OULU 2008
C 289

ACTA UNIVERSITATIS OULUENSIS

Mauri Nissilä

ITERATIVE RECEIVERS FOR DIGITAL COMMUNICATIONS VIA VARIATIONAL INFERENCE AND ESTIMATION

FACULTY OF TECHNOLOGY,
DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING,
INFOTECH OULU,
UNIVERSITY OF OULU

ACTA UNIVERSITATIS OULUENSIS
C Technica 289

MAURI NISSILÄ

ITERATIVE RECEIVERS FOR DIGITAL COMMUNICATIONS VIA VARIATIONAL INFERENCE AND ESTIMATION

Academic dissertation to be presented, with the assent of the Faculty of Technology of the University of Oulu, for public defence in Raahensali (Auditorium L10), Linnanmaa, on January 18th, 2008, at 12 noon

OULUN YLIOPISTO, OULU 2008

Acta Univ. Oul. C 289, 2008

Supervised by
Research Professor Aarne Mämmelä

Reviewed by
Professor Riccardo Raheli
Professor Bhaskar D. Rao

ISBN 978-951-42-8685-8 (Paperback)
ISBN 978-951-42-8686-5 (PDF)
http://herkules.oulu.fi/isbn9789514286865/
ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)
http://herkules.oulu.fi/issn03553213/

Cover design
Raimo Ahonen

OULU UNIVERSITY PRESS
OULU 2008

Nissilä, Mauri, Iterative receivers for digital communications via variational inference and estimation
Faculty of Technology, University of Oulu, P.O.Box 4000, FI-90014 University of Oulu, Finland, Department of Electrical and Information Engineering, Infotech Oulu, University of Oulu, P.O.Box 4500, FI-90014 University of Oulu, Finland
Acta Univ. Oul. C 289, 2008
Oulu, Finland

Abstract

In this thesis, iterative detection and estimation algorithms for digital communications systems in the presence of parametric uncertainty are explored and further developed. In particular, *variational* methods, which have been extensively applied in other research fields such as artificial intelligence and machine learning, are introduced and systematically used in deriving approximations to the optimal receivers in various channel conditions. The key idea behind the variational methods is to transform the problem of interest into an optimization problem via an introduction of extra degrees of freedom known as *variational parameters*. This is done so that, for fixed values of the free parameters, the transformed problem has a simple solution, solving approximately the original problem.

The thesis contributes to the state of the art of advanced receiver design in a number of ways. These include the development of new theoretical and conceptual viewpoints of iterative turbo-processing receivers as well as a new set of practical joint estimation and detection algorithms. Central to the theoretical studies is to show that many of the known low-complexity turbo receivers, such as linear minimum mean square error (MMSE) soft-input soft-output (SISO) equalizers and demodulators that are based on the Bayesian expectation-maximization (BEM) algorithm, can be formulated as solutions to the variational optimization problem. This new approach not only provides new insights into the current designs and structural properties of the relevant receivers, but also suggests some improvements on them.

In addition, SISO detection in multipath fading channels is considered with the aim of obtaining a new class of low-complexity adaptive SISOs. As a result, a novel, unified method is proposed and applied in order to derive recursive versions of the classical Baum-Welch algorithm and its Bayesian counterpart, referred to as the BEM algorithm. These formulations are shown to yield computationally attractive *soft decision-directed* (SDD) channel estimators for both deterministic and Rayleigh fading intersymbol interference (ISI) channels.

Next, by modeling the multipath fading channel as a *complex bandpass* autoregressive (AR) process, it is shown that the statistical parameters of radio channels, such as frequency offset, Doppler spread, and power-delay profile, can be conveniently extracted from the estimated AR parameters which, in turn, may be conveniently derived via an EM algorithm. Such a joint estimator for all relevant radio channel parameters has a number of virtues, particularly its capability to perform equally well in a variety of channel conditions.

Lastly, adaptive iterative detection in the presence of phase uncertainty is investigated. As a result, novel iterative joint Bayesian estimation and symbol *a posteriori* probability (APP) computation algorithms, based on the variational Bayesian method, are proposed for both constant-phase channel models and dynamic phase models, and their performance is evaluated via computer simulations.

Keywords: Bayesian inference and estimation, EM algorithm, frequency offset estimation, turbo receivers

Acknowledgements

I would not have been able to complete this work without help. I am grateful to my advisor Professor Subbarayan Pasupathy from the University of Toronto for his continual support and encouragement over the years. His pursuit for the highest standards in research, combined with his insights, enthusiasm, and knowledge, has been a great inspiration for my work. I am also grateful for his hospitality during my research visit at the University of Toronto. I would also like to thank my supervisor, Research Professor Aarne Mämmelä from VTT, for his patient and persistent efforts to help me become a better writer and researcher. I am also grateful to him for giving me the opportunities to carry out research for this thesis during my stay at VTT, while providing me with the freedom to pursue topics of my interest. In addition, I would like to express my gratitude to the reviewers of the thesis, Professor Bhaskar D. Rao from the University of California, San Diego, USA, and Professor Riccordo Raheli from the University of Parma, Italy, for their constructive comments and suggestions which have significantly helped me to improve the quality of this thesis.

Although presented as a monograph, this thesis is a collection of papers that have been published or accepted for publication in international journals. The text of Chapter 3, in part, is a reprint of the material as it appears in: M. Nissilä and S. Pasupathy, "Soft-input soft-output equalizers for turbo receivers: A statistical physics perspectives," *IEEE Transactions on Commununications*, vol. 55, pp. 1300–1307, Jul. 2007. The text of Chapter 4, in part, is a reprint of the material as it appears in: M. Nissilä and S. Pasupathy, "Adaptive Bayesian and EM-based detectors for frequency-selective fading channels," *IEEE Transactions on Commununications*, vol. 51, pp. 1325–1336, Aug. 2003. The text of Chapter 5, in part, is a reprint of the material as it appears in: M. Nissilä and S. Pasupathy, "Joint estimation of carrier frequency offset and statistical parameters of the multipath fading channel," *IEEE Transactions on Commununications*, vol. 54, pp. 1038–1048, Jun. 2006. The text of Chapter 6, in part, is based on the material which has been accepted for publication as: M. Nissilä and S. Pasupathy, "Adaptive iterative detectors for phase-uncertain channels via variational bounding," *IEEE Transactions on Commununications*. I was the primary author for all these papers, and the co-author, Prof. Pasupathy, provided me with criticism and technical as well as editorial advice, for which he is acknowledged. In addition, the anonymous reviewers of these

papers as well as the editors, who were nominated to handle these papers, provided many valuable comments and suggestions, which has greatly influenced the above papers and this thesis as well. I am grateful for this help. Also, I would like to thank Anna Shepherd for proofreading the thesis.

The financial support provided by the Infotech Oulu Graduate School in Electronics, Telecommunications, and Automation, VTT Electronics, the Nokia Foundation, the Tauno Tönning Foundation, the Foundation for the Promotion of Technology (TES), the Oulu University Scholarship Foundation, and the Emil Aaltonen Foundation is gratefully acknowledged. This support also partly enabled my research visit to the University of Toronto.

I have been fortunate to be able to work in various research and development groups at different places in the course of my graduate studies, while the topics of my interest and work have ranged from an analog and mixed-signal circuit theory to various aspects of digital signal processing and digital communication theory. My line managers in all of these places have been highly supportive and friendly. In this respect, I am indebted to Juha Pikkarainen, Kari Rikkinen, and Ilkka Niva from Nokia Mobile Phones, Professor Pentti Leppänen from the University of Oulu, Kyösti Rautiola from VTT, and Markku Vainikka from my current employer, Nokia Siemens Networks (NSN). In particular, my special thanks are due to Kyösti Rautiola and Markku Vainikka, who have provided me with flexibility that has been needed in writing the research papers and finalizing this thesis. Also, I am grateful to many colleagues at VTT, at Nokia, and most lately at NSN for creating a pleasant atmosphere to work. From NSN, I am especially indebted to my colleague, Dr. Vinh Van Phan, for the valuable editorial suggestions he gave me at the final preparation phase of the thesis. My parents have always taught me the value of education and dedication to work. I appreciate that.

Lastly, I owe my deepest gratitude to my family. Especially, thanks to my wife, Minna, for her invaluable company, motivation, and support, which have made this a very pleasant journey. Our six children, Heikki (12 years), Tuomo (10 years), Karoliina (8 years), Hanna (6 years), Elias (4 years), and Alma (1 year), have occasionally been successful in slowing down this work with so much more important things. Nevertheless, I managed to complete it, finally. Without you all, that would have maybe never happened. I dedicate this thesis to my family.

Oulu, November 2007 Mauri Nissilä

Symbols and abbreviations

$\underline{\boldsymbol{A}}$	extended (baseband) AR coefficient matrix ($p(L+1) \times p(L+1)$)
$\boldsymbol{A}$	extended (complex bandpass) AR coefficient matrix ($p(L+1) \times p(L+1)$)
$\bar{\mathbf{A}}_i$	ith order (baseband) AR coefficient matrix ($(L+1) \times (L+1)$)
$\mathbf{A}_i$	ith order (complex bandpass) AR coefficient matrix ($(L+1) \times (L+1)$)
a_l	lth diagonal element of $\mathbf{A}_1$
$\tilde{\mathbf{A}}_l$	matrix including the AR coefficients of the lth channel tap ($p \times p$)
$\mathbf{a}_l$	lth column of $\boldsymbol{A}^T$
$\mathcal{A}_{k,l}(j)$	a set used in defining the approximate trellis branch probabilities
$\mathrm{APP}_d(s(k))$	*a posteriori* probability of the symbol s_k in the presence of the deterministic channel
$\mathrm{APP}_p(s(k))$	*a posteriori* probability of the symbol s_k in the presence of the random channel
$\boldsymbol{B}$	matrix of the MA model parameters ($(L+1) \times (q+1)(L+1)$)
$\mathbf{B}_i$	ith submatrix of $\boldsymbol{B}$ ($(L+1) \times (L+1)$)
$\mathbf{b}_l$	lth column of $\boldsymbol{B}^T$
B_d	Doppler spread of the fading channel
$\mathcal{B}_k(l)$	a set used in defining the forward-backward processing SISO (BCJR) algorithm
$\mathcal{C}_k(j)$	a set used in defining the forward-backward processing SISO (BCJR) algorithm
$\mathbb{C}$	set of complex numbers
C, c, c_k	constant
$\mathcal{C}$	set of code words (i.e., code)
$\mathcal{C}_n$	code applied to the signal at the transmitter antenna n
$\mathbf{C}_h$	covariance matrix of the FS fading channel
$\mathbf{C}_{h_l}$	normalized covariance matrix of the lth fading channel tap
$C_{h_{m,n,l}}(i)$	autocorrelation function of l-th channel tap of the subchannel between antennas m and n
$\mathbf{C}_{\bar{\mathbf{w}}(k)}(i)$	covariance matrix of the AR process noise vector $\bar{\mathbf{w}}(k)$ ($(L+1) \times (L+1)$)

$\mathbf{c}_{n,k,\text{CMF}}$	filter coefficients of the LCMF SISO detector ($M(K+L)\times 1$)
$d_{n,k}$	degree of the symbol node for $s_n(k)$
$D(\cdot\|\cdot)$	Kullback Leibler divergence
$\tilde{\mathbf{D}}(k)$	data matrix of the state-space model related to the SDD-KS ($J^{L+1}\times p(L+1)$)
$\tilde{\mathbf{D}}_{\text{e}}(k)$	data matrix of the state-space model related to the SDD-RCKS ($(L+1)J^{L+1}\times p(L+1)$)
$\mathcal{D}_k(j)$	a set used in defining the forward-backward processing SISO (BCJR) algorithm
$e(k)$	output of the FED at time k
$\mathbf{e}_{n,k}$	unit vector ($NK\times 1$)
$\mathcal{E}(\cdot)$	posterior log likelihood of the argument
$E_{n,k}$	energy of the symbol $s_n(k)$ at the receiver
E_b	energy of information bit
E_s	received energy per symbol
$\mathbf{F}$	set of CIR snapshots
$\tilde{\mathbf{F}}_l$	set of channel values belonging to the lth channel path
$\mathcal{F}(Q)$	variational free energy (defined with respect to trial distribution Q)
$\mathcal{F}_{\text{Bethe}}(Q)$	variational Bethe free energy
$\mathcal{F}_{\text{Helmholtz}}$	Helmholz free energy
$\mathcal{F}_k(l)$	a set used in defining the forward-backward processing SISO (BCJR) algorithm
$\boldsymbol{f}(k)$	vector obtained by concatenating samples of the frequency shifted channel impulse response and its delayed versions at time k (in a single antenna system) ($p(L+1)\times 1$)
$\mathbf{f}(k)$	vector of samples of the frequency shifted impulse response at time k ($(L+1)\times 1$)
$\tilde{\mathbf{f}}_l(k)$	vector of channel values belonging to the lth channel tap at time k ($p\times 1$)
$f_l(k)$	value of the lth frequency shifted channel tap in a single antenna system at time instance k
G_k	Kalman gain factor
$\mathcal{G}(\cdot,\cdot)$	inverse variational free energy functional
$\mathcal{G}_k(\cdot,\cdot,\cdot)$	inverse variational free energy functional corresponding to the kth received signal sample
$\boldsymbol{G}$	extended matrix of standard deviations of the AR process noise samples

	$(p(L+1)\times(L+1))$
$\mathbf{G}$	diagonal matrix of standard deviations of the AR process noise samples $((L+1)\times(L+1))$
g_l	variance of the l-th element of the process noise vector $\bar{\mathbf{w}}(k)$
$g(t)$	time-continuous impulse response of the receiver front-end filter
$\mathcal{H}(\cdot)$	entropy of the argument
$\mathcal{H}_{\text{Bethe}}(\cdot)$	Bethe entropy
$\mathbf{h}$	vector of complex time-varying channel gains ($MNL_{\text{ch}}(K+L)\times 1$) or impulse response of the time-invariant channel ($MN(L+1)\times 1$)
$\breve{\mathbf{h}}_{m,n}$	vector of complex time-varying channel gains for the subchannel between the mth receiver and the nth transmitter antenna ($L_{\text{ch}}(K+L)\times 1$)
$\mathbf{h}_{m,n,l}$	vector of complex time-varying channel gains for the lth tap of the subchannel between the mth receiver and the nth transmitter antenna $((K+L)\times 1)$
$h_{m,n,l}(k)$	value of the lth channel tap of the subchannel between antennas n and m at time instance k
$h_l(k)$	value of the lth channel tap in a single antenna system at time instance k
$\mathbf{h}(k)$	vector of samples of the impulse responses of all subchannels at time k $(MN(L+1)\times 1)$
$\boldsymbol{h}(k)$	vector obtained by concatenating samples of the channel impulse response and its delayed versions at time k (in a single antenna system) $(p(L+1)\times 1)$
$\mathbf{h}_m(k)$	vector of samples of the impulse responses from all transmitter antennas to the mth receiver antenna at time k $(N(L+1)\times 1)$
$\mathbf{h}_{m,n}(k)$	vector of samples of the impulse response from the nth transmitter antenna to the mth receiver antenna at time k $((L+1)\times 1)$
$\bar{\mathbf{h}}_{m,n}(k)$	kth column vector of the subchannel convolution matrix $\mathbf{H}_{m,n}$ $((K+L)\times 1)$
$\mathbf{H}$	channel convolution matrix including the effects of frequency instabilities $(M(K+L)\times NK)$
$\mathbf{H}_{m,n}$	channel convolution matrix for the subchannel between the mth receiver and the nth transmitter antenna $((K+L)\times K)$
i	discrete index
$I_0(\cdot)$	zeroth order modified Bessel function of the first kind

$I_1(\cdot)$	first order modified Bessel function of the first kind
$\mathbf{I}_M$	identity matrix ($M \times M$)
$I_C(\cdot)$	indicator function for the code C
$\mathcal{I}(\cdot)$	information function
$\mathcal{I}_e(\cdot)$	empirical information function
$\boldsymbol{\mathcal{I}}$	information matrix
$\boldsymbol{\mathcal{I}}_c$	complete-data information matrix
$\boldsymbol{\mathcal{I}}_m$	missing information matrix
j	index for the constellation points
J	number of symbol constellation points
$\mathcal{J}_0$	zero-order Bessel function of the first kind
k	discrete time index
K	number of symbols in the data block
K_{data}	number of data symbols in a transmitted block of symbols
K_{pre}	number of symbols in the preamble
K_{prefix}	number of symbols in the cyclic prefix
K_{pilot}	number of pilot symbols in the transmission frame
K_{tail}	number of tail symbols
$\mathbf{k}(k,j)$	running gain vector of the SDD-RLS algorithm ($(L+1)\times 1$)
l	channel tap index
L	degree of channel memory
L_{ch}	length of the channel impulse response in each subchannel
$\mathcal{L}(\cdot)$	log likelihood function
$\mathcal{L}_c(\cdot)$	complete-data log likelihood function
m	receiver antenna index
M	number of receiver antennas
n	transmitter antenna index
$\tilde{n}_l(k)$	receiver noise of the state-space model corresponding to the lth channel path
N	number of transmitter antennas
N_0	receiver noise spectral density
$N_{innerIter}$	number of inner iterations in the turbo receiver
$N_{outerIter}$	number of outer iterations in the turbo receiver
p	order of the autoregressive model describing the fading dynamics
$p(\cdot\|\cdot)$	conditional probability distribution

$p(\cdot)$	probability density function
$P(\cdot)$	probability mass function
$p_r(\cdot)$	conditional distribution of the argument given the received signal sequence $\mathbf{r}$
$\mathbf{P}(k,j)$	running inverse auto-correlation matrix of the SDD-RLS algorithm ($(L+1)\times(L+1)$)
$\mathbf{P}(k\|K)$	error covariance of FI Kalman smoother
$p(t)$	time-continuous impulse response of the pulse-shaping filter
q	order of the moving averaging model describing the fading dynamics
$q_{n,k}$	trial "factor" distribution corresponding to the symbol $s_n(k)$
q_k	trial "factor" distribution corresponding to the symbol vector $\mathbf{s}(k)$
$q_x(x)$	trial "factor" distribution which is a function of variable x
$q_{\mathbf{x}}(\mathbf{x})$	trial "factor" distribution which is a function of variable vector $\mathbf{x}$
$Q(\cdot)$	variational distribution function
$Q^{(n)}(\cdot)$	variational distribution function for the nth substructure of structured MF approximation
$Q_{\text{BMF}}(\cdot)$	variational distribution corresponding to Bayesian mean field approximation
$Q_{\text{MF}}(\cdot)$	variational distribution corresponding to naive mean field approximation
$Q(\cdot\|\cdot)$	objective function of the EM algorithm
$\mathbf{r}$	vector of received signal samples over the processing window ($M(K+L)\times 1$)
$\mathbf{r}_m$	vector of received signal samples belonging to the mth receiver antenna ($(K+L)\times 1$)
$\mathbf{r}(k)$	vector of received signal samples at time instance k ($M\times 1$)
$\tilde{\mathbf{r}}(k)$	received signal vector of the state-space model related to the SDD-KS ($J^{L+1}\times 1$)
$r_m(k)$	received signal sample at the mth receiver antenna at time instance k
$r(k)$	received signal sample at a single receiver antenna at time instance k
$\mathbb{R}$	set of real numbers
$\mathbf{s}$	vector of transmitted information symbols (in one data block) collected from all transmitter antennas ($NK\times 1$)
$\mathbf{s}_n$	vector of transmitted information symbols (in one data block) from the nth transmitter antenna ($K\times 1$)

