and G. B. Giannakis (2004, September).
Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels.
IEEE Transactions on Information Theory 50(9), 2138–2145.

Papoulis, A. and S. U. Pillai (2002).
Probability, Random Variables and Stochastic Processes (4 ed.).
McGraw-Hill Higher Education.

Paulraj, A., R. Nabar, and D. Gore (2003).
Introduction to Space-Time Wireless Communications.
Cambridge University Press.

Råde, L. and B. Westergren (2000).
Mathematics Handbook for Science and Engineering.
Studentlitteratur.

Rappaport, T. S. (2002).
Wireless Communications Principles & Practice (2 ed.).
Prentice Hall.

Ratnarajah, T. (2006, April).
Spatially correlated multiple-antenna channel capacity distributions.
IEE Proceedings-Communications 153(2), 263–271.

Sampei, S. and T. Sunaga (1993, May).
Rayleigh fading compensation for QAM in land mobile radio communications.
IEEE Transactions on Vehicular Technology 42(2), 137–147.

Sandhu, S. and A. Paulraj (2000, December).
Space-time block codes: A capacity perspective.
IEEE Communication Letters 4(12), 384–386.

Simon, M. K. and M.-S. Alouini (2005).
Digital Communication over Fading Channels: A Unified Approach to Performance Analysis (2 ed.).
John Wiley & Sons Inc.

Smith, P. J. and M. Shafi (2004, June).
The impact of complexity in MIMO channel models.
In *Proc. IEEE International Conference on Communications (ICC)*, pp. 2924–2928.

Stüber, G. L. (1996).
Principles of Mobile Communications.
Kluwer Academic Publisher.

Tang, X., M.-S. Alouini, and A. J. Goldsmith (1999, December).
Effect of channel estimation error on *M*-QAM BER performance in Rayleigh fading.
IEEE Transactions on Communications 47(12), 1856–1864.

Tarokh, V., H. Jafarkhami, and A. R. Calderbank (1999, July).
Space-time block codes from orthogonal designs.
IEEE Transactions on Information Theory 45(5), 1456–1467.

Temme, N. M. (1996).
Special Functions: An Introduction to the Classical Functions of Mathematical Physics.
New York: John Wiley & Sons Inc.

Therrien, C. W. (1992).
Discrete Random Signals and Statistical Signal Processing.
Signal Processing Series. Englewood Cliffs, NJ: Prentice Hall.

Torrance, J. M. and L. Hanzo (1995, September).
Comparative study of pilot symbol assisted modem schemes.
In *IEE Sixth International Conference on Radio Receivers and Associated Systems*, Bath, pp. 36–41.

Torrance, J. M. and L. Hanzo (1996, June).
Optimisation of switching levels for adaptive modulation in slow Rayleigh fading.
IEE Electronics Letters 32(13), 1167–1169.

Ue, T., S. Sampei, N. Morinaga, and K. Hamaguchi (1998, November).
Symbol rate and modulation level-controlled adaptive modulation/TDMA/TDD system for high-bit-rate wireless data transmission.
IEEE Transactions on Vehicular Technology 47, 1134–1147.

Vucetic, B. (1991, May).
An adaptive coding scheme for time-varying channels.
IEEE Transactions on Communications 39(5), 653–662.

Wang, Z. and G. B. Giannakis (2003, August).

A simple and general parameterization quantifying performance in fading channels.
IEEE Transactions on Communications 51(8), 1389–1398.

Webb, W. T. and R. Steele (1995, July).
Variable rate QAM for mobile radio.
IEEE Transactions on Communications 43(7), 2223–2230.

Winters, J. H. (1984, July).
Optimum combining in digital mobile radio with cochannel interference.
IEEE Journal on Selected Areas in Communications SAC-2(4), 528–539.

Winters, J. H. (1998, February).
Smart antennas for wireless systems.
IEEE Personal Communications 5, 23–27.

You, R., H. Li, and Y. Bar-Ness (2005, October).
Diversity combining with imperfect channel estimation.
IEEE Transactions on Communications 53(10), 1655–1662.

Zhou, S. and G. B. Giannakis (2002, October).
Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback.
IEEE Transactions on Signal Processing 50(10), 2599–2613.

Zhou, S. and G. B. Giannakis (2004a, September).
Adaptive modulation for multi-antenna transmissions with channel mean feedback.
IEEE Transactions on Wireless Communications 3(5), 1626–1636.

Zhou, S. and G. B. Giannakis (2004b, July).
How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?
IEEE Transactions on Wireless Communications 3(4), 1285–1294.

Øien, G. E., R. K. Hansen, D. V. Duong, H. Holm, and K. J. Hole (2002, October).
Bit error rate analysis of adaptive coded modulation with mismatched and complexity-limited channel prediction.
In *Proc. IEEE Nordic Signal Processing Symposium (NORSIG)*, Hurtigruten, Norway.

Øien, G. E. and K. J. Hole (2001, February).
Maximum average spectral efficiency of slowly varying rayleigh fading channels with pilot-symbol-assisted channel estimation.
In *Proc. European Personal Mobile Communications Conference*, Vienna.

Øien, G. E., H. Holm, and K. J. Hole (2002a, September).
Adaptive coded modulation with imperfect channel state information: System design and performance analysis aspects.
In *Proc. International Symposium on Advances in Wireless Communications (ISWC)*, Victoria, BC, Canada.

Øien, G. E., H. Holm, and K. J. Hole (2002b, September).
Channel prediction for adaptive coded modulation in rayleigh fading.
In *Proc. European Signal Processing Conference (EUSIPCO)*, Toulouse.

Øien, G. E., H. Holm, and K. J. Hole (2004, May).
Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading.
IEEE Transactions on Vehicular Technology 53(3), 758–769.