$\mathbf{s}_n^{\mathrm{e}}$	cyclically extended vector of transmitted information symbols from the nth transmitter antenna ($(K + K_{\text{prefix}}) \times 1$)
$\mathbf{s}_n^{\mathrm{cp}}$	cyclic prefix for the nth transmitter antenna ($K_{\text{prefix}} \times 1$)
$s_n(k)$	information symbol transmitted from the nth transmitter antenna at time instance k
$\overline{\mathbf{s}}$	vector of soft symbols (soft input to the linear SISO detector) ($NK \times 1$)
$\overline{\mathbf{s}}_n$	vector of soft symbols (soft input to the linear SISO detector) corresponding to the nth transmitter antenna signal ($K \times 1$)
$\overline{s}_n(k)$	soft symbol value (soft input to the linear SISO detector)
$\mathbf{s}(k)$	vector of all symbols that are contributing to the received signal at time k ($N(L+1) \times 1$)
$\boldsymbol{s}(k)$	zero-padded symbol vector at time instance k (in a single antenna system) ($p(L+1) \times 1$)
$\mathbf{s}_n(k)$	vector of symbols transmitted from the nth transmitter antenna that are contributing to the received signal at time k ($(L+1) \times 1$)
$\acute{\mathbf{s}}_n(k)$	trellis state vector for the nth Markov chain at time k ($L \times 1$)
$\mathcal{S}$	discrete symbol space
T	temperature
T_s	length of a symbol period
$T_k^{(n)}(\cdot,\cdot,\cdot)$	trellis check function at time k for the nth subsystem
$\mathcal{U}(Q)$	variational average energy (defined with respect to trial distribution Q)
$\mathbf{v}$	vector of receiver noise samples over the processing window ($M(K + L) \times 1$)
$\mathbf{v}_m$	vector of receiver noise samples pertaining to the mth receiver antenna ($(K + L) \times 1$)
$\mathbf{v}(k)$	vector of receiver noise samples at time instance k ($M \times 1$)
$\tilde{\mathbf{v}}(k)$	noise vector of the state-space model related to the SDD-KS ($J^{L+1} \times 1$)
$\tilde{\mathbf{v}}_{\mathrm{e}}(k)$	noise vector of the state-space model related to the SDD-RCKS ($(L + 1)J^{L+1} \times 1$)
$v_m(k)$	receiver noise sample at the mth receiver antenna at time instance k
$v(k)$	receiver noise sample at a single receiver antenna at time instance k
$\ddot{v}_l(k)$	noise component pertaining to the lth multipath at time instance k
$\bar{\mathbf{w}}(k)$	vector of AR process noise samples at time instance k ($(L+1) \times 1$)
$\mathbf{w}(k)$	vector of complex bandpass AR process noise samples at time instance k ($(L+1) \times 1$)

$\bar{w}_l(k)$	AR process noise sample pertaining to the lth channel tap at time instance
$x_l(k)$	lth multipath component of the received signal
$\tilde{\mathbf{x}}(k)$	received signal vector of the state-space model related to the SDD-RCKS ($(L+1)J^{L+1} \times 1$)
$\tilde{\mathbf{x}}_l(k)$	received signal vector of multipath l of the state-space model related to the SDD-RCKS ($J^{L+1} \times 1$)
$\tilde{x}_{l,j}(k)$	received signal sample (pertaining to the lth multipath component and the jth hypothetic value of the data vector) of the state-space model related to the SDD-RCKS
$\hat{x}_{l,j}(k)$	pseudo-observation pertaining to the lth multipath component and the jth hypothetic value of the data vector
$\hat{\mathbf{x}}(k)$	vector of averaged pseudo-observations ($(L+1) \times 1$)
$\mathbf{X}_n$	matrix for the transmitted symbols from the nth transmitter antenna ($(K+L) \times L_{\text{ch}}(K+L)$)
$\mathbf{X}_{n,l}$	diagonal matrix of the transmitted symbols from the nth transmitter antenna delayed by l symbol period ($(K+L) \times (K+L)$)
$\bar{\mathbf{X}}$	matrix for the transmitted symbols distorted by the frequency instabilities ($M(K+L) \times MNL_{\text{ch}}(K+L)$)
$\bar{\mathbf{X}}_n$	matrix for the transmitted symbols from the nth transmitter antenna which are distorted by the frequency instabilities ($M(K+L) \times ML_{\text{ch}}(K+L)$)
$\mathcal{X}$	complete data set of the EM algorithm
$\mathcal{Y}$	incomplete data set of the EM algorithm
$y_l(k)$	output of the lth channel path at time k
$\mathbf{z}(k,j)$	running cross-correlation vector of the SDD-RLS algorithm ($(L+1)\times 1$)
Z	normalizing constant
α_j	discrete (complex) symbol value
$\alpha(k,l)$	forward running product sum variable of the BCJR algorithm
$\beta(k,l)$	backward running product sum variable of the BCJR algorithm
$\gamma(k,l,m)$	conditional state probability of the trellis diagram
$\gamma_{n,k}$	normalizing constant of the detector output pdf corresponding to the symbol $s_n(k)$
$\delta(i)$	Dirac delta function
$\Delta(k)$	random phase increment at time k

ζ	a scaling factor used in the variational Bayesian-based soft symbol mapper
$\boldsymbol{\Theta}$	parameter vector
$\boldsymbol{\theta}_i$	ith parameter set
λ	forgetting factor
Λ	diagonal matrix of forgetting factors
Λ_k	diagonal matrix of time-varying forgetting factors at time k
$\lambda_{1,n,k}$	extrinsic information on the symbol $s_n(k)$ produced by the SISO detector
$\lambda_{2,n,k}$	extrinsic information on the symbol $s_n(k)$ produced by the channel decoder
μ	step size parameter
ν_l	carrier frequency shift pertaining to the l-th channel tap
ν	carrier frequency offset equal to all channel taps
ξ	attenuation of the AR spectrum relative to the peak value
$\boldsymbol{\xi}_j$	jth component vector in the vector space spanned by the symbol space $\mathcal{S}$ $((L+1)\times 1)$
$\boldsymbol{\Xi}(k,j)$	running auto-correlation matrix of the SDD-RLS algorithm $((L+1)\times(L+1))$
ρ	signal-to-noise ratio
ϱ_l^2	relative power level of the lth channel tap
σ_Δ^2	variance of the random phase increment
$\sigma_{h_{m,n,l}}^2$	variance of the fading gain $h_{m,n,l}$
$\sigma_{\tilde{n}_l}^2$	variance of the noise $\tilde{n}_l$
σ_v^2	variance of receiver noise
$\sigma_{\ddot{v}_l}^2$	variance of noise component pertaining to the lth multipath
$\sigma_{s_k}^2$	variance of the soft symbol estimate at time k
$\sigma_{i,n,k}^2$	variance of the soft symbol estimate (input to the detector)
$\sigma_{o,n,k}^2$	variance of the soft symbol estimate (output of the detector)
$\bar{\sigma}_i$	vector of soft symbol variances (input to the detector) $(NK\times 1)$
$\bar{\sigma}_o$	vector of soft symbol variances (output of the detector) $(NK\times 1)$
$\boldsymbol{\Sigma}_i$	diagonal covariance matrix of soft symbol estimates (input to the detector) $(NK\times NK)$
$\boldsymbol{\Sigma}_o$	diagonal covariance matrix of soft symbol estimates (output of the detector) $(NK\times NK)$

$\mathbf{\Sigma}_k$	diagonal covariance matrix of soft symbol estimates at time k ($(L+1)\times(L+1)$)
$\mathbf{\Upsilon}$	diagonal matrix needed for computation of the CRLB for the CFO ($K\times K$)
ϕ	carrier phase
$\phi(k)$	carrier phase at time k
$\mathbf{\Phi}_k$	conditional covariance of the CIR at time k given the received signal up to time K
$\mathbf{\Phi}_{k\|k}$	conditional covariance of the CIR at time k given the received signal up to time k
$\mathbf{\Phi}_{k-1\|k}$	conditional covariance of the CIR at time $k-1$ given the received signal up to time k
$\varphi(k,j)$	jth value of the trellis branch APP at time k
$\varphi'(k,j)$	jth value of the approximate trellis branch APP computed by the forward processor of the BCJR algorithm at time k
$\breve{\varphi}(k,j)$	jth value of the symbol APP at time k
$\mathbf{\Psi}_k$	conditional covariance between the CIR at time $k-1$ and the CIR at time k given the received signal up to time K (for the complex bandpass signal model)
$\mathbf{\Psi}_{k\|k}$	conditional covariance between the CIR at time $k-1$ and the CIR at time k given the received signal up to time k (for the complex bandpass signal model)
$\boldsymbol{\psi}_{l,k}$	lth column of $\mathbf{\Psi}_k$
$\boldsymbol{\psi}_{l,k\|k}$	lth column of $\mathbf{\Psi}_{k\|k}$
$\bar{\mathbf{\Psi}}_k$	conditional covariance between the CIR at time $k-1$ and the CIR at time k given the received signal up to time K (for the low-pass signal model)
$\bar{\mathbf{\Psi}}_{k\|k}$	conditional covariance between the CIR at time $k-1$ and the CIR at time k given the received signal up to time k (for the low-pass signal model)
$\omega_{m,n,l}(k)$	frequency instability function related to the l-th transmission path between antennas n and m
$\omega_l(k)$	frequency instability function related to the l-th transmission path in a single antenna system
$\omega_{m,n}(k)$	frequency instability function related to the carrier between antennas n

	and m (equal for all channel taps)
$\mathbf{\Omega}_{m,n}$	diagonal matrix of complex phasors due to carrier frequency instabilities for the subchannel between the mth receiver antenna and the nth transmitter antenna ($(K+L) \times (K+L)$)
$\mathbf{\Omega}(k)$	diagonal matrix of complex phasors due to frequency instabilities at time instance k ($MN(L+1) \times MN(L+1)$)
$\bar{\mathbf{\Omega}}(k)$	diagonal matrix of complex phasors padded with a vector of ones at time instance k (in a single antenna system) ($p(L+1) \times p(L+1)$)
$\mathbf{\Omega}_m(k)$	diagonal matrix of complex phasors due to frequency instabilities at the mth receiver antenna at time instance k ($N(L+1) \times N(L+1)$)
$\mathbf{\Omega}_{m,n}(k)$	diagonal matrix of complex phasors due to frequency instabilities for the subchannel between the mth receiver antenna and the nth transmitter antenna at time instance k ($(L+1) \times (L+1)$)
Ω_D	Doppler angular frequency
3GPP	Third Generation Partnership Project
A-SISO	adaptive soft-input soft-output
ABW	adaptive Baum-Welch
APP	*a posteriori* probability
AR	autoregressive
ARMA	autoregressive moving average
AWGN	additive white Gaussian noise
BB	burst builder
BCJR	Bahl, Cocke, Jelinek, Raviv
BD	block deinterleaver
BEM	Bayesian expectation maximization
BER	bit error rate
BI	block interleaver
BMFA	Bayesian mean field approximation
BP	belief propagation
BPSK	binary phase shift keying
BWA	Baum-Welch algorithm
CA	code-aided
CC	convolutional code
CDMA	code division multiple access
CE	channel encoder

CFO	carrier frequency offset
CIR	channel impulse response
CP	cyclic prefix
CPM	continuous phase modulation
CRLB	Cramer-Rao lower bound
CSI	channel state information
CT	continuous-time
DA	data-aided
DPLL	digital phase locked loop
EM	expectation maximization
EP	expectation propagation
EKS	extended Kalman smoother
EXIT	extrinsic information transfer
FDE	frequency domain equalizer
FE	front-end
FED	frequency error detector
FER	frame error rate
FF	flat-fading
FFT	fast Fourier transform
FG	factor graph
FI	fixed interval
FL	fixed lag
FS	frequency selective
FSM	finite state machine
HDD	hard decision-directed
HMM	hidden Markov model
ISI	intersymbol interference
KL	Kullback Leibler
KF	Kalman filter
KS	Kalman smoother
LCMF	linear channel matched filter
LDPC	low-density parity-check
LFG	loopy factor graph
LMMSE	linear minimum mean square error
LMS	least mean square

LPF	low-pass filter
MA	moving average
MAP	maximum *a posteriori*
MAPSD	maximum *a posteriori* symbol detector
MCE	minimum cross-entropy
MCRLB	modified Cramer-Rao lower bound
MF	mean field
MFA	mean field annealing
MF-EM	mean field expectation maximization
MFB	matched filter bound
MIMO	multiple-input multiple-output
ML	maximum likelihood
MLSD	maximum likelihood sequence detector
MM-AHE	Morelli and Mengali ah-hoc estimator
MMSE	minimum mean square error
MMV	Morelli, Mengali and Vitetta (estimator)
M-QAM	M-level quadrature amplitude modulation
MSE	mean square error
NCO	numerically controlled oscillator
NDA	non-data-aided
OFDM	orthogonal frequency-division multiplexing
OSA	optimal soft-output algorithm
pdf	probability density function
PLL	phase-locked loop
PO	pilot only
PS	product-sum
PSAE	pilot symbol aided estimator
PSK	phase shift keying
PSP	per-survivor processing
QPSK	quadrature phase shift keying
RLS	recursive least square
RC	reduced complexity
RC-AR	reduced complexity autoregressive
RCKF	reduced complexity Kalman filter
RCKS	reduced complexity Kalman smoother

RM	resource management
rms	root mean square
SAGE	space alternating generalized EM algorithm
SC	single carrier
SDD	soft decision-directed
SIC	soft interference canceller
SISO	soft-input soft-output
SNR	signal-to-noise ratio
SP	sum-product
ST	space-time
TFDE	frequency domain turbo equalizer
uKF	unweighted Kuo and Fitz (estimator)
uMMV	unweighted Morelli, Mengali and Vitetta (estimator)
VA	Viterbi algorithm
VB	variational Bayesian
VBE	variational Bayesian E-step
VBM	variational Bayesian M-step
VCO	voltage controlled oscillator
VD	Viterbi detector
VE-LMS	Viterbi estimator combined with least mean square channel estimator
VFEM	variational free energy minimization
VMP	variational message passing
WMF	whitened matched filter
WSS	wide sense stationary
WSSUS	wide sense stationary uncorrelated scattering
ZF	zero forcing
$\mathbf{0}_{l\times k}$	zero matrix or vector ($l \times k$)
$\mathbf{0}_L$	square zero matrix ($L \times L$)
$\arg$	argument
$\arg(\cdot)$	phase of the argument after reduced to interval $[0\ 2\pi[$
$E[\cdot]$	expectation operator
$E[a\|b]$	expectation of a given b
$E_Q[\cdot]$	expectation under the distribution Q
$\hat{\mathbf{f}}$	estimate of the vector $\mathbf{f}$
$\hat{\mathbf{f}}(k\|t)$	estimate of the time-varying vector process $\mathbf{f}(k)$ or time-invariant vector

	$\mathbf{f}$ at time k given the received signal up to time t (moreover, $\hat{\mathbf{f}}(k\|k) \equiv \hat{\mathbf{f}}(k)$)
$(\cdot)^*$	complex conjugate of the argument
$(\cdot)^T$	transpose of the argument
$(\cdot)^H$	conjugate transpose of the argument
$\|\cdot\|$	Euclidean vector norm
$[\mathbf{A}]_{l,k}$	element at the l-th row and kth column of matrix $\mathbf{A}$
$[\mathbf{a}]_l$	l-th element of vector $\mathbf{a}$
$\otimes$	Kronecker product
$*$	convolution for time-continuous signals
diag$(\cdots)$	(with multiple arguments) diagonal or block diagonal matrix with the arguments on the main diagonal or (with single vector argument) diagonal matrix with the vector elements on the main diagonal
max	maximum
$\Im(\cdot)$	imaginary part of the argument
$\Re(\cdot)$	real part of the argument
sign$(\cdot)$	sign of the argument
tr$(\cdot)$	trace of the matrix in the argument

Contents

1 Introduction

Over the last decades, we have witnessed an explosive growth in cellular telephony, portable computing, and, most recently, the mobile Internet. These trends suggest a convergence of mobility and different data applications and, thus, a strong potential for broadband wireless data communication. The ever increasing demand for new data services together with device cost factors, network coverage issues, and a constantly growing customer base set major challenges for system or network designers. The spectrum efficiency of the network can be increased in a number of ways, including designs for efficient resource management (RM), advanced radio access methods, various interference mitigation methods (e.g., using RM and/or various smart antenna concepts), joint optimization of transmitter and receiver (e.g., link adaptation), and advanced receiver technology. The focus of this thesis is on the advanced receiver technology and, in particular, iterative turbo-processing receivers and their adaptation to varying channel conditions.

The increased demand for high data rate wireless communications will pose great challenges for the system design to achieve high-throughput wireless transmission in radio channels where the signals are corrupted by multipath propagation. The multipath dispersion causes intersymbol interference (ISI) where the number of interfering symbols grows linearly with the data rate. Orthogonal frequency-division multiplexing (OFDM) has been proposed as an effective anti-multipath technique, primarily because it offers favorable trade-offs between performance and signal processing complexity in severe multipath environments [44]. A strong competitor to the OFDM technique is a cyclic prefix-based single carrier (CP-SC) transmission technique which enables similar frequency domain equalization as the OFDM. Moreover, it does not possess the property of a high peak-to-average power ratio, which is the major drawback of the OFDM technique when considered for personal radio communications [69]. Multiple transmitter and receiver antennas are also likely to be a viable solution in the future broadband wireless systems, as the bandwidth efficiency of such a multiple-input multiple-output (MIMO) channel is expected to increase enormously in a rich-scattering environment [77].

In this thesis, a single carrier system with or without cyclic prefix is assumed as a system model. The same system model with slightly varying details is used through-

out the thesis for the sake of consistence and unity. However, many of the concepts and algorithms developed in the thesis may be systematically extended to other system models, including the OFDM and the code-division multiple access (CDMA) methods, as well.

The main aim of this thesis is to develop likelihood-based joint detection and estimation algorithms for generic single carrier digital communications systems. The likelihood-based methods have a number of favorable features, particularly their uniform treatment of uncertainty at all levels of detection and estimation processes. The formalism also allows ready incorporation of prior knowledge into detection and estimation processes as well as the seamless combination of such knowledge with received data. The elegance of the likelihood-based processing, however, often comes at a substantial computational cost. For example, posterior distributions for data symbols resulting from the incorporation of the received samples from the channel and prior information from the decoder must be represented and updated, and this generally involves high-dimensional summations or integrations.

Exact computation of posterior distributions, i.e., exact inference, may not be a viable solution even in relatively simple settings, such as transmission of data symbols over the additive white Gaussian noise (AWGN) channel in the presence of a single unknown parameter. Therefore, the deterministic approximation methods, generally known as *variational* methods, have widely been applied for inference and estimation in various applications. For example, they have been extensively used in statistical physics [198] and statistics [216]. In addition, these methods have also found many applications in machine learning and artificial intelligence areas, particularly in the context of graphical models [125, 124, 264]. In this thesis, variational methods are applied to obtain feasible approximations to optimal receivers in digital communications.[1] In particular, many of the existing receiver structures are rigorously justified through the variational bounding technique, and plenty of new ones are suggested.

The graphical models cited above facilitate understanding of complicated problems at hand, serving as a representational platforms for exact or approximate probabilistic calculations in the complex multivariate systems. In a fascinating way, they bring together graph theory and probability theory in a powerful formalism, providing a natural

[1] An optimal receiver is, in general, understood as a receiver that minimizes a bit error rate (BER) or a symbol error rate (SER). However, in the presence of nonstochastic parametric uncertainty, the problem of optimal detection defined in this strict sense may be ill-posed. Hence, another criterion for receiver optimization is needed in that case.

framework for formulating *variations* on the classical statistical signal processing architectures, as well as for exploring entirely new families of statistical models. In fact, algorithms for computing basic statistical quantities such as likelihoods and marginal probabilities have long been expressed in terms of recursions operating on associated graphs. Examples include Markov models, Markov random fields, the forward-backward processing algorithms [207, 199], as well as Kalman smoothing and the iterative turbo decoding algorithms [137, 136, 159].

The basic intuition underlying variational methods is that complex systems can in fact be probabilistically simple. For example, in systems characterized by densely connected graphs, a central limit theorem may come into play, rendering the nodes relatively insensitive to particular settings of values of their neighbors. Taking advantage of these averaging phenomena can lead to simple yet accurate approximate inference algorithms. Particularly the mean field inference algorithms have proven to be useful in providing practical solutions to many complex statistical inference and estimation problems [217].

The rest of this chapter is organized in the following way. First, the receiver techniques for known and unknown channels are reviewed. Since excellent reviews on the topic already exist in the open literature [232, 157], only the major steps in the development of the receiver techniques are pointed out here. Then, the main contributions of the thesis are explicitly stated and the organization for the rest of the thesis is outlined.

1.1 Receivers for known channels

In this section, a short historical survey of optimal receivers in known channels is provided first, followed by a survey of more recent developments of powerful iterative turbo-processing receivers for coded signals in known channels. Major development steps in these areas are pointed out, emphasizing theoretical and conceptual aspects of receiver techniques more than applications.

1.1.1 Historical perspectives on optimal receivers

The start of the development of the optimal receivers for digital communications can be traced back to the formulation of optimal detection algorithms for one-shot receivers in the additive white Gaussian noise (AWGN) channel, based on various optimization criteria. Perhaps the most celebrated of these receivers is the one which aims to maxi-

mize the *a posteriori* probability of the transmitted data symbol, generally referred to as the MAP receiver. The MAP receiver minimizes a symbol error rate, and that is why it has been so popular over the years. An essential part of the MAP receiver is a matched filter, which forms a set of sufficient statistics for symbol detection [204]. In fact, the MAP receiver in the AWGN channel cross-correlates the received signal with every possible transmitted signal. Therefore, the receiver is composed of a correlator or, alternatively, a matched filter bank followed by a decision devise. Moreover, for signals with unequal energies, bias terms have to be added to the output of each matched filter. Early developments of these one-shot receivers are summarized in [127].

The optimal one-shot MAP receiver in a known frequency-selective channel is conceptually a trivial modification of the MAP receiver in the AWGN channel. In this case, the received signal is cross-correlated with every possible transmitted and channel filtered signal [127]. Importantly, these same detection principles can be directly applied also to modern sequence detection receivers, including the MAP sequence and symbol detectors for ISI channels. The MAP sequence detector is equivalent to the more commonly used maximum likelihood sequence detector (MLSD) for equiprobable data sequences. The MLSD receiver is essentially a one-shot receiver for the whole transmitted data sequence and, hence, in the known frequency selective channel, it can be implemented with the matched filter bank, which correlates the received signal with every possible transmitted and channel distorted symbol sequence [155].

For brevity, the MAP symbol detector is hereafter referred to as the MAPSD. Similarly to MLSD receivers, the MAPSD can also be formulated in terms of *sequence metrics* which are further processed by exposing them to an exponential nonlinearity and then to a multidimensional summation prior to making symbol decisions. On the other hand, by using the causality and finite-memory properties associated with the ISI channel, both MLSD and MAPSD can alternatively be formulated in terms of the classical *branch metrics* [74].

A recursive version of the MLSD receiver, in a known time-invariant frequency-selective channel, was first described in [76]. The receiver includes a cascade of a matched filter and a noise-whitening filter prior to a Viterbi algorithm (VA). This kind of whitened matched filter, when sampled at the symbol rate, was shown to form a set of sufficient statistics to detect the sequence of data symbols [76]. A new formulation of the MLSD receiver, avoiding the need for an explicit noise-whitening filter, was presented in [241]. Comparison and some unification of different versions of the MLSD receiver was presented in [34].

Further extension of the MLSD receiver to known time- and frequency-selective channels has lately been proposed in [101]. Particularly in the fast fading channels, the Doppler diversity provided by the fading channel was shown to improve the performance of the MLSD receiver. Similarly, a joint multipath-Doppler diversity gain has been shown to be attainable by the use of spread spectrum signaling over known frequency-selective time-varying channels [32, 220]. Specifically, using the canonical channel decomposition technique, optimal and approximate "one-shot" receiver structures for doubly-spread channels were derived in [32] and in [220], respectively.[2]

There are in the literature at least three different recursive or sequential algorithms for the calculation of the MAPSD in a known frequency-selective channel. First, a fixed-delay MAPSD algorithm which updates the decision metrics of the trellis using only forward recursions was proposed as early as in 1970 by Abend and Fritchman [2]. The computational complexity of the algorithm, however, grows exponentially with a decision delay. As a remedy, a reduced complexity version of this algorithm with parallel structures and some simplifications was proposed in [68]. Second, Chang and Hancock derived a sequential block-based MAPSD algorithm, which uses both forward and backward recursions, in the 1960's [37]. Along with this invention, implementation for long sequences became more feasible, since the complexity of this algorithm grows only linearly with the number of the symbols in the block. In fact, an essentially identical algorithm was independently proposed in [24] and, ever since, has been extensively used in processing of hidden Markov models (HMM) [206]. Another similar type of algorithm was applied to the decoding of linear convolutional codes in [16]. Recently, this forward-backward processing MAPSD algorithm has often been referred to as the BCJR algorithm, according to authors of [16]. Third, a recursive version of the MAPSD algorithm, which was called an optimal soft-output algorithm (OSA), was proposed in [144]. It is a recursive fixed-delay algorithm like the detector by Abend and Fritchman, but its complexity grows only linearly with the decision delay. The OSA algorithm in [144] was derived for linearly modulated signals but has later been extended to the continuous phase modulated (CPM) signals in [18].

Optimal decoding of coded signals is also based on the principle of cross-correlation [170]. Again, the received signal is correlated with every possible transmitted signal, which at this time must fulfill the constraints imposed by the channel code. When multiple component codes, possibly combined into a single coding scheme, are imposed

[2] One-shot receivers were obtained by assuming ideal autocorrelation properties of pseudorandom codes.

on the transmitted signal, which is then exposed to a known frequency-selective channel, the cross-correlation at the receiver has to be performed over the signal set which fulfills all code and channel constraints simultaneously.

Under the usual assumption of the received signal having finite memory in time, the channel can be modeled as a finite state machine (FSM) and, hence, be described by a trellis diagram. Also, if an encoder or a chain of encoders at the transmitter has a FSM description, the behavior of the demodulator and the decoder can be described jointly with a single composite trellis. Thus, an optimal sequence detector in such cases is obtained simply by the VA operating on the composite trellis. Unfortunately, in many practical cases, the number of states in the composite trellis becomes very large, and, consequently, the implementation complexity becomes prohibitive. This is particularly true if interleaving is used in between the encoder and the modulator at the transmitter. Moher and Gulliver [170] proposed an iterative minimum cross-entropy (MCE) decoding algorithm to reduce the complexity of direct decoding. The iterative MCE decoder essentially projects the estimated bit sequence distribution alternatively onto the subspaces representing those distributions which satisfy one or more of the constraints imposed by the respective FSM blocks. Unfortunately, the complexity of the MCE decoder is still too high for many practical applications. Therefore, computationally more efficient decoders, based on certain independence assumptions, were proposed. In effect, these suboptimal decoders were shown to be conceptually similar to the turbo-decoders which were invented and first introduced by Berrou *et al.* [28] on a more ad hoc basis.

1.1.2 Iterative turbo-processing receivers

Early experiments with the turbo codes, constructed by arranging two simple convolutional codes in parallel and permutating the input bit sequence to the second one, provided a performance surprisingly close to the channel capacity in the AWGN channel [28] when decoded by an iterative decoding algorithm. Ever since, there has been a huge upsurge of research interest in original turbo and related codes and their decoding algorithms, in theoretical analysis as well as in the applications domain. To the great astonishment of many researchers and practitioners, the turbo-principle as applied to decoding was shown to render the decoding of complex composite codes feasible while, at the same time, retaining a good performance. Yet more importantly, the applicability of the turbo-principle is not limited to the decoding of composite codes. It can be

applied and, in fact, has been applied to a multitude of tasks gaining from the iterative likelihood-based processing in one way or another. Iterative receivers, where different subblocks like soft-input soft-output (SISO) demodulator/detector, SISO decoder, and parameter estimators are iteratively interconnected by using turbo-processing, are often called turbo receivers. Indeed, a typical paradigm for the turbo receivers is the interconnection of special SISO modules which exchange probabilistic information with each other, most commonly in the form of *extrinsic* information [50].

In regard to the theoretical aspects of the turbo codes, at the time they were invented very little was understood about convergence properties and optimality of the associated iterative decoding algorithm. Since then, however, many approaches have been taken to shed more light onto these issues. These include analysis based on the nonlinear dynamics of the iterative decoding systems [134, 3, 63, 235], encompassing density evolution [63, 229], and extrinsic information transfer (EXIT)-chart-based [235] approaches as special cases. On the other hand, a analysis method that is based on the signal-space characterization of the iterative decoding algorithm [81] has been shown to be particularly useful for analyzing the decoding of finite-length codes. The properties of turbo and low-density parity-check (LDPC) codes were examined by using an information geometrical framework in [116].

Furthermore, various analysis methods that are associated with the graphical modeling techniques have turned out to be useful in exploring the properties of the iterative decoding algorithms [4, 137, 136, 159, 259]. For instance, the standard turbo decoding algorithm has been shown to be an instance of the generic belief propagation (BP) algorithm, operating on the graphical model of the code, whereas the BP algorithm has been linked to the minimization of the statistical physics functional called Bethe free energy [265, 110]. Further insight into BP and its generalizations has been obtained via the tree-based reparametrization framework [250].

The graphical modeling of the turbo code, or any turbo-processing receiver for that matter, is based on the observation that the global objective function associated with the data symbols or bits can be factorized into the product of conditional probability density functions. In the graph, the conditional independencies between the symbols are encoded by the missing edges between the corresponding nodes. Indeed, the turbo-processing algorithms have been shown to be specific instances of more generic message passing algorithms, variants of which are called generalized distributive law [4], sum-product (SP) algorithms [137] or, BP algorithms. These operate on the "junction graphs," the "factor graphs," and belief networks, respectively, all providing a graphical

representation of the associated function factorization. In this thesis, the terms SP and BP are used interchangeably.

As is well known, the message-passing algorithms realize an exact inference only in the graphs which are cycle-free. However, the graphs representing the turbo-receivers unfortunately include cycles and, therefore, in contrast to the iterative MCE algorithms [170], there is no notion of "termination," since the messages can travel around the cycles indefinitely. There is, nevertheless, strong experimental evidence suggesting that the turbo-processing algorithms in many cases attain a very close approximation of the desired objective function. Moreover, it has been proven recently that any stationary point of the SP algorithm operating on the loopy factor graph (LFG) can be interpreted as a solution to the constrained free energy minimization problem [265, 110]—a result which provides theoretical evidence for the fact that the turbo-processing algorithms converge into something which makes sense.

In regard to the applications of turbo processing, the turbo receivers tuned up for combating against the detrimental effects of ISI have widely been applied in different forms in recent years. Generally speaking, in all of these schemes the equalization/detection and decoding tasks are iteratively combined on the same set of received data so that feedback information from the decoder is incorporated in one form or another into the equalization/detection process. Since the frequency-selective channel can be viewed as a rate-1 encoder which maps the discrete input alphabet into continuous-valued complex numbers, the channel encoder at the transmitter concatenated with such a channel can, from the receiver point of view, be considered a concatenated coding system. Consequently, the turbo-processing can readily be applied [65]. In its standard form, the SISO demodulator and the SISO decoder computes the symbol or bit *a posteriori* probabilities (APPs) exactly while the information transfer between the blocks takes place in the form of *extrinsic* information.

Regarding the SISO demodulator, exact computing of symbol APPs may, however, in some cases become computationally infeasible. Difficulties due to high complexity arise particularly in the wideband MIMO channels and in the code division multiple access (CDMA) channels. Therefore, various approaches to reduce the computational complexity of SISO demodulation in both single antenna and multiple antenna channels have appeared in recent years. Without even trying to be exhaustive, we refer to [140, 238, 211, 51, 139, 79, 23, 221, 1, 226, 64, 98, 112] and references therein. Specifically, the prefiltering of the received signal before APP computation was proposed as a method for complexity reduction in [23], whereas in [79] the prefilter was tailored

for the M-BCJR algorithm. In [51], it was shown that the SP algorithm operating on the LFG, where the data symbols form a set of variable nodes, converges to a solution which provides a good approximation for the exact APPs as long as the factor graph has girth of at least 6. In [139], a similar approach was taken. Since the girth of the LFG of the ISI channel is typically less than 6, a stretching technique was proposed in [51] as a feasible method to obtain an equivalent girth-6 graph. In principle, the stretching technique can be used to transform an arbitrary factor graph to an equivalent cycle-free factor graph for which the SP algorithm produces exact APPs [137]. Of course, the complexity of the SP algorithm operating on the transformed cycle-free factor graph is considerably higher than in the original case.

Furthermore, linear SISO equalizers were proposed for single antenna systems in [140, 255, 238, 211, 253] and for the multiple-input multiple-output (MIMO) channels in [221, 1].[3] They consist of forward and feedback finite impulse response filters which are jointly optimized by using the minimum mean square error (MMSE) criterion. In [140], the filter coefficients were computed under the assumption that perfect *a priori* information on transmitted symbols is available, whereas in other cases the estimated *a priori* information obtained from the decoder is taken into account in the computation of the filter coefficients. Just recently, an interpretation of these devices as instances of the SP algorithm operating on the suitably defined factor-graphs was given in [66].

Interestingly, the linear SISO detectors actually trace their history back to the work of Taylor in the early 1970's [233]. Specifically, an estimate feedback equalizer, as it was called in [233], consists of essentially the same structural elements as the modern block-processing linear SISO detectors. However, judged as being unrealizable, it was replaced by a recursive MMSE decision feedback equalizer in practical receivers for decades.

Other techniques to reduce the computational complexity of the turbo receivers include the list sphere decoding [112], the semi-definite relaxation of the maximum-likelihood (ML) detection problem [226], or sequential Monte Carlo methodology [64, 98]. In CDMA systems, the idea of combining the multiuser detection process iteratively with the channel decoding process has sparked off substantial research work.

[3] Although terms such as SISO equalization and turbo equalization are commonly used in the literature, it may, however, be more appropriate to talk about SISO detection. Equalization refers to the techniques used to cancel or reduce ISI at the receiver in order to allow memoryless or reduced-memory detection. In this respect, the turbo receivers are rather dealing with SISO *detection* than *equalization*. Hence, the following terminological agreement is made: the wording "linear SISO detection/detector" is used later in this thesis when referring to the linear SISO devices that were previously known as SISO equalizers or turbo equalizers.

As a result, a wide range of turbo multiuser receivers have emerged in recent years, some of which are given in [169, 210, 5, 255, 67, 91, 92] and references therein. A unifying framework, based on the factor graph representation of a multiuser channel and the SP algorithm, was provided in [35], enabling an asymptotic performance analysis of many existing receivers via the density evolution technique.

1.2 Receivers for channels with parametric uncertainty

In the presence of parametric uncertainty, detection must be adapted according to either implicitly or explicitly estimated parameter values. Optimal receivers, as well as most of their recursive and iterative approximations, apply parameter estimation implicitly, so that the estimation is optimized for detection [102]. More practical adaptive receivers can be obtained by carrying out estimation and detection either jointly or separately. Typically, estimation can be separated from detection when the parameter values change slowly in time. In that case, the parameter estimation can be done either from the known data segment using pilot symbols or blindly, i.e., in the non-data-aided fashion. The estimated parameter values are then exploited in detection as if they were true parameter values.

In this section, a short overview, also encompassing some historical notes, of the optimal receivers and receivers which are closely related to the optimal receivers is given first. Then, adaptive receiver techniques for unknown ISI as well as for phase uncertain channels are discussed, emphasizing the estimation methods for SISO blocks. Finally, estimation methods for fading channel statistics and frequency offset are briefly reviewed.

1.2.1 On optimal receivers and their approximations

An optimal receiver in the presence of parametric uncertainties, where the optimality of the receiver is defined in the sense of providing *exact* MAP symbol decisions, exists only if the unknown parameters are stochastic and the associated probability distribution functions are known. Otherwise the optimal receiver is not defined because unknown parameters cannot be averaged out from the decision metrics, i.e., the problem of optimal detection is ill-posed in the presence of nonstochastic uncertainty. However, as will be discussed later, some related detection criterions can be applied in order to derive practical receivers also in these channels. In most cases, optimality implies high com-

plexity which is due to an infinite memory effect induced by the unknown parameters [42].

The optimal MLSD and MAPSD receivers in the presence of random parameters maximize the conditional likelihood function of data averaged over all random parameters [244]. The optimal MAPSD receiver in the presence of uniformly distributed carrier phase and timing phase was studied in the late 1970's [70]. In particular, the MAP symbol detector was formulated as a maximum of the averaged likelihood function, but a closed-form solution to the maximum was not found. In general, the MAP detection algorithm in the presence of random parameters can be obtained in closed-form only in some special cases. Perhaps the most important of these special cases is formed by linear Gaussian disturbances, which are ubiquitous and which typically emerge as complex Gaussian distortion, such as Rayleigh or Ricean fading channels.

Thus, in the presence of a Rayleigh fading channel, the optimal MLSD and MAPSD receivers are obtained by averaging the conditional likelihood function over the complex Gaussian channel impulse response (CIR), assuming that the mean and covariance matrix of the Gaussian fading process are known. Early developments of these fading channel receivers trace their history back to the work of Kailath [127], who derived an optimal one-shot MAP detector in the Rayleigh fast fading channel. The detector, which is also often called the quadratic receiver, has an estimator-correlator structure, where a minimum mean-square error (MMSE) estimation of the channel impulse response is performed for every possible transmitted signal. In effect, the received signal is cross-correlated with the estimated received signal for each possible transmitted data signal. Importantly, Kailath's estimator-correlator receiver structure can be readily generalized into the modern data sequence detectors. The optimal sequence detectors on the Rayleigh fading channel have to calculate an MMSE estimate of the received signal for every possible transmitted symbol sequence—a fact that was explicitly stated for the MLSD in [42] and for the MAPSD in [222].

While the optimal ML detection poses no difficulty in known channels, the channel-induced memory in the received signal samples causes remarkable computational challenges in the presence of unknown channels. In fact, no fixed complexity receiver is optimal under these conditions [42]. Nevertheless, there have been many attempts to fold the optimal quadratic receiver into a fixed-size trellis structure by force. One of those was presented in [173]. In effect, the general likelihood-ratio formula by Kailath [128] was combined with the Viterbi algorithm (VA) when assuming that the received signal has a finite memory. This recursive MLSD receiver in the complex Gaussian

fading channel was derived by using analog signal processing. Therefore, contrary to what was claimed, the VA does not apply in this case, as was later shown in [42]. Similar type of receivers were later derived for sampled receivers in [150, 266, 246, 247]. Specifically, a force-folded MLSD receiver in the Rayleigh fading channel for nonlinearly modulated CPM signals was presented in [150]. It was further generalized to any modulation format and to diversity systems by using the concept of an innovation process in [266]. Interestingly, almost identical MLSD receivers were also derived in [246, 247], where the trellis structure of the associated Viterbi algorithm was called a super-trellis. However, as shown in [105], a forced folding of the MLSD receiver into the fixed-size trellis may in fast fading channels cause an error floor where the receiver's BER levels out at a high signal-to-noise ratio. Also, three other sources of error floors were identified in [105].

Optimal MLSD receivers and their approximations were considered for the frequency-flat and frequency-selective Rayleigh fading channels in [130, 99] and in [55, 117], respectively, by applying Kalman filters to channel estimation. As may be readily perceived, the per-sequence Kalman filters in the MLSD results from the assumption that the fading statistics can be faithfully described by the autoregressive (AR) channel model and that the AR parameters are known. As explained in [99], the Kalman filter replaces the traditional phase-locked loop (PLL) for carrier recovery in frequency-flat Rayleigh fading channels. Furthermore, the symbol timing was estimated along with the channel impulse response (CIR) in [117] by replacing the bank of Kalman filters with the bank of extended Kalman filters (EKF) in the computation of the decision metrics of the MLSD. Due to the nonlinear nature of the joint estimation of symbol timing and CIR, this receiver is inherently suboptimal but, nevertheless, gives a fairly good approximation of the optimal receiver when the signal-to-noise ratio (SNR) is high enough.

An iterative expectation-maximization (EM)-based MLSD receiver in the presence of random Gaussian parameters was proposed in [87]. In particular, the implementation of MLSD receivers was studied in the presence of the Gaussian noise channel with uniformly distributed time-invariant phase-shift and in the presence of the Rayleigh fading channel. The EM algorithm was shown to provide an efficient methodology for close to optimal receiver implementation in both channel models. Since convergence of the EM algorithm into the global maximum of the likelihood surface, a prerequisite for optimal detection, cannot be guaranteed, the EM-based receivers should in general be regarded as suboptimal receivers. Moreover, an iterative EM-based MLSD receiver

in the presence of a random timing offset with uniform distribution was considered in [88]. While an exact EM-based implementation was found to be infeasible, a high SNR approximation of the receiver was shown to lead to the use of maximum likelihood (ML) timing estimators. In fact, a similar approximation was made earlier in [70] for direct MLSD receivers in the presence of an unknown timing offset.

The optimal MAPSD receiver in a frequency-flat Rayleigh fading channel was considered in [222], but in order to avoid exponential growth in complexity, a recursive fixed-lag MAPSD, conceptually similar to the receiver by Abend and Fritchman, was proposed as a suboptimal approach. Equivalent MAPSD receiver structures were later derived for nonlinearly modulated CPM signals in [17]. Furthermore, a sequential MAPSD based on a forward-backward processing algorithm was derived for the frequency-flat Rayleigh fading channels in [89]. Trellis structure for the MAPSD was again obtained by restricting the dependency of the received signal samples on past observations into a finite number of observations. These MAPSD receivers were extended to the frequency-selective Rayleigh channels and analyzed in an unified manner by using the concept of the innovations process in [103]. Yet another interesting approach to the formulation of the optimal MAPSD receiver was proposed in [8], where an exact expression for the symbol APPs in the presence of an unknown frequency-selective Rayleigh fading channel was derived in an elegant way that enabled the decoupling of the complexity and observation length. Then, based on this decoupling property, a number of computationally efficient suboptimal SISO modules were also derived.

1.2.2 Adaptive receivers for unknown ISI channels

Most of the CIR estimators can be roughly categorized into ML-based estimators or MMSE-based estimators. The recursive versions of the ML-based CIR estimators need to know only the bandwidth of the Doppler spectrum of the fading channel so that the tracking parameters of the estimators can be selected properly. Instead, the MMSE-estimators rely on the assumption that the fading process is Gaussian and its statistics are known. The most common ML-based recursive estimators with built-in channel tracking capabilities are the least mean square (LMS) estimator and the recursive least square (RLS) estimator, whereas the Kalman filter is the optimal recursive MMSE estimator for the fading Gaussian channel, assuming that an AR description of the fading channel is available [106]. Naturally the MMSE-based estimators are expected to perform better than the ML-based estimators in time-varying channels since they exploit

the prior information about the channel dynamics in a more efficient way. Typically, the statistical parameters of the channel are unknown to the receiver, and, hence, a coupling of the AR parameter estimator for fading statistics and the Kalman CIR estimator was proposed in [60].

Although in many cases the initial channel estimate can be obtained with the aid of pilot symbols, further noise averaging as well as tracking of the time-varying channel are often needed during the transmission of the information symbols. Then, detection and estimation has to be done somehow jointly. In addition to optimal detectors that perform channel estimation implicitly together with detection, a popular approach for combining detection and estimation is to perform joint ML detection and estimation. This is done by maximizing the joint likelihood function over the data symbols and the channel parameters [43]. In frequency-selective time-varying channels, significant complexity reduction can be obtained by folding the tree-search detection algorithm into a fixed size trellis and applying, e.g., a PSP-technique for channel estimation [42, 209]. A related receiver structure is given in [269]. Iterative versions of the joint ML detection and estimation algorithms, obtained via the EM algorithm, are discussed in [86, 87, 71]. Moreover, an excellent review and a novel classification of different channel estimation methods and of different joint detection and estimation schemes is presented in [157]. Another excellent review of optimal and joint estimation and detection schemes is given in [232].

The channel parameters can also be estimated in a non-data-aided (NDA) manner. From many different blind estimation methods reported in the literature [240, 236], only the probabilistic methods, i.e., the blind methods which use likelihood processing, are of interest in this thesis, and, therefore, only these methods are further surveyed in the sequel. A probabilistic approach was used in [129], where an iterative algorithm for blind ML channel estimation and MAP symbol detection via the Baum-Welch algorithm was proposed. The same receiver structure was also used in [45],where an approximate Cramer-Rao lower bound (CRLB) was derived for the parameter estimator of the Baum-Welch algorithm and was then used to find an approximate expression for the probability of symbol error.

The Baum-Welch algorithm is applicable only when the channel is fixed during the transmission of the information symbols. In order to avoid this shortcoming, a recursive approximation of the Baum-Welch algorithm with channel tracking capability was proposed in [13]. Furthermore, the NDA MMSE estimate of the frequency-flat Rayleigh fading channel was obtained jointly with MAP symbol decisions via the Bayesian EM

(BEM) algorithm in [38]. A conceptually similar scheme was proposed in [262]. Specifically, the BEM algorithm was applied to the dynamic system model, describing the data transmission over the flat-fading channel, and it was then shown to yield a soft decision-directed (SDD) Kalman smoother interconnected with the symbol APP computation. On the other hand, an SDD Kalman smoother for the generic vector dynamic system model was proposed in [152], but the complexity of this scheme is very high. In [14, 15], the properties of Martingale difference sequences were exploited in order to derive nonlinear Kalman-like recursive channel estimators which were also able to take advantage of the probabilistic information about the data symbols.

Common to all these probabilistic algorithms is the fact that the unknown channel parameters are estimated by using the *soft* information about the data symbols, instead of the hard symbol decisions. In essence, these estimators can be said to be blind in the sense that they aim to maximize either the averaged likelihood function or the averaged *a posteriori* density function where the averaging in both cases is performed over the unknown data symbols. However, in order to avoid phase ambiguity problems caused by the rotational symmetry of the symbol constellation, the probabilistic estimators must be operated in a *semiblind* manner where some of the transmitted symbols, e.g., pilot symbols, are known by the receiver. Specifically, the phase ambiguity problem is related to channel identifiability. In general, the channel can be made identifiable by imposing some code structure on the transmitted (measurement) signal, while the insertion of pilot symbols into the signal frame structure can be considered a special type of coding.

Turbo receivers also have to somehow deal with typically continuous valued unknown channel parameters. Besides using the standard estimation methods, one would be tempted to do estimation and inference jointly by somehow exploiting the graphical modeling techniques. In particular, extending the traditional graphical modeling concept, the unknown channel parameters could be included in the graph as additional nodes. Then, the generic message passing algorithm could be operated on the resulting graph. The main obstacle in pursuing this attractive idea further, however, is that the descriptions for messages associated with the continuous valued variables require infinitely many parameters. This is because the channel constraints are defined in a continuous domain. In practice, only partial descriptions for these messages are feasible, and, hence, the exploitation of canonical distributions has been suggested in [261] as a generic tool for handling the continuous valued variables in graphical models (see also [151]). In fact, many of the adaptive turbo receivers, which treat the parameter

estimators as separate subblocks in computing the symbol APPs, can be interpreted as message passing algorithms where the messages outgoing from the continuous valued variable nodes are approximated with appropriate canonical distributions. Examples of such adaptive turbo receivers are given in [242, 29, 19].

An alternative method for inference in a hybrid dynamic system with both continuous and discrete variables was obtained via the generalized loopy belief propagation algorithm, called an expectation propagation (EP) algorithm, in [168]. Instead of propagating the exact "beliefs" about the variables of the graph, the EP algorithm retains only expectations of the "beliefs" and iterates until these expectations are consistent throughout the network. As compared to many existing methods, the EP algorithm was shown to provide a solution with improved accuracy in the selected statistical models. Moreover, the EP algorithm has been successfully applied to the adaptive detection in flat-fading channels [205].

Yet another related approach for joint estimation and detection is to embed the parameter estimators inside the SISO demodulators in a more traditional way. In effect, this approach has been adopted in [89, 103, 8, 113, 108]. These adaptive SISOs (A-SISOs) are essentially based on the principle of "parameter-first combining," where the sequence conditioned channel estimation is performed along all or a selected number of transmitted sequences and then the corresponding sequence metrics are combined appropriately to produce the desired soft output. Related A-SISOs using the super-trellis technique and a "sequence-first combining" method have been reported in [61] and [94], respectively.

1.2.3 Adaptive receivers for phase uncertain channels

Data detection in the presence of the AWGN channel distorted by phase uncertainties has been a subject of active research already for decades. Such phase uncertainties are typically caused by the instabilities at the transmitter and receiver oscillators. The recent research interest in this topic has been stimulated by the discovery of the efficient turbo coding schemes as well as the rediscovery of the low-density parity-check (LDPC) coding schemes [83],[4] which both operate at signal-to-noise ratio (SNR) regions close

[4]Indeed, the LDPC codes were invented by Gallager in 1962 [83], but they were forgotton for decades until rediscoved recently. This rediscovery was promoted by the popularity of the turbo codes whose iterative decoding algorithm was noticed to be related to the decoding algorithm of LDPC codes. An interesting subclass of LDPC codes is formed by irregular repeat accumulate codes which can be encoded with an

to the capacity of the AWGN channel. The decoding of the turbo and LDPC codes call for an iterative likelihood-based processing and, thus, an incorporation of the phase estimation into this iterative processing is an intriguing idea. In fact, it has already triggered the development of various iterative methods for joint phase estimation and data detection, with a common aim that both estimation and detection could take full advantage of the code structure.

Existing methods for joint phase estimation and data detection can be divided into Bayesian and non-Bayesian (classical) approaches. By modeling the phase as an unknown deterministic constant, the latter class of receivers computes the symbol APPs based on the *generalized* likelihood processing [188, 9, 154, 189]. In [9], the unknown phase is estimated in a per-survivor fashion on the trellis, whose size is a design parameter and is determined by appropriately pruning the sequence tree associated with the component decoder of the turbo-decoder. The soft output of the component decoder is then computed by using the standard forward-backward processing techniques. Due to restrictions caused by the trellis-based processing, this methodology is readily applicable only to the turbo-codes which are based on the interaction between finite state machines. In contrast, the iterative joint computation of the phase estimate and symbol APPs via the EM algorithm, as first proposed in [188] and later independently in [189], is a more general approach and, hence, applicable to the decoding of any turbo-like codes, including the LDPC codes. An essentially identical algorithm was also derived in [154], without using the EM formalism. Moreover, many of these classical receivers can be made adaptive by applying a "sliding window averaging."

The Bayesian receivers are based on the assumption that the unknown phase is a random variable or a random process with a known *a priori* probability density function [49, 192, 227, 48, 47]. In particular, a noncoherent trellis processing was proposed in [49], where a finite size trellis was obtained by truncating the memory of the channel. Furthermore, widely applicable joint Bayesian estimation and data decoding via the SP algorithm operating on the appropriately defined factor graphs (FG) have been proposed by a number of authors [192, 227, 48, 47, 21, 58, 59]. With the exception of [47], these algorithms execute message passing between detection and estimation by applying the generic approach presented in [261]. In particular, the problems associated with the continuous valued variables in the FG processing are circumvented by enforcing various *canonical* distributions for the continuous valued phase. For example, in [48], a

extremely simple encoder and decoded by a low-complexity iterative decoder, see e.g. [267] for further details.

Tikhonov parametrization of the phase distribution was shown to provide a good trade-off between complexity and performance. In [47], another type of Bayesian approach was taken to the LDPC decoding in the presence of unknown phase—the approach which was based on the memory truncation rather than the explicit representation of the channel parameters in the FG. Moreover, a theoretical framework for iterative synchronizers obtained by combining the SP and EM algorithms has recently been proposed in [109], and an excellent overview of various iterative detection algorithms is provided in [73].

1.2.4 Estimation of unknown frequency offset and fading statistics

An important part of a receiver design is to develop accurate models for the channel, along with advanced estimation methods to estimate the model parameters. But the system performance can also be improved by a more optimal signal design. In particular, various adaptive transmission schemes have received considerable research interest in recent years. In these schemes, the structure of the transmitted signal is made to vary as a function of time and/or frequency according to the prevailing channel conditions. Therefore, accurate channel modeling and advanced estimation methods are of great importance also from the adaptive transmitter point of view.

Many of the transmitter functionalities and parameters, such as channel coding schemes, interleaving schemes, and modulation schemes as well as the network control algorithms like the soft handover algorithm and adaptive power control algorithm, may be adjusted by using the channel's second order statistics rather than making adjustments purely on the basis of instantaneous signal-to-noise ratio values [135]. Under the usual assumption of wide-sense stationary uncorrelated scattering (WSSUS) channel, the second order statistics of the channel can be described by the scattering function (also called a fading spectrum). From the transceiver adaptation point of view, important parameters describing the fading spectrum are the normalized bandwidth of the spectrum (hereafter referred to as the Doppler spread) and the center frequency of the spectrum (hereafter referred to as the frequency offset) as well as the relative power levels of the different multipath components. The frequency offset is the total difference between the transmitted carrier frequency and the local reference frequency at the receiver, and it is cumulatively caused by the Doppler scattering of the communication

channel (especially a Doppler shift) and transmitter and receiver oscillator instabilities. In general, the frequency offset and other channel parameters can be different for different multipath signal components due to different scattering environments or due to nonsynchronous transmitters in systems where space-time diversity is exploited.

Some of the earlier contributions to the estimation of fading channel statistics are reported in [135, 104, 60, 237, 274, 224, 269, 167]. In particular, an iterative joint estimator of synchronization parameters and second order statistics of the Rician fading channel was presented in [104], but the estimator is based on an ad hoc "minimization" scheme and is very complex. In [60] and [237], a multichannel AR description was given to the frequency-selective time-varying channel, and the AR model parameters were estimated by a RLS estimator and by a higher-order statistics-based blind estimator, respectively. In [274], the AR model parameters were estimated along with the complex Gaussian channel process using an extended Kalman filter. The autoregressive moving average (ARMA) model parameters along with the CIR estimates were obtained via the sequential EM algorithm in [269]. Furthermore, an iterative EM-based nonparametric estimation of the scattering function and the auto-covariance matrix of the WSSUS channel were considered in [224] and [167], respectively.

The most important earlier contributions to the block-based estimation of the frequency offset in the flat-fading or frequency-selective fading channels have been reported in [138, 172, 30, 122, 123, 171, 107, 245]. Many of these estimation methods rely either on the assumption of known fading statistics or on some specific pilot signal structure. Moreover, the estimation range is typically either limited or forced to be traded off for the estimation accuracy. Furthermore, a simple frequency offset tracker exploiting the differential Kalman filter was presented in [84]. There are also classical text books [164, 165] which can be used to find more information on the traditional approaches as well as more advanced approaches for frequency offset estimation.

1.3 Aim and contribution of the thesis

In the research work reported in this thesis, the main focus is on iterative turbo-processing receiver algorithms that are explored and further developed by using the variational optimization methods. The variational methods provide a systematic approach for deriving approximations to optimal receivers. The quality of such approximations can be quantified with an *entropic* measure of error, which is inherently available within this methodology. This research is especially aimed at providing new insights and en-

hancements into the existing algorithms and at deriving a set of new variants of iterative turbo-processing receivers. The new receiver algorithms are derived under various assumptions on the available *a priori* information about the properties of the communication medium. As a result of these explorations, this thesis contributes to state of the art of advanced receiver designs in a number of ways. The main contributions are summarized as follows:

1. It is shown that well-known linear SISO detectors, such as the channel matched filter-based linear SISO detector [140] and the linear MMSE SISO detector [238, 239], can be formulated as specific instances of the so-called generic variational free energy minimization (VFEM) algorithm. This novel formulation provides a new set of insights into the design and performance of these known detectors and allows several suggestions of enhancing adjustments on them. This new formulation also shows an interesting link between the linear SISO detectors and the demodulators that are based on exact symbol APP computations. In addition to elucidating that connection, it also conveys a sense of when a particular variational approximation is likely to perform well and when it is bound to fail. These issues are discussed in detail in Chapter 3 of this thesis.
2. The EM algorithm can be interpreted as an approximate inference algorithm in the light of variational optimization, as shown in [176]. This finding is further extended in this thesis by showing that the Bayesian EM algorithm can also be interpreted as a solution to the variational lower bounding. In particular, when this approach is applied to inference in the Rayleigh fading channel, it provides an enlightening link between the Bayesian EM-based adaptive demodulator and the demodulator that computes exact symbol APPs. Imposing further restrictions on the set of distributions over which the variational optimization is performed, the linear SISO detectors combined with SDD channel estimators are shown to ensue. Details of the above findings are presented in Chapter 4 of this thesis.
3. Recursive SDD-LMS and SDD-RLS estimators for unknown deterministic ISI channels and the SDD-Kalman estimator for unknown frequency-selective Rayleigh fading channels are derived. This is based on the recursive version of the EM algorithm and the Bayesian EM algorithm, respectively. The obtained new formulations are shown to yield computationally attractive estimators which are then integrated with the standard forward-backward processing SISO blocks, designed for known channels. In particular, matrix inversions in these SDD estimators are avoided by apply-

ing sequential processing over the *time* and *trellis branch* indices. Futher details can be found in Chapter 4 of this thesis.

4. The problem of iterative detection in the presence of phase uncertainty is considered from the variational Bayesian perspective, bridging the gap between the classical EM-based receivers and the optimal noncoherent receivers. In particular, efficient and computationally attractive iterative joint estimation and detection/decoding schemes for both constant phase models and dynamic phase models are obtained as a result of the variational bounding. Moreover, versions of the proposed adaptive iterative detection schemes exhibiting accelerated convergence are shown to be obtained by applying *incremental* scheduling for the free energy minimization. These issues are discussed in more detail in Chapter 6 of this thesis.
5. Joint estimation of radio channel parameters, such as frequency offset, Doppler spread and, power delay profile, are considered. By modeling the multipath Rayleigh fading channel as a *complex bandpass* autoregressive (AR) process, it is shown how these important parameters can be conveniently extracted from the estimated *complex* AR parameters obtained via the EM algorithm. The main advantage of the proposed new estimation scheme, as opposed to the existing ones, is that it is designed to operate equally well in a wide variety of channel conditions, while there are no restrictions pertaining to the pilot signal structure. This is valid even in extremely high-mobility environments. Furthermore, it is shown that by applying the variational (mean field) bounding technique, the complexity of the AR parameter estimator in the presence of a multipath channel can be significantly reduced, while just slightly trading off performance. Chapter 5 of this thesis is referred for furthur details.
6. A feedforward adaptation scheme, in which estimators for the channel statistical parameters are recursively updated along with the estimator of the CIR, is derived. This derivation applies the stochastic gradient approximation to the score statistics (i.e., gradient of the likelihood function) of the AR parameters. Moreover, based on the complex bandpass AR modeling and the score statistics of the frequency offset, a novel frequency error detector, operating in parallel with the fixed-delay Kalman-smoother for the CIR, is proposed. When incorporated into the feedback frequency recovery scheme, the proposed frequency error detector is shown to provide excellent performance at a low computational cost, being well suited also for very fast fading channel conditions. Both of the aforementioned adaptation schemes can operate in the hard-decision or in the soft-decision directed modes, in addition to the

pilot mode. Detailed derivations of these algorithms are presented in Chapter 5 of this thesis.

7. The Cramer Rao lower bound (CRLB) is well known as a lower bound for the error variance of any unbiased estimator. However, the CRLB is often difficult to obtain analytically when there are nuisance (unwanted) parameters interfering with estimation. The problem of CRLB computation in the presence of nuisance parameters is revisited in two different contexts. First, the CRLB of the frequency offset and receiver noise variance in the presence of a multipath time-varying Rayleigh fading channel is formulated in Chapter 5. The applied analytical method not only allows for feasible computation but also makes it possible to quantify the effect of the fading phenomenon itself and the effect of incomplete knowledge of the fading channel realization on the CRLB separately. Second, computation of the CRLB of the carrier phase in the presence of unknown data symbols is considered in Chapter 6. In this computation, the Fisher information of the carrier phase, expressed in terms of the symbol APPs, is replaced with the empirical Fisher information. Thus, universally applicable numerical approximation for the exact CRLB can be obtained, rendering the evaluation of the CRLB feasible also in cases where information symbols are encoded.

The thesis, in part, is based on 12 original publications of the author [177, 178, 180, 179, 188, 182, 183, 184, 186, 185, 187, 181]. The author of this thesis was the primary contributor of all these papers; he invented and developed the main ideas, performed all numerical simulations, and wrote the papers. Professor Subbarayan Pasupathy provided motivation, criticism, and suggestions related to various technical as well as editorial issues. In the case of [188], Research Professor Aarne Mämmelä also provided criticism and editorial suggestions. For clarity, the thesis is presented as a monograph, and the original publications are not reprinted as such.

1.4 Outline of the thesis

This chapter has so far presented the history and the state of the art of receiver algorithms, especially detection and estimation algorithms and their interactions. The rest of this thesis is organized as follows. Chapter 2 provides a detailed system model, encompassing a parametric description for the frequency-selective fading channel. The variational bounding technique is introduced via four simple examples to demonstrate

its potential in producing sensible approximations. In addition, a VFEM framework is introduced and suggested as an unifying tool for obtaining approximate inference algorithms. In Chapter 3, linear SISO detectors are formulated as instances of the variational bounding algorithms. This chapter, in part, is a reprint of paper [180]. In Chapter 4, the standard EM and Bayesian EM algorithms are shown to fit into a generic framework of variational optimization. Their recursive versions are then exploited in deriving low-complexity SDD channel estimators for frequency-selective fading channels. This chapter, in part, is a reprint of paper [177]. In Chapter 5, EM-based iterative estimation schemes for radio channel parameters, including power delay profile, Doppler spread, and frequency offset, are developed. Also, forward- and backward-processing adaptation schemes for tracking the changes in fading statistics and the frequency offset are presented. This chapter, in part, is a reprint of paper [178]. In Chapter 6, iterative detection schemes for the phase-uncertain channels are discussed, resorting to the EM and variational Bayesian techniques for obtaining approximations for the optimal receiver. This chapter, in part, is based on paper [179]. Finally, Chapter 7 concludes the thesis by providing a summary of the main results and contributions of the thesis. Furthermore, some potential future directions for the research of advanced receiver technology are provided.

2 Preliminaries

This chapter starts with a description of the generic system model, followed by detailed descriptions of the AR and complex bandpass AR-based channel models. Then, a variational bounding technique is introduced with a few illustrative examples providing a framework. The chapter also includes an introduction of the generic variational free energy minimization framework that will be extensively utilized in the upcoming chapters.

2.1 Generic system model

In this thesis, the transmission of linearly modulated information symbols over the radio channel with parametric uncertainty is considered. Therefore, a generic multiple-antenna system model, where the radio channel between each transmitter and receiver antenna pair is assumed to be both frequency- and time-selective, is given next. The system is assumed to employ N transmitter antennas and M receiver antennas. Each transmitter antenna is assumed to transmit independent data. System models with other channel realizations, e.g., the AWGN channel and the block-fading channel, can be obtained as special cases of this generic system model.

In the presence of an unknown time-varying channel, an optimal receiver front-end (FE) processor is somewhat difficult to define. This is due to fact that the received continuous-time (CT) signal is in all practical cases time-limited and, therefore, not strictly band-limited. This renders the sampling of the received CT signal with any sampling rate inherently suboptimal.[5] However, assuming that the received signal is strictly band-limited (for all practical purposes), sampling at the proper rate provides a set of sufficient statistics for data detection and parameter estimation [166] (see also the Appendix A of [73] for a concise overview on discretization by sampling). On the other hand, assuming that the channel evolution in time is slow compared to the signaling interval T_s and that the CT physical channel can be accurately modeled with a tapped delay line model, the fractional sampling (at the proper rate) of the output of the receiver FE-filter, which is matched to the transmitter pulse-shaping filter and combined with

[5] A stationary random CT process may be represented exactly by discrete observables, regardless of the finite length of the observed CT signal, but the number of required observables can approach infinity [100].

the noise-whitening filter, provides a set of sufficient statistics [43]. Such a receiver filter is generally referred to as the whitened matched filter (WMF). Under the further assumption of close to perfect timing synchronization,[6] a good suboptimal FE processor is obtained by symbol rate sampling of the output of the WMF [43].

Therefore, based upon the assumptions made above, a discrete-time system model with only one sample per symbol is employed in this thesis. This will, however, cause no loss of generality in terms of the developed algorithms since they can be easily extended to support fractionally sampled system models. The received signal sample at the mth receiver antenna at the kth time point can be expressed as

$$r_m(k) = \sum_{n=1}^{N} \sum_{l=0}^{L} \exp(j2\pi\omega_{m,n,l}(k)) s_n(k-l) h_{m,n,l}(k) + v_m(k), \tag{1}$$

where $\omega_{m,n,l}(k)$ incorporates all the phase and frequency instabilities related to l-th path of the subchannel between the nth transmitter and the mth receiver antenna, $s_n(k)$ is the kth linearly modulated information symbol transmitted from the nth transmitter antenna, $h_{m,n,l}(k)$ denotes the l-th sample of the channel impulse response (also called a l-th channel tap) between antennas n and m at time instance k, L is the degree of channel memory, and $v_m(k)$ is a sample of complex circular additive white Gaussian noise (AWGN) at the mth receiver antenna with the variance σ_v^2. Each symbol $s_n(k)$ takes values in the discrete space $\mathcal{S} = \{\alpha_1, \ldots, \alpha_J\}$, where $J = |\mathcal{S}|$ denotes the cardinality of the symbol space, i.e., the number of complex constellation points of the symbol mapper.

The vector of received signal samples at time instant k can be expressed as

$$\mathbf{r}(k) = \big(\mathbf{I}_M \otimes \mathbf{s}^T(k)\big)\mathbf{\Omega}(k)\mathbf{h}(k) + \mathbf{v}(k), \tag{2}$$

where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, operator $\otimes$ denotes the Kronecker product, and

$$\mathbf{r}(k) = \big[r_1(k), r_2(k), \ldots, r_M(k)\big]^T \in \mathbb{C}^M, \tag{3}$$

$$\mathbf{v}(k) = \big[v_1(k), v_2(k), \ldots, v_M(k)\big]^T \in \mathbb{C}^M. \tag{4}$$

[6] Sampling at the proper rate always allows one to recover the timing information in the digital domain [166].

In addition, symbol and channel vectors are defined as

$$\mathbf{s}(k) = \left[\mathbf{s}_1^T(k), \mathbf{s}_2^T(k), \ldots, \mathbf{s}_N^T(k)\right]^T \in \mathcal{S}^{N(L+1)}, \tag{5}$$

$$\mathbf{s}_n(k) = \left[s_n(k), s_n(k-1), \ldots, s_n(k-L)\right]^T \in \mathcal{S}^{(L+1)}, \tag{6}$$

$$\mathbf{h}(k) = \left[\mathbf{h}_1^T(k), \mathbf{h}_2^T(k), \ldots, \mathbf{h}_M^T(k)\right]^T \in \mathbb{C}^{MN(L+1)}, \tag{7}$$

$$\mathbf{h}_m(k) = \left[\mathbf{h}_{m,1}^T(k), \mathbf{h}_{m,2}^T(k), \ldots, \mathbf{h}_{m,N}^T(k)\right]^T \in \mathbb{C}^{N(L+1)}, \tag{8}$$

$$\mathbf{h}_{m,n}(k) = \left[h_{m,n,0}(k), h_{m,n,1}(k), \ldots, h_{m,n,L}(k)\right]^T \in \mathbb{C}^{(L+1)}, \tag{9}$$

and diagonal phasor matrices are defined as

$$\mathbf{\Omega}(k) = \mathrm{diag}\big(\mathbf{\Omega}_1(k), \mathbf{\Omega}_2(k), \ldots, \mathbf{\Omega}_M(k)\big) \in \mathbb{C}^{MN(L+1)\times MN(L+1)}, \tag{10}$$

$$\mathbf{\Omega}_m(k) = \mathrm{diag}\big(\mathbf{\Omega}_{m,1}(k), \mathbf{\Omega}_{m,2}(k), \ldots, \mathbf{\Omega}_{m,N}(k)\big) \in \mathbb{C}^{N(L+1)\times N(L+1)}, \tag{11}$$

$$\mathbf{\Omega}_{m,n}(k) = \mathrm{diag}\big(\exp(j2\pi\omega_{m,n,0}(k)), \ldots, \exp(j2\pi\omega_{m,n,L}(k))\big) \in \mathbb{C}^{(L+1)\times(L+1)}. \tag{12}$$

Before transmission, the information symbols at each transmitter antenna branch are collected into a block of K symbols, and the successive blocks are separated in time so that the interblock interference can be assumed to be zero. At the receiver, the block of received signal samples can be expressed as

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{v}, \tag{13}$$

where

$$\mathbf{r} = \left[\mathbf{r}_1^T, \mathbf{r}_2^T, \ldots, \mathbf{r}_M^T\right]^T \in \mathbb{C}^{M(K+L)}, \tag{14}$$

$$\mathbf{r}_m = \left[r_m(1), r_m(2), \ldots, r_m(K+L)\right]^T \in \mathbb{C}^{(K+L)}, \tag{15}$$

$$\mathbf{s} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, \ldots, \mathbf{s}_N^T\right]^T \in \mathcal{S}^{NK}, \tag{16}$$

$$\mathbf{s}_n = \left[s_n(1), s_n(2), \ldots, s_n(K)\right]^T \in \mathcal{S}^{K}, \tag{17}$$

$$\mathbf{v} = \left[\mathbf{v}_1^T, \mathbf{v}_2^T, \ldots, \mathbf{v}_M^T\right]^T \in \mathbb{C}^{M(K+L)}, \tag{18}$$

$$\mathbf{v}_m = \left[v_m(1), v_m(2), \ldots, v_m(K+L)\right]^T \in \mathbb{C}^{(K+L)}. \tag{19}$$

Moreover, it is assumed that all transmission paths in each subchannel experience the same frequency instabilities, i.e., $\omega_{m,n,l} \equiv \omega_{m,n}\ \forall l$. Under this assumption, reflecting the properties of typical transmission environments, the channel convolution matrix $\mathbf{H}$

is given as

$$\mathbf{H} = \begin{bmatrix} \boldsymbol{\Omega}_{1,1}\mathbf{H}_{1,1} & \cdots & \boldsymbol{\Omega}_{1,N}\mathbf{H}_{1,N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}_{M,1}\mathbf{H}_{M,1} & \cdots & \boldsymbol{\Omega}_{M,N}\mathbf{H}_{M,N} \end{bmatrix} \in \mathbb{C}^{M(K+L)\times NK}. \tag{20}$$

In (20), subchannel matrices are defined as

$$\mathbf{H}_{m,n} = \left[\bar{\mathbf{h}}_{m,n}(1), \ldots, \bar{\mathbf{h}}_{m,n}(K)\right] \in \mathbb{C}^{(K+L)\times K}, \tag{21}$$

$$\begin{aligned} \bar{\mathbf{h}}_{m,n}(k) = \big[&\mathbf{0}_{1\times(k-1)}, h_{m,n,0}(k), h_{m,n,1}(k+1), \ldots \\ &, h_{m,n,L}(k+L), \mathbf{0}_{1\times(K-k)}\big]^T \in \mathbb{C}^{(K+L)}, \end{aligned} \tag{22}$$

where the matrix (vector) $\mathbf{0}_{l\times k}$ denotes the $l \times k$ zero matrix (vector), and the diagonal phasor matrices $\boldsymbol{\Omega}_{m,n}$ are given as

$$\boldsymbol{\Omega}_{m,n} = \mathrm{diag}\big(\exp(j2\pi\omega_{m,n}(1)), \ldots, \exp(j2\pi\omega_{m,n}(K+L))\big) \in \mathbb{C}^{(K+L)\times(K+L)}. \tag{23}$$

One way to separate successive data blocks in time is to extend each data block with a cyclic prefix (CP). In particular, an extended data block transmitted from the nth transmitter antenna is given as

$$\mathbf{s}_n^{\mathrm{e}} = \left[\left(\mathbf{s}_n^{\mathrm{cp}}\right)^T, \mathbf{s}_n^T\right]^T \in \mathcal{S}^{(K+K_{\mathrm{prefix}})}, \tag{24}$$

where the cyclic prefix $\mathbf{s}_n^{\mathrm{cp}}$ is defined as

$$\mathbf{s}_n^{\mathrm{cp}} = \left[s_n(K - K_{\mathrm{prefix}} + 1), s_n(K - K_{\mathrm{prefix}} + 2), \ldots, s_n(K)\right]^T \in \mathcal{S}^{K_{\mathrm{prefix}}}. \tag{25}$$

If the channel is static (time-invariant) over the block and $K_{\mathrm{prefix}} > L$, the subchannel convolution matrix $\mathbf{H}_{m,n}$ becomes, after the removal of the cyclic prefix from the received signal, a circulant matrix. The $K \times K$ circulant subchannel convolution matrix is obtained from the matrix defined in (21) by removing the last L rows and adding them to the first L rows. The resulting block-circulant structure of the convolution matrix $\mathbf{H}$ facilitates the receiver implementation in wideband channels by allowing the employment of a similar low-complexity frequency-domain equalizer structure as in the OFDM systems. Another method to separate the successive data blocks in time is obtained simply by zero-padding the transmitted data blocks.

An alternative description for the samples of the received signal is given as

$$\mathbf{r} = \bar{\mathbf{X}}\mathbf{h} + \mathbf{v}, \tag{26}$$

where the channel vector $\mathbf{h}$ is defined as

$$\mathbf{h} = \big[\breve{\mathbf{h}}_{1,1}^T, \ldots, \breve{\mathbf{h}}_{M,1}^T, \breve{\mathbf{h}}_{1,2}^T, \ldots, \breve{\mathbf{h}}_{M,2}^T, \ldots, \breve{\mathbf{h}}_{1,N}^T, \ldots, \breve{\mathbf{h}}_{M,N}^T\big]^T \in \mathbb{C}^{MNL_{\text{ch}}(K+L)}, \tag{27}$$

$$\breve{\mathbf{h}}_{m,n} = \big[\mathbf{h}_{m,n,0}^T, \mathbf{h}_{m,n,1}^T, \ldots, \mathbf{h}_{m,n,L}^T\big]^T \in \mathbb{C}^{L_{\text{ch}}(K+L)}, \tag{28}$$

$$\mathbf{h}_{m,n,l} = \big[h_{m,n,l}(1), h_{m,n,l}(2), \ldots, h_{m,n,l}(K+L)\big]^T \in \mathbb{C}^{(K+L)}, \tag{29}$$

and the system matrix $\bar{\mathbf{X}}$ is given as

$$\bar{\mathbf{X}} = \big[\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \ldots, \bar{\mathbf{X}}_N\big] \in \mathbb{C}^{M(K+L)\times MNL_{\text{ch}}(K+L)}, \tag{30}$$

$$\bar{\mathbf{X}}_n = \text{diag}\big(\mathbf{\Omega}_{1,n}, \mathbf{\Omega}_{2,n}, \ldots, \mathbf{\Omega}_{M,n}\big)\big(\mathbf{I}_M \otimes \mathbf{X}_n\big) \in \mathbb{C}^{M(K+L)\times ML_{\text{ch}}(K+L)}, \tag{31}$$

$$\mathbf{X}_n = \big[\mathbf{X}_{n,0}, \mathbf{X}_{n,1}, \ldots, \mathbf{X}_{n,L}\big] \in \mathbb{C}^{(K+L)\times L_{\text{ch}}(K+L)}, \tag{32}$$

$$\mathbf{X}_{n,l} = \text{diag}\Big(\big[\mathbf{0}_{1\times l}, s_n(1), \ldots, s_n(K), \mathbf{0}_{1\times(L-l)}\big]\Big) \in \mathbb{C}^{(K+L)\times(K+L)}. \tag{33}$$

The received signal vector $\mathbf{r}$ and the noise vector $\mathbf{v}$ are defined in (14)-(15) and in (18)-(19), respectively. In addition, for notational simplicity, the following *general notational convention* is made: indexes for transmitter and receiver antennas are omitted whenever only single antenna systems are considered, e.g., $r(k) \equiv r_1(k)$, $v(k) \equiv v_1(k)$, $h_l(k) \equiv h_{1,1,l}(k)$, $\omega_l(k) \equiv \omega_{1,1,l}(k)$, and so on.

The discrete-time channel impulse response $\mathbf{h}_{m,n}(k)$ for each subchannel is modeled either as a deterministic time-varying process or as a random time-varying Rayleigh fading process. The latter modeling of the radio channel implies that the samples of the l-th channel tap $h_{m,n,l}(k)$ are distributed according to a zero-mean complex Gaussian distribution. If the channel is additionally assumed to be wide sense stationary (WSS), its second order statistics are time-invariant. Furthermore, provided that the fading process is circular, the distribution of each channel tap is completely characterized alternatively by the covariance matrix or the power spectral density [201], both of which give nonparametric descriptions of the underlying fading statistics. The power spectral density of the fading process is often modeled using Clarke's U-shaped function [46, 121] corresponding to a uniform angular distribution of the received energy at the receiver antenna (i.e., the isotropic scattering model). In this case, the autocorrelation function of the l-th channel tap is given by

$$C_{h_{m,n,l}}(i) = E\big[h_{m,n,l}(k)h_{m,n,l}^*(k-i)\big] = \sigma_{h_{m,n,l}}^2 \mathcal{J}_0(\Omega_D i), \tag{34}$$

where $\sigma_{h_{m,n,l}}^2 = C_{h_{m,n,l}}(0)$, $\Omega_D = 2\pi B_d T_s$, B_d is the Doppler spread of the fading channel, and $\mathcal{J}_0(\cdot)$ denotes the Bessel function of the first kind and zero-order. Since

the random fading gain $h_{m,n,l}(k)$ is strictly band-limited to the maximum Doppler frequency B_d, its autocorrelation function has infinite length.

An alternative description of the fading channel dynamics can be obtained by means of the parametric model. The goal of the parametric modeling is to postulate a stochastic model, sometimes referred to as a hypermodel [149], to the fading channel so that an acceptable representation of the underlying fading process is obtained with as few parameters as possible. Despite an infinitely long correlation in time, the Rayleigh fading channel gain can, in theory, be approximated with an arbitrary small error by an autoregressive (AR) model of a sufficiently high order [106]. Consequently, the channel process $\mathbf{h}_{m,n}(k)$ can be approximately modeled as a pth-order Gauss-Markov process. The resulting dynamic discrete-time state-space model will be described in the next section.

2.2 Parametric modeling of fading statistics

Purely for notational convenience, the AR description for the fading statistics and the resulting dynamic state-space system models are given only for the single antenna ($M = 1$ and $N = 1$) case. In addition, it is assumed that the frequency instabilities at each transmission path are exclusively due to the constant frequency shifts ν_l, i.e, $\omega_l(k) = \nu_l k$, while the constant phase offset and slow phase drifting are lumped with the complex-valued CIR.

2.2.1 Lowpass AR modeling

When the pth-order autoregressive (AR(p)) hypermodel for a Rayleigh fading frequency-selective channel is employed, transmission of information symbols over the radio channel can be described with the following state-space model:

$$\boldsymbol{h}(k) = \underline{\boldsymbol{A}}\,\boldsymbol{h}(k-1) + \boldsymbol{G}\bar{\mathbf{w}}(k-1) \tag{35}$$

$$r(k) = \boldsymbol{s}^T(k)\bar{\boldsymbol{\Omega}}(k)\boldsymbol{h}(k) + v(k). \tag{36}$$

In this model, the zero-padded symbol vector $\boldsymbol{s}(k)$ is given by

$$\boldsymbol{s}(k) = \left[\mathbf{s}^T(k), \mathbf{0}_{1\times L_{\text{ch}}(p-1)}\right]^T, \in \mathbb{C}^{pL_{\text{ch}}} \tag{37}$$

and the vector channel process $\boldsymbol{h}(k)$ is defined as

$$\boldsymbol{h}(k) = \left[\mathbf{h}^T(k), \mathbf{h}^T(k-1), \ldots, \mathbf{h}^T(k-p+1)\right]^T, \in \mathbb{C}^{pL_{\text{ch}}} \tag{38}$$

where $L_{\text{ch}} = L + 1$ denotes the length of the CIR. The extended time-varying phasor matrix $\bar{\mathbf{\Omega}}(k)$ is defined as

$$\bar{\mathbf{\Omega}}(k) = \begin{bmatrix} \mathbf{\Omega}(k) & \mathbf{0}_{L_{\text{ch}} \times (p-1)L_{\text{ch}}} \\ \mathbf{0}_{(p-1)L_{\text{ch}} \times L_{\text{ch}}} & \mathbf{I}_{(p-1)L_{\text{ch}}} \end{bmatrix} \in \mathbb{C}^{pL_{\text{ch}} \times pL_{\text{ch}}}. \tag{39}$$

Moreover, the complex circular Gaussian process noise vector $\bar{\mathbf{w}}(k)$ is defined as

$$\bar{\mathbf{w}}(k) = \left[\bar{w}_1(k), \bar{w}_2(k), \ldots, \bar{w}_{L_{\text{ch}}}(k)\right]^T \in \mathbb{C}^{L_{\text{ch}}} \tag{40}$$

and is assumed to be independent of the complex circular receiver noise process $v(k)$. The covariance matrix of $\bar{\mathbf{w}}(k)$ is given as $\mathbf{C}_{\bar{\mathbf{w}}(k)}(i) = E\left[\bar{\mathbf{w}}(k)\bar{\mathbf{w}}(k-i)^H\right] = I_{L_{\text{ch}}}\delta(i)$, where $\delta(i)$ denotes Kronecker delta function. The deterministic model matrices $\underline{\boldsymbol{A}}$ and $\boldsymbol{G}$ are defined as

$$\underline{\boldsymbol{A}} = \begin{bmatrix} \bar{\mathbf{A}}_1 & \bar{\mathbf{A}}_2 & \cdots & \bar{\mathbf{A}}_p \\ \mathbf{I}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}}} & \cdots & \mathbf{0}_{L_{\text{ch}}} \\ & \ddots & \mathbf{0}_{L_{\text{ch}}} & \vdots \\ & \mathbf{0}_{L_{\text{ch}}} & \mathbf{I}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}}} \end{bmatrix} \in \mathbb{C}^{pL_{\text{ch}} \times pL_{\text{ch}}}, \quad \boldsymbol{G} = \begin{bmatrix} \mathbf{G} \\ \mathbf{0}_{L_{\text{ch}}} \\ \vdots \\ \mathbf{0}_{L_{\text{ch}}} \end{bmatrix} \in \mathbb{R}^{pL_{\text{ch}} \times L_{\text{ch}}},$$

where $\bar{\mathbf{A}}_1, \ldots, \bar{\mathbf{A}}_p$ are $L_{\text{ch}} \times L_{\text{ch}}$ submatrices with real diagonal elements, $\mathbf{G}$ is a $L_{\text{ch}} \times L_{\text{ch}}$ diagonal matrix with l-th diagonal element being $[\mathbf{G}]_{l,l} = \sqrt{g_l}$, and $\mathbf{0}_{L_{\text{ch}}}$ is a $L_{\text{ch}} \times L_{\text{ch}}$ zero matrix. The nonzero complex valued off-diagonal elements of the submatrices $\bar{\mathbf{A}}_1, \ldots, \bar{\mathbf{A}}_p$ reflect the correlation between the channel taps.

2.2.2 Complex bandpass AR modeling

The carrier frequency offset rotates the received signal samples around the time axis of the complex domain description of the received signal. Alternatively, the carrier frequency offset can be incorporated into the CIR so that the resulting modified CIR is correspondingly rotated around the time axis. Specifically, by doing the replacement

$$\boldsymbol{f}(k) \triangleq \bar{\mathbf{\Omega}}(k)\boldsymbol{h}(k) \tag{41}$$

into (36) (correspondingly, $\mathbf{f}(k) \triangleq \mathbf{\Omega}(k)\mathbf{h}(k)$ and $f_l(k) \triangleq \exp(j2\pi\nu_l k)h_l(k)$) and multiplying both sides of (35) by $\bar{\mathbf{\Omega}}(k)$, a new state-space model is obtained:

$$\boldsymbol{f}(k) = \boldsymbol{A}\boldsymbol{f}(k-1) + \boldsymbol{G}\mathbf{w}(k-1) \tag{42}$$

$$r(k) = \boldsymbol{s}^T(k)\boldsymbol{f}(k) + v(k). \tag{43}$$

Provided that the noise components $\bar{w}_i(k)\ \forall i$ are circular as well as independent and identically distributed with respect to both indices i and k, the Gaussian process noise vector $\mathbf{w}(k) = \boldsymbol{\Omega}(k)\bar{\mathbf{w}}(k)$ is statistically equivalent to $\bar{\mathbf{w}}(k)$ [201, 202]. Likewise, the state-space models (42) and (43) are statistically equivalent to (35) and (36).

The deterministic model matrix $\boldsymbol{A}$ can now be expressed as

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_p \\ \mathbf{I}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}}} & \cdots & \mathbf{0}_{L_{\text{ch}}} \\ & \ddots & \mathbf{0}_{L_{\text{ch}}} & \vdots \\ & \mathbf{0}_{L_{\text{ch}}} & \mathbf{I}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}}} \end{bmatrix} \in \mathbb{C}^{pL_{\text{ch}} \times pL_{\text{ch}}},$$

where

$$\mathbf{A}_i = \boldsymbol{\Omega}(k)\bar{\mathbf{A}}_i\boldsymbol{\Omega}^H(k-i) \in \mathbb{C}^{L_{\text{ch}} \times L_{\text{ch}}}, \quad i = 1, \cdots, p. \tag{44}$$

For a static frequency offset, the model matrix $\boldsymbol{A}$ is time-invariant, and, thus, the state equation (42) is stationary. In particular, the l-th diagonal element of $\mathbf{A}_i$ is given as

$$[\mathbf{A}_i]_{l,l} = \left[\bar{\mathbf{A}}_i\right]_{l,l} \exp(j2\pi i\nu_l), \tag{45}$$

and, since the diagonal elements of $\bar{\mathbf{A}}_i$ are real numbers, the phase of $[\mathbf{A}_i]_{l,l}$ now depends exclusively on the frequency shift ν_l. Moreover, in typical channel circumstances, where a frequency shift in different paths of the channel can be regarded as equal, i.e., $\nu_l \equiv \nu\ \forall l$, the model matrix $\mathbf{A}_i$ can be written as

$$\mathbf{A}_i = \bar{\mathbf{A}}_i \exp(j2\pi i\nu), \quad i = 1, \cdots, p. \tag{46}$$

This frequency offset modulated channel model (hypermodel) is hereafter referred to as the *complex bandpass* AR channel model, since it represents a frequency shifted version of the original baseband (lowpass) model. Data transmission over a flat-fading channel in the presence of a frequency offset is illustrated in Fig. 1, when the channel is modeled with a first order complex bandpass AR model.

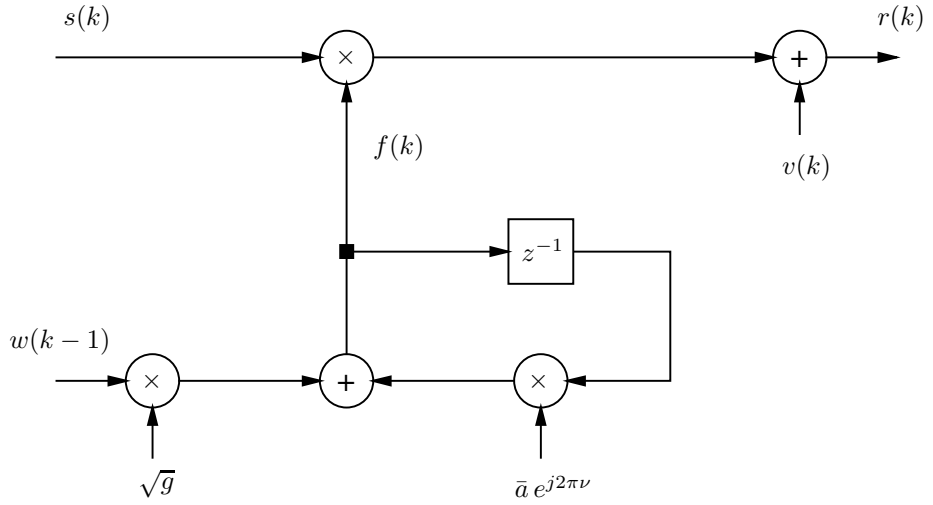

Fig 1. Signal model for transmission over flat-fading channels in the presence of a frequency offset.

2.3 Variational bounding: An introduction via simple examples

Basically, the main objective of the variational methods is to transform the problem of interest into an optimization problem via the introduction of extra degrees of freedom known as *variational parameters* [125]. The optimization problem should be defined so that, for fixed values of variational parameters, the transformed problem has a simple closed form solution, solving approximately the original problem. The variational parameters, in turn, are adjusted so as to yield a sequence of improved approximations.

The variational methodology often yields approximations that provide bounds on the probabilities of interest [125, 217, 119]. The basics of variational bounding techniques are next introduced with four simple yet illuminating examples so that the great potential behind these methods could be fully appreciated. Mathematically rigorous and more general treatment of the variational optimization methods with theory and applications can be found in many tutorial expositions [124, 118, 80, 252].

Example 1: Consider any continuously differentiable *convex* function $f(x)$. The

tangent line of $f(x)$ at the point $x = x_0$ is given by

$$L(x; x_0) = f(x_0) + \frac{\partial}{\partial x} f(x)|_{x=x_0}(x - x_0).$$

Convexity of $f(x)$ guarantees by definition that any tangent line $L(x; x_0)$ always remains below the function $f(x)$ itself, i.e., $f(x) \geq L(x; x_0)$ for all x and x_0 and $f(x_0) = L(x_0; x_0)$.

Thus, a collection of tangent lines of a convex function can be interpreted as a parameterized family of lower bounds for that function. Naturally, the tangents are parameterized by their locations in this case. Thus, any one of these tangent lines can be considered to be a global approximation for the function $f(x)$. In the terminology of the variational methods, $L(x; x_0)$ is termed as a variational lower bound of $f(x)$, whereas the parameter x_0 is called a variational free parameter. Put in another way, the variational free parameter x_0 can now be *freely* adjusted so as to make $L(x; x_0)$ an as accurate approximation of $f(x)$ as possible, around the point of interest. The lower bound $L(x; x_0)$ coincides with $f(x)$ at a single point x_0 and the approximation degrades as x recedes from x_0 at a rate which depends on the curvature of $f(x)$. If the curvature is low, $L(x; x_0)$ can be regarded as a relatively good approximation of the objective function $f(x)$ at a wide range of values around x_0.

For a continuously differentiable *concave* function $f(x)$, the tangent line $L(x; x_0)$ gives an upper bound of $f(x)$. As a specific example, suppose that $f(x) = \ln(x)$. The variational bounding via parameterized tangent line readily yields

$$\ln(x) \leq \lambda x - \ln(\lambda) - 1,$$

where $\lambda = 1/x_0$ denotes the variational free parameter. The upper bounding of $\ln(x)$ is illustrated in Fig. 2 by using tangent lines with different slopes. As may be noticed from the figure, for each value of x it is possible to find a value of λ that gives a very tight upper bound for $\ln(x)$ in some neighborhood of x. In consequence, the concave function $\ln(x)$ can be interpreted as a solution to the optimization problem,

$$\ln(x) = \min_{\lambda} \big\{ \lambda x - \Phi(\lambda) \big\},$$

where $\Phi(\lambda) = \ln(\lambda) + 1$ is called a *conjugate* or *dual* function [125]. On the other hand, it can be easily verified that the conjugate function $\Phi(\lambda)$ can be obtained from the following dual expression:

$$\Phi(\lambda) = \min_{x} \big\{ \lambda x - \ln(x) \big\},$$

which is also known as a Legendre transformation of $\ln(x)$. In general, the relationship between the concave (or convex) function $f(x)$ and its dual $\Phi(\lambda)$ is called *convex duality* in convex analysis [214]. □

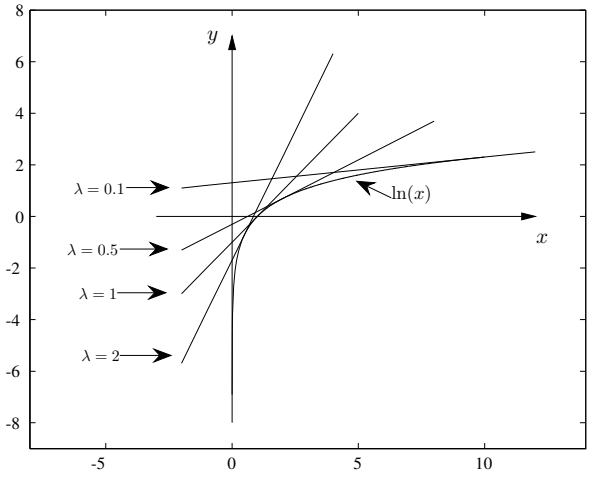

Fig 2. Variational bounding of $\ln(x)$ by using parametrized tangent lines.

An illustrative geometrical interpretation for variational bounding of a concave function via the principle of convex duality is provided in [125]. In fact, many of the variational transformations that can be found in the literature of graphical models are based on the general principle of convex duality [125].

Example 2: Let the received data be $r = s + v$, where the symbol s is a random variable, taking values in the discrete set $\mathcal{S} \in \{1, -1\}$ with equal probability, and v is a zero-mean, Gaussian random variable with unknown variance denoted as σ_v^2. Suppose that we want to estimate the variance σ_v^2, given the received data r, via ML estimation. The log-likelihood function for this simple estimation problem is easily obtained as

$$\mathcal{L}(\sigma_v^2) = -\frac{1}{2}\ln(8\pi) - \frac{1}{2}\ln(\sigma_v^2) - \frac{1}{2\sigma_v^2}(r^2+1) + \ln\Big(e^{\frac{r}{\sigma_v^2}} + e^{-\frac{r}{\sigma_v^2}}\Big).$$

Unfortunately, the maximization of $\mathcal{L}(\sigma_v^2)$ over σ_v^2, a prerequisite for ML estimation, does not seem to yield any analytical, closed-form solution. So, let us try to find a

variational lower bound for the above log-likelihood function so that a more tractable estimator could be obtained.

Towards this end, let us consider the function

$$g(y) = \ln\left(e^{\frac{y}{2}} + e^{-\frac{y}{2}}\right),$$

which can alternatively be expressed as

$$g(y) = -\frac{y}{2} + \ln(1 + e^{y}).$$

Since $\ln(1 + e^{y})$ is a convex function of y (as can be readily verified by calculating the second derivative), the principle of convex duality applies, and the function $g(y)$ can be lower bounded with linear functions. In particular, we obtain that

$$g(y) \geq -\frac{y}{2} + \lambda y + H(\lambda),$$

where $H(\lambda)$ is a binary entropy function, $H(\lambda) \triangleq -\lambda \ln(\lambda) - (1-\lambda)\ln(1-\lambda)$. Inserting $y = \frac{2r}{\sigma_v^2}$, the log-likelihood function of the noise variance can now be expressed *variationally* as

$$\mathcal{L}(\sigma_v^2) = \max_{\lambda}\left\{-\frac{1}{2}\ln(8\pi) - \frac{1}{2}\ln(\sigma_v^2) - \frac{1}{2\sigma_v^2}(r^2+1) - \frac{r}{\sigma_v^2} + \frac{2\lambda r}{\sigma_v^2} + H(\lambda)\right\}.$$

Consequently, the ML estimate of the noise variance can be expressed by using double maximization as follows:

$$\left(\widehat{\sigma_v^2}\right)_{\text{ML}} = \arg\max_{\sigma_v^2}\max_{\lambda}\left\{-\frac{1}{2}\ln(\sigma_v^2) - \frac{1}{2\sigma_v^2}(r^2+1) - \frac{r(1-2\lambda)}{\sigma_v^2} + H(\lambda)\right\}.$$

This double maximization can be carried out by an iterative coordinate ascent algorithm, maximizing over one parameter at a time, while keeping the other fixed. Thus, an overall procedure for estimating σ_v^2 consists of two alternating steps, given at the ith iteration by

$$\begin{aligned}\lambda^{(i)} &= \frac{1}{1 + e^{-2r/\widehat{\sigma_v^2}^{(i-1)}}} = P\left(s = 1 \middle| r, \widehat{\sigma_v^2}^{(i-1)}\right) \in]0,1[\\ \widehat{\sigma_v^2}^{(i)} &= r^2 + 2r\left(1 - 2\lambda^{(i)}\right) + 1.\end{aligned}$$

As may be readily noticed, the likelihood is guaranteed to monotonically increase at each iteration—a highly favorable property that is shared with the standard EM algorithm. □

Example 3: Let us consider the same system model and problem setting as in the previous example. This time the log-likelihood function is, however, lower bounded as follows:

$$\mathcal{L}(\sigma_v^2) = \ln \sum_s p(r,s|\sigma_v^2) = \ln \sum_s \lambda(s) \frac{p(r,s|\sigma_v^2)}{\lambda(s)}$$
$$\geq \sum_s \lambda(s) \ln \frac{p(r,s|\sigma_v^2)}{\lambda(s)} = \sum_s \lambda(s) \ln p(r,s|\sigma_v^2) - \sum_s \lambda(s) \ln \lambda(s),$$

where $p(r,s|\sigma_v^2) = p(r|s,\sigma_v^2)P(s)$ denotes the joint probability density function of r and s given the noise variance σ_v^2 and $\lambda(s)$ denotes the variational "trial" distribution of s. The second line in the above equation is obtained by making an appeal to Jensen's inequality [52]. In this case, the distribution $\lambda(s)$ for different values of s forms a vector of free parameters to be optimized.

Based on this variational formulation, the ML estimate of the noise variance can alternatively be obtained as the maximizing argument of the following optimization:

$$\left(\widehat{\sigma_v^2}\right)_{\text{ML}} = \arg\max_{\sigma_v^2} \max_{\lambda(s)} \left\{ \sum_s \lambda(s) \ln p(r,s|\sigma_v^2) - \sum_s \lambda(s) \ln \lambda(s) \right\}.$$

As can be easily verified, the coordinate ascent on this lower bound yields the standard EM algorithm, expressed at the ith iteration as

$$\lambda(s)^{(i)} = P\left(s \middle| r, \widehat{\sigma_v^2}^{(i-1)}\right)$$
$$\widehat{\sigma_v^2}^{(i)} = r^2 - 2r \sum_s \lambda(s) s + 1.$$

Although the EM algorithm is conventionally presented as an alternation between an expectation step (E-step) and a maximization step (M-step), both steps can, from a variational optimization perspective, be viewed as maximization steps. This fact has formally been proven, e.g., in [54, 176].

Importantly, also this version of the variational transformation can be justified by appealing to the general principle of convex duality, defined now in terms of vectors [125]. Specifically, let us view the log probability function $\ln p(r,s|\sigma_v^2)$ as a vector of the real functions of r, defined on the set of values of s, and treat it as a variable vector. Then, the key question is whether the log-likelihood function $\mathcal{L}(\sigma_v^2)$ is a convex function in $\ln p(r,s|\sigma_v^2)$. By writing

$$\mathcal{L}(\sigma_v^2) = \ln\left(\sum_s e^{\ln p(r,s|\sigma_v^2)} \right),$$

it is easy to show that $\mathcal{L}(\sigma_v^2)$ indeed fulfills the convexity requirement. Moreover, it can be noticed that the Legendre transformation of $\mathcal{L}(\sigma_v^2)$ is the negative of the variational entropy function: $\Phi(\lambda) = \sum_s \lambda(s) \ln \lambda(s)$. □

Basically, the variational algorithms presented in Examples 2 and 3 are similar, except regarding the number of free parameters to be optimized. Convergence properties of these two iterative estimators are studied here by numerically evaluating the evolution of the log-likelihood values in a simple exemplary setting. According to simulations, both algorithms produce exactly the same chain of values for $\mathcal{L}(\sigma_v^2)$ when started from equivalent initial conditions.[7]

Table 1. Exemplary convergence table for the iterative bounding algorithms of Examples 2 and 3. The exact value of the log-likelihood is $\mathcal{L}_{\text{true}}(\sigma_v^2) = -1.5453$.

iteration no.	$\mathcal{L}(\sigma_v^2)$
1	-2.6799
2	-1.9399
3	-1.8592
4	-1.8046
5	-1.7299
6	-1.6089
7	-1.5467
8	-1.5453
9	-1.5453

A chain of log-likelihood values obtained from an example simulation exercise, in which 9 iterations were executed, is presented in Table 1. The tabulated likelihood values represent those obtained by using algorithms of Examples 2 and 3 in separate test runs. The true value of the noise variance was $\sigma_v^2 = 0.6310$, while the ML estimate turned out to be $\left(\widehat{\sigma_v^2}\right)_{\text{ML}} = 0.3223$. The initial value of the free parameter was set at $\lambda^{(0)} = 0.1$ in the first implementation (Example 2), whereas $\lambda(s)^{(0)} = [0.1\ 0.9]^T$ was used in the second one (Example 3). In addition, the received signal assumed the value

[7] In general, however, endowing the variational optimization problem with extra degrees of freedom may render the formulations more flexible.

$r = 1.5674$ in both test cases.

Example 4: Let the received data be $r = hs + v$, where the symbol s is a random variable, taking values in the discrete set $\mathcal{S} \in \{1, -1\}$ with equal probability, h and v are zero-mean, independent Gaussian random variables with variances σ_h^2 and σ_v^2, respectively. Suppose again that we want to estimate the variance σ_v^2, given the received data r. The log-likelihood function for this problem is given by

$$\mathcal{L}(\sigma_v^2) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_h^2 + \sigma_v^2) - \frac{1}{2}\frac{r^2}{\sigma_h^2 + \sigma_v^2}.$$

In this case, the ML estimate of the noise variance is readily obtained as

$$\left(\widehat{\sigma_v^2}\right)_{\text{ML}} = \begin{cases} r^2 - \sigma_h^2 & \text{if } r^2 > \sigma_h^2, \\ 0 & \text{otherwise,} \end{cases}$$

and, hence, there is no need for estimation via variational bounding. Justifiably, one may question the meaningfulness of this example as a whole. In any case, let us consider, for merely illustrative purposes, the following approximation for the log-likelihood function:

$$\begin{aligned}
\mathcal{L}(\sigma_v^2) &= \ln \int_h \sum_s p(r|s,h,\sigma_v^2)p(s)p(h)dh \\
&\geq -\frac{1}{2}\ln(8\pi\sigma_v^2) - \frac{1}{2\sigma_v^2}\int_h q(h)(r^2+h^2)dh + \int_h q(h)\Big(e^{\frac{rh}{\sigma_v^2}} + e^{-\frac{rh}{\sigma_v^2}}\Big)dh \\
&\quad + \int_h q(h)\ln p(h) - \int_h q(h)\ln q(h) \\
&\geq -\frac{1}{2}\ln(8\pi\sigma_v^2) - \frac{1}{2\sigma_v^2}\big(r^2 + \langle h^2\rangle\big) - \frac{r(1-2\lambda)\langle h\rangle}{\sigma_v^2} + H(\lambda) \\
&\quad + \int_h q(h)\ln p(h) - \int_h q(h)\ln q(h),
\end{aligned}$$

where the following definition is used: $\langle h\rangle \triangleq \int_h q(h)\, h\, dh$. The "trial" distribution $q(h)$ and λ now form a set of free parameters to be optimized. In the above formulation, the variational bounding via the principle of convex duality is applied twice in two different forms introduced in the previous examples. Based on this bounding, the expression for the ML estimate of σ_v^2 can alternatively be presented as

$$\begin{aligned}
\left(\widehat{\sigma_v^2}\right)_{\text{ML}} = \arg\max_{\sigma_v^2}\max_{q(h)}\max_{\lambda}\Big\{ &-\frac{1}{2}\ln(\sigma_v^2) - \frac{1}{2\sigma_v^2}\big(r^2 + \langle h^2\rangle\big) \\
&- \frac{r(1-2\lambda)\langle h\rangle}{\sigma_v^2} + H(\lambda) + \int_h q(h)\ln p(h) - \int_h q(h)\ln q(h)\Big\}.
\end{aligned}$$

An iterative coordinate ascent at the ith iteration yields

$$\lambda^{(i)} = \frac{1}{1+\exp\left\{-\frac{2r\langle h\rangle^{(i-1)}}{\widehat{\sigma_v^2}^{(i-1)}}\right\}}$$

$$\langle h\rangle^{(i)} = \int_h q^{(i)}(h)h\ dh = \frac{\sigma_h^2}{\sigma_h^2+\widehat{\sigma_v^2}^{(i-1)}}\left(1-2\lambda^{(i)}\right)r$$

$$\langle h^2\rangle^{(i)} = \int_h q^{(i)}(h)h^2\ dh = \frac{\sigma_h^2}{\sigma_h^2+\widehat{\sigma_v^2}^{(i-1)}}\left(\widehat{\sigma_v^2}^{(i-1)} + \frac{\left(1-2\lambda^{(i)}\right)^2 r^2\sigma_h^2}{\sigma_h^2+\widehat{\sigma_v^2}^{(i-1)}}\right)$$

$$\widehat{\sigma_v^2}^{(i)} = r^2 + 2r\langle h\rangle^{(i)}\left(1-2\lambda^{(i)}\right) + \langle h^2\rangle^{(i)}.$$

Again, at each iteration, the likelihood is guaranteed to monotonically increase. A numerical example of the convergence of this iterative optimization algorithm is presented in Table 2. □

Table 2. Exemplary convergence table for the iterative bounding algorithm of Example 4. The exact value of the log-likelihood is $\mathcal{L}_{\text{true}}(\sigma_v^2) = -1.0355$. In addition, the following parameter values were assumed: $r = 0.2$, $\lambda^{(0)} = 0.01$, $\widehat{\sigma_v^2}^{(0)} = 10$.

iteration no.	$\mathcal{L}(\sigma_v^2)$
1	-2.0362
2	-1.2027
3	-1.1466
4	-1.1253
5	-1.1163
6	-1.1124
7	-1.1107
8	-1.1099
9	-1.1094
10	-1.1093
11	-1.1093
12	-1.1093

Through these rather simple and somewhat unrealistic examples, the feasibility of the joint inference and estimation via the variational bounding technique has been illus-

trated. While aiming for the ML estimate of the noise variance, the symbol APPs were obtained as a by-product of the iterative bounding algorithm. The variational bounding can also be applied to approximate inference alone, as will be discussed in the sequel. Moreover, the variational formula inferred from the convex duality principle may be modified so that it does not provide a lower bound on the probability of interest any more, but may, however, produce more accurate approximation. Examples include the Bethe approximation method [263] and its generalizations [265]. Next, the variational methods will be looked at from the statistical physics point of view.

2.4 Variational free energy minimization framework

In this section, a generic variational free energy minimization (VFEM) framework will be introduced, building upon the concept of a "free energy" from statistical physics. At the same time, the interplay between the principle of convex duality and the concept of free energy minimization will be emphasized along the way.

In fact, the origin of the variational methods is in statistical physics [263], where these methods have been applied in order to obtain a computationally feasible approach for evaluating the quantity known as a *Helmholtz* free energy. However, concentrating more on the modern applications of these methods in the areas of statistical signal processing and digital communications, a principled goal of the VFEM framework, provided in this section, will be stated so as to give a computationally oriented meaning to the generic problem of variatioanal inference—a meaning that reposes on basic concepts in the field of convex analysis.

Suppose that a set of K discrete-valued variables is given as $\{X_1, \ldots, X_K\}$, while x_i denotes a realization of X_i. In addition, let $\mathbf{x}$ stand for $\{x_1, \ldots, x_K\}$, and let the joint probability mass function $P(X_1 = x_1, \ldots, X_K = x_K)$ be succintly expressed as $p(\mathbf{x})$. In the physical systems, the vector $\mathbf{x}$ can represent an overall state of the system, whereas each element of $\mathbf{x}$, i.e., $x_i \ \forall i$, label a state of the ith particle of the system. Such a particle could represent, for example, an atom in a magnetic crystal, and the state of the atom in that case describes whether a given electron has an "up" spin or "down" spin [265]. If the system is in thermal equilibrium with a large reservoir, it can be shown that, under the fundamental assumptions of thermal physics, the probability that the system will be in a configuration $\mathbf{x}$ is given by the Boltzmann distribution:

$$p(\mathbf{x}) = \frac{e^{-\mathcal{E}(\mathbf{x})/T}}{Z}, \tag{47}$$

where $\mathcal{E}(\mathbf{x})$ denotes the energy of the system at the state configuration $\mathbf{x}$, T is the temperature, and Z is a normalization constant, also known as the *partition function* (see, e.g., [133]). The partition function can be expressed as

$$Z = \sum_{\mathbf{x}} e^{-\mathcal{E}(\mathbf{x})/T}. \tag{48}$$

Statistical physicists have devoted much of their energy to justifying Boltzmann's law [133, 203, 265]. On the other hand, for nonphysical systems where $p(\mathbf{x})$ simply denotes the joint probability mass function for the discrete-valued elements of a vector $\mathbf{x}$, Boltzmann's law (47) can be viewed as a postulate that defines an energy for the system and where the temperature T can be set arbitrarily since it only sets a scale for the units in which the energy is measured.

The Helmholtz free energy, defined as

$$\mathcal{F}_{\text{Helmholtz}} = -\ln Z, \tag{49}$$

is a fundamentally important quantity in statistical physics since it can be used to characterize the underlying physical system [36, 203]. Physicists have, therefore, developed a wide range of techniques to calculate the Helmholtz free energy, either exactly or approximately. The most important approximate methods are Monte Carlo probabilistic resampling algorithms [156] and deterministic variational inference algorithms [265].

Towards describing the variational inference algorithms in terms of statistical physics, let us first consider the problem of computing the Helmholtz free energy as a variational optimization problem. Specifically, using Jensen's inequality, $\mathcal{F}_{\text{Helmholtz}}$ can be bounded as

$$\begin{aligned}\mathcal{F}_{\text{Helmholtz}} &= -\ln \sum_{\mathbf{x}} e^{-\mathcal{E}(\mathbf{x})/T} = -\ln \sum_{\mathbf{x}} Q(\mathbf{x}) \frac{e^{-\mathcal{E}(\mathbf{x})/T}}{Q(\mathbf{x})} \\ &\leq -\sum_{\mathbf{x}} Q(\mathbf{x}) \ln \frac{e^{-\mathcal{E}(\mathbf{x})/T}}{Q(\mathbf{x})} = \frac{1}{T}\sum_{\mathbf{x}} Q(\mathbf{x})\mathcal{E}(\mathbf{x}) + \sum_{\mathbf{x}} Q(\mathbf{x}) \ln Q(\mathbf{x}),\end{aligned} \tag{50}$$

where $Q(\mathbf{x})$ is called a *variational* (trial) distribution.[8] Similarly to the variational bounding in Example 3 of the previous section, the lower bound (50) for the Helmholtz free energy can be justified by making an appeal to convex duality theory. So, the classical variational bounding technique in statistical physics and the principle of convex duality in convex analysis meet each other in the formulation given in (50).

[8] Since $Q(\mathbf{x})$ is a distribution, it should, of course, be normalized and obey $0 \leq Q(\mathbf{x}) \leq 1$ for all $\mathbf{x}$.

Let us define the *variational average energy* as

$$\mathcal{U}(Q) = \frac{1}{T} \sum_{\mathbf{x}} Q(\mathbf{x}) \mathcal{E}(\mathbf{x}), \tag{51}$$

and the *variational entropy* as

$$\mathcal{H}(Q) = -\sum_{\mathbf{x}} Q(\mathbf{x}) \ln Q(\mathbf{x}). \tag{52}$$

In addition, let us define the *variational free energy* (also sometimes called the *Gibbs free energy* [265]) as

$$\mathcal{F}(Q) = \mathcal{U}(Q) - \mathcal{H}(Q). \tag{53}$$

Based on these definitions, the variational free energy can be interpreted as an upper bound for the Helmholtz free energy, i.e., $\mathcal{F}_{\text{Helmholtz}} \leq \mathcal{F}(Q)$. In fact, it can be further elaborated as

$$\begin{aligned} \mathcal{F}(Q) &= -\sum_{\mathbf{x}} Q(\mathbf{x}) \ln \frac{e^{-\mathcal{E}(\mathbf{x})/T}}{Q(\mathbf{x})} = -\sum_{\mathbf{x}} Q(\mathbf{x}) \ln \frac{p(\mathbf{x})Z}{Q(\mathbf{x})} \\ &= -\ln Z + \sum_{\mathbf{x}} Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})}{p(\mathbf{x})} = \mathcal{F}_{\text{Helmholtz}} + D\big(Q(\mathbf{x}) \| p(\mathbf{x})\big), \end{aligned} \tag{54}$$

where the Kullback-Leibler (KL) divergence $D\big(Q(\mathbf{x}) \| p(\mathbf{x})\big)$ (also known as the cross-entropy or the relative entropy) is defined as

$$D\big(Q(\mathbf{x}) \| p(\mathbf{x})\big) = \sum_{\mathbf{x}} Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})}{p(\mathbf{x})}. \tag{55}$$

Since $D\big(Q(\mathbf{x}) \| p(\mathbf{x})\big)$ is always nonnegative and is zero if and only if $Q(\mathbf{x}) = p(\mathbf{x})$ [52], it can be readily noticed that the upper bound in (50) achieves an exact value of $\mathcal{F}_{\text{Helmholtz}}$ precisely when $Q(\mathbf{x}) = p(\mathbf{x})$. Moreover, it follows from (54) that the original ("target") distribution $p(\mathbf{x})$ is recovered as a distribution that minimizes the variational free energy, i.e.,

$$p(\mathbf{x}) = \arg\min_{Q(\mathbf{x})} \mathcal{F}(Q(\mathbf{x})). \tag{56}$$

Thus, minimizing the variational free energy $\mathcal{F}(Q)$ over an unconstrained set of $Q(\mathbf{x})$ distributions is an exact procedure for computing $\mathcal{F}_{\text{Helmholtz}}$. But the computational complexity related to this kind of optimization is, of course, also excessive for large values of K, rendering this approach totally intractable as such.

Basically, at least two strategies exist for obtaining practical algorithms for computing the Helmholtz free energy approximately: i) minimizing $\mathcal{F}(Q)$ over a restricted class of variational distributions and ii) modifying the functional $\mathcal{F}(Q)$ itself so that the optimization over the variational distributions becomes computationally less demanding. The practical algorithms can be obtained on the basis of one of these strategies or by using both of them. Computational algorithms that are obtained on the basis of the first strategy are generally called *mean field* (MF) algorithms [217]. In particular, a minimization of $\mathcal{F}(Q)$ over a class of fully factorized trial distributions leads to the solution that is often referred to as a *naive mean field* solution [263]. On the other hand, the *Bethe* method, whose origins date back to 1935 [31], is obtained by modifying the variational entropy functional $\mathcal{H}(Q)$ and setting constraints on the variational distributions $Q(\mathbf{x})$ [265]. So, recapping, computationally attractive algorithms for approximating the Helmholtz free energy are obtained by casting the exact computational procedure as a solution to the optimization problem and then solving the perturbed version of this optimization problem.

The utility of a variational free energy minimization (VFEM) framework is not limited to facilitating the evaluation of the Helmholtz free energy, but it equally applies to a broad class of computational problems in the context of nonphysical systems that are frequently encountered in the areas of statistical signal processing and digital communications. Such problems must often deal with the computation of the probabilities of some hidden variables given the observations. Typically, these computations consist of marginalization of some global function of all (or a large number of) variables.

In general, there are K marginal functions $p_i(x_i)$, associated with the joint pdf $p(\mathbf{x})$, that are defined as

$$p_i(x_i) \triangleq \sum_{\mathbf{x}:x_i} p(\mathbf{x}). \tag{57}$$

The notation used above should be interpreted so that the value of $p_i(X_i)$ is obtained by summing the value of $p(\mathbf{x})$ over all configurations of variables that have $x_i = X_i$.

From the computational procedures perspective, there is a direct analogy between the log-partition function in physics and the marginal function $p_i(x_i)$ in probability theory—both are obtained through the process of *marginalization*. The key idea behind the VFEM algorithm when applied to computation of $\mathcal{F}_{\text{Helmholtz}}$ is to find the trial distribution $Q(\mathbf{x})$ which resembles closely the Boltzmann distribution $p(\mathbf{x})$ and, at the same time, makes the marginalization required for the computation of the relevant upper bound (right hand side of (50)) computationally feasible. Similarly, the key insight

behind the *variational inference* in probabilistic models is that the exact marginal functions $p_i(x_i)$ can be approximated by the marginals

$$q_i(x_i) = \sum_{\mathbf{x}:x_i} Q(\mathbf{x}), \tag{58}$$

where the trial distribution $Q(\mathbf{x})$ is optimized so that it resembles the target distribution $p(\mathbf{x})$ as closely as possible. This optimization is achieved by minimizing the variational free energy $\mathcal{F}(Q)$ with respect to $Q(\mathbf{x})$, under the given constraints on $Q(\mathbf{x})$. On the other hand, the minimization of $\mathcal{F}(Q)$ amounts to the minimization of the KL divergence $D\big(Q(\mathbf{x})\|p(\mathbf{x})\big)$—a "distance" measure that is often used to measure differences between two distributions.

The VFEM framework presented above can naturally be extended to include the joint parameter estimation and variational inference, as already demonstrated in the previous section via some examples. Similarly to the procedures outlined above, the dual representations of convex/concave functions can be used to transform estimation algorithms into optimization problems which are then tuned into approximations by somehow relaxing the optimizations involved. As in the context of the variational formulation of the Helmholtz free energy, this relaxation has two main consequences: i) it simplifies the computations involved and ii) it yieds a bound, both of which are desirable properties. The formal treatment of the joint estimation and inference within the VFEM framework will be introduced in the following chapters when found relevant.

3 Statistical physics perspective of turbo receivers

In this chapter, the problem of joint detection and decoding via turbo-processing in known channels is addressed by looking at it from the statistical physics perspective. Based on the VFEM framework, it will be shown that the well-known linear SISO detectors, such as the linear channel matched filter (LCMF)-based SISO detectors [140] and linear minimum mean square error (LMMSE) SISO detectors [238], can be obtained as solutions to the variational optimization problem.

As already noted, the variational inference algorithms have their origin in statistical physics [36] but have, in recent years, been a subject for active research and development also within neural computation, artificial intelligence, and machine learning communities [125, 126, 265, 263, 160]. In addition, methods of statistical physics have been applied the analysis of the performance of CDMA systems in a number of papers, for example in [97, 228]. An extension of these into the realm of nonlinear soft interference cancellation (SIC) receivers in the context of CDMA systems was obtained in [145].

The contribution [145] can essentially be regarded as a parallel work to the one reported in this chapter, while both of these works were pursued independently. Though conceptually similar approaches were used, details in algorithms and different focuses in analysis may be used to differentiate these independent contributions from one another. In addition to these, a set of bit-level SISO detectors, dealing directly with the posterior distributions of the information *bits*, was derived for Gray-coded M-ary quadrature amplitude modulated (M-QAM) symbols via variational optimization in [146].

The variational optimization methods, in the context of turbo receivers, will be shown to provide an interesting link between the 'optimal' APP demodulator and the linear SISO detectors, characterizing the close relationship between them. Further, the new formulation of the linear SISO detectors has even more far-reaching implications since such a formulation essentially supplements the earlier contributions [265, 159, 110], which have connected the BP algorithm with the variational optimization methodology. Altogether, these results indicate that the many ideas underlying turbo receivers can be understood, unified, and generalized within the formalism of variational optimization.

In [137], clustering and stretching techniques were introduced as effective means to

modify a factor graph with cycles in such a way that some or all the cycles may be eliminated (further information can be found in [51]). Such modifications are obtained by increasing the complexity of local functions and/or the domains of the variable nodes. The modified factor graph typically entails improved performance, although often at a great expense in overall complexity. Importantly, the local function factorization via variational optimization as proposed in this chapter may also be seen as a generic low complexity method to eliminate some or all cycles in the fraction of the factor graph describing the channel behavior. This, together with clustering and stretching techniques, can provide further flexibility in designing approximate receiver structures for the model at hand.

Moreover, the properties of the exact APP demodulator and linear SISO detectors are examined from the information divergence and the entropy viewpoints. Although the algorithms themselves are mostly known, this examination may be justified by noting that such fresh viewpoints can offer a new set of insights and opportunities. As an example, some extensions of the linear SISO detectors, providing an improved performance with an acceptable increase in complexity, are proposed in the context of MIMO systems.

This chapter is structured as follows. In Section 3.1, the turbo-processing principles are summarized and design criteria for SISO demodulation are addressed. In Section 3.2, the familiar linear SISO detectors are formulated as instances of the VFEM algorithm. In section 3.3, a discussion on the performance of the APP demodulators and linear SISO detectors in the light of information divergence is given. In Section 3.4, some examples of advanced receiver structures based on variational optimization are discussed. Then, complexities of the discussed receiver structures are considered at a general level in Section 3.5, and numerical simulation results are given in Section 3.6. Finally, conclusions are drawn in Section 3.7.

3.1 Turbo processing and variational optimization

The turbo-processing principle is first summarized, emphasizing the factors which make turbo receivers inherently suboptimal. Then, an information theoretic justification for APP demodulators as effective building blocks of the turbo receivers is provided. Finally, the VFEM framework as an avenue to low-complexity alternatives is proposed.

3.1.1 Turbo-processing principle

An optimal decoding for the coded MIMO signal is given as

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} p(\mathbf{s}|\mathbf{r}), \tag{59}$$

where $p(\mathbf{s}|\mathbf{r})$ represents the *a posteriori* probability distribution of the codewords, and it can be expressed as

$$p(\mathbf{s}|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s})I_C(\mathbf{s})}{\sum_{\mathbf{s}} p(\mathbf{r}|\mathbf{s})I_C(\mathbf{s})}. \tag{60}$$

In (60), $p(\mathbf{r}|\mathbf{s})$ denotes the conditional probability density function (pdf) of $\mathbf{r}$ given $\mathbf{s}$, and $I_C(\mathbf{s})$ denotes the indicator (characteristic) function for the code $\mathcal{C}$. In a case where a separate channel code ($\mathcal{C}_n$) is applied to different transmitter antenna signals (e.g., a multiuser MIMO case), the indicator function can be decomposed as

$$I_C(\mathbf{s}) = I_{C_1}(\mathbf{s}_1)I_{C_2}(\mathbf{s}_2)\cdots I_{C_N}(\mathbf{s}_N), \tag{61}$$

where $I_{C_n}(\mathbf{s}_n)$ denotes an indicator function for the channel code imposed on the data to be transmitted from transmitter antenna n. The optimal receiver is in most cases prohibitively complex since the symbol sequence probabilities have to be computed for all permissible sequences.

Following the formulation in [170], it may be possible to find an iterative minimum cross-entropy (MCE) algorithm that applies only a subset of constraints (either channel or some of the code constraints) at a time, and yet is guaranteed to converge to the distribution $p(\mathbf{s}|\mathbf{r})$. But even if this could be done, the complexity would not be essentially reduced since the symbol sequence probabilities must be updated at each iteration for all permissible sequences, taking values in the Cartesian product space $\mathcal{S}^{KN}$, where K denotes the length of $\mathbf{s}_n$. In turbo processing, the complexity issue is addressed in two different ways: i) by iterative processing where the symbol sequence distribution is updated iteratively in separate demodulation and decoding blocks that exchange probabilistic information with each other in the form of *extrinsic* information, and ii) by assuming that the symbol probabilities at the output of each block are independent in distinct symbol intervals. In consequence of the assumed independence between the symbols, the symbol sequence distribution at the output of the demodulator and the decoder takes a fully factorized form, where each factor is a function of a single symbol only. Since this can be regarded as the most essential assumption made in turbo-processing [170], its validity may eventually determine how well any particular paradigm of the turbo receiver performs with respect to the optimal receiver.

Basically, there are a number of ways to improve the validity of the symbol independence assumption. The role of an interleaving and deinterleaving between the demodulation and decoding functions is crucial in this respect. Secondly, the exchange of only the *extrinsic* information (information difference between the input and output of each block) between different blocks improves the independence of the symbols. Finally, the method used to compute the soft output distribution at each stage of the turbo processing may have a great impact on the validity of the independence assumption as well as on the convergence properties of the whole iterative receiver.

3.1.2 SISO demodulation, information divergences, and variational optimization

The basic problem in SISO demodulation as well as in SISO decoding is finding a factorization which approximates the symbol sequence distribution $p(\mathbf{s}|\mathbf{r})$ as closely as possible. An intuitively appealing choice is to approximate the symbol sequence distribution with a product of marginal distributions. In fact, this used to be a common practice in the appearance of turbo receivers [170] and is still an only known feasible method to implement the SISO decoding. According to this strategy, the SISO demodulator approximates the symbol sequence distribution as follows:

$$p_r(\mathbf{s}) \stackrel{\triangle}{=} \frac{1}{p(\mathbf{r})} p(\mathbf{r}|\mathbf{s}) P(\mathbf{s}) \approx \prod_n \prod_k p_r(s_n(k)), \tag{62}$$

where $P(\mathbf{s})$ denotes the *a priori* distribution of $\mathbf{s}$.[9] The marginals are computed as $p_r(s_n(k)) = \sum_{\mathbf{s} \setminus s_n(k)} p_r(\mathbf{s})$, where the summation is performed over all symbol sequences $\mathbf{s}$ that are consistent with $s_n(k)$ (called "summary for $s_n(k)$" in [137]). In the literature, computation of marginal distributions is often called *exact inference* as opposed to *approximate inference*, which refers to approximate methods of calculating the marginals. It is important to note, that the exact inference, performed in different processing blocks of the turbo receiver, does not imply an optimality of the receiver, since the optimal receiver must always manipulate the whole symbol sequence distributions, as already pointed out earlier.

In accordance with the turbo principle, the *a priori* distribution of the symbol se-

[9] In general, $P(\mathbf{s})$ is known by the receiver since it can be obtained from the coding rule. However, using exact values for $P(\mathbf{s})$ entails exponentially growing processing and data storage requirements. Hence, distribution $P(\mathbf{s})$ is often approximated by means of extrinsic output from the SISO decoder.

quence $P(\mathbf{s})$ is assumed to be fully factorized uniform distribution at the first iteration, whereas at the following iterations it is postulated to be

$$P(\mathbf{s}) = \prod_n \prod_k \lambda_{2,n,k}(s_n(k)), \tag{63}$$

where $\lambda_{2,n,k}(s_n(k))$ represents the extrinsic information on the symbol $s_n(k)$ delivered by the channel decoder via the interleaver. Due to these simplifications, the distribution $p_r(\mathbf{s})$ is only an approximation of $p(\mathbf{s}|\mathbf{r})$ when the channel encoding is employed. The focus of this chapter is, however, on the latter approximation in (62), i.e., on the factorization performed by the SISO demodulator. A theoretical justification for the factorization in (62) can be obtained by using the concept of *information divergence*, as shown in the following lemma.

Lemma 1 *The product of marginal distributions can be interpreted as a factorization that is obtained by minimizing the Kullback-Leibler (KL) divergence between the object distribution $p_r(\mathbf{s})$ and any fully factorized distribution, i.e.,*

$$\prod_n \prod_k p_r(s_n(k)) = \arg\min_{Q(\mathbf{s})} D\big(p_r(\mathbf{s}) \| Q(\mathbf{s})\big), \tag{64}$$

where the trial distribution $Q(\mathbf{s})$ is constrained to be a product of arbitrary "factor" probabilities $q_{n,k}$, i.e.,

$$Q(\mathbf{s}) = \prod_n \prod_k q_{n,k}. \tag{65}$$

Proof. See Appendix 1. □

The KL-divergence can be regarded as a distance measure between two distributions although it is generally asymmetrical, i.e., $D(p_r\|Q) \neq D(Q\|p_r)$. In fact, the KL-divergence is a special case in a large family of information divergences, called α-divergences [6].

The turbo receiver based on the factorization (62) can be interpreted as an instance of the SP algorithm, operating on the factor graph where the demodulator and the decoders are represented by cycle-free Wiberg-type graphs [137, 159]. The symbol *a posteriori* probabilities (APPs) $p_r(s_n(k))$ are efficiently computed by the standard forward-backward processing (BCJR) algorithm [16], with a complexity which scales exponentially as a function of LN. When the channel memory length or the number of transmitter antennas or both increase, even this APP demodulator becomes computationally infeasible, and one is forced to pursue other approximations. A pertinent

question at this point is whether the minimization of $D\big(Q(\mathbf{s})\|p_r(\mathbf{s})\big)$, another form of the information divergence, produces a computationally more feasible solution. The answer is in the affirmative, as will be shown in the next section.

Meanwhile, considering the situation where the channel convolution matrix $\mathbf{H}$ is known by the receiver, let us rewrite the distribution $p_r(\mathbf{s})$ in the form of Boltzmann distribution as follows:

$$p_r(\mathbf{s}) = \frac{1}{Z} e^{-\frac{1}{T}\mathcal{E}(\mathbf{s})}, \tag{66}$$

where the posterior log likelihood $\mathcal{E}(\mathbf{s})$, defining the energy of the system at hand, is given by

$$\mathcal{E}(\mathbf{s}) = -\mathbf{r}^H\mathbf{H}\mathbf{s} - \mathbf{s}^H\mathbf{H}^H\mathbf{r} + \mathbf{s}^H\mathbf{H}^H\mathbf{H}\mathbf{s} - T\ln P(\mathbf{s}). \tag{67}$$

As usual, Z is a normalizing constant (called the partition function in statistical physics) and the temperature T corresponds (and is nominally set) to the variance σ_v^2 of the thermal noise.[10] For Boltzmann distributions, the minimization of the KL divergence $D(Q(\mathbf{s})\|p_r(\mathbf{s}))$ amounts to the minimization of the corresponding variational free energy. Thus, the minimization of $D(Q(\mathbf{s})\|p_r(\mathbf{s}))$, with the trial distribution $Q(\mathbf{s})$ assuming a fully factorized form, falls into the purview of the VFEM framework. It is this statistical physics perspective of the turbo receivers that will be further explored in the sequel.

3.2 Formulation of linear SISO detectors as instances of VFEM algorithm

In this section, two classes of existing linear SISO detectors are formulated as devices which aim to minimize the variational free energy over the family of fully factorized distributions. In particular, the first class of receivers handle the information symbols as discrete valued unknown variables. In this case, the linear SISO detector is recovered by using the conventional mean field approach. Meanwhile, the second class of linear SISO detectors model the information symbols as random Gaussian variables, and, hence, the probabilistic calculations within this class fit perfectly into the Bayesian formalism.

[10] The noise variance σ_v^2 is deliberately replaced by the temperature T for future reference.

3.2.1 Linear channel matched filter-based SISO detectors

A new formulation for the LCMF SISO detectors, proposed earlier by several authors (see, e.g., [140, 238, 194]), is given by the following proposition.

Proposition 1 *Consider the minimization of* $\mathcal{F}(Q) = -\ln Z + \sum_{\mathbf{s}} Q(\mathbf{s}) \ln \frac{Q(\mathbf{s})}{p_r(\mathbf{s})}$ *subject to constraint that the variational distribution* $Q(\mathbf{s})$ *is fully factorized. The fixed point equations for solving this VFEM problem results in the output distribution of the LCMF SISO detector [140, 238], given as* $Q_{MF}(\mathbf{s}) = \prod_n \prod_k q_{n,k}$*, where*

$$q_{n,k} = \lambda_{2,n,k} \, \frac{1}{\gamma_{n,k}} \exp\left[-\frac{E_{n,k}}{T} \big| s_n(k) - \hat{s}_n(k) \big|^2 \right]. \tag{68}$$

Here, $\lambda_{2,n,k}$ *denotes an extrinsic information delivered by the SISO decoder,* $\gamma_{n,k}$ *is a normalizing constant, and* $\hat{s}_n(k)$ *denotes the output of the linear SISO detector given as*

$$\begin{aligned} \hat{s}_n(k) &= \frac{1}{E_{n,k}} \mathbf{e}_{n,k}^H \mathbf{H}^H \big(\mathbf{r} - \mathbf{H}\overline{\mathbf{s}} + \mathbf{H}\mathbf{e}_{n,k} \overline{s}_n(k) \big) \\ &= \mathbf{c}_{n,k,CMF}^H \big(\mathbf{r} - \mathbf{H}\overline{\mathbf{s}} + \mathbf{H}\mathbf{e}_{n,k} \overline{s}_n(k) \big), \end{aligned} \tag{69}$$

where $E_{n,k} = \mathbf{e}_{n,k}^H \mathbf{H}^H \mathbf{H} \mathbf{e}_{n,k}$ *is the energy of the symbol* $s_n(k)$ *at the receiver (when the transmitted signal power is normalized to one) and*

$$\mathbf{c}_{n,k,CMF} \triangleq \mathbf{H}\mathbf{e}_{n,k} \tag{70}$$

denotes the coefficient vector of the channel matched filter. Furthermore, $\mathbf{e}_{n,k}$ *denotes a unit vector defined as*

$$\mathbf{e}_{n,k} = \big[\mathbf{0}_{1\times((n-1)K+k-1)}, 1, \mathbf{0}_{1\times((N-n+1)K-k)} \big]^T, \tag{71}$$

and the vector of the soft symbols $\overline{\mathbf{s}}$ *is defined as*

$$\overline{\mathbf{s}} = \big[\overline{\mathbf{s}}_1^T, \overline{\mathbf{s}}_2^T, \ldots, \overline{\mathbf{s}}_N^T \big]^T \in \mathbb{C}^{NK}, \tag{72}$$

$$\overline{\mathbf{s}}_n = \big[\overline{s}_n(1), \overline{s}_n(2), \ldots, \overline{s}_n(K) \big]^T \in \mathbb{C}^K, \tag{73}$$

where the value of the soft symbol $\overline{s}_n(k)$ *is obtained as*

$$\overline{s}_n(k) = E_{q_{n,k}} \big[s_n(k) \big] = \sum_{\alpha_i \in \mathcal{S}} \alpha_i q_{n,k} \big(s_n(k) = \alpha_i \big). \tag{74}$$

In (74), $E_{q_{n,k}}[\cdot]$ *denotes expectation under the distribution* $q_{n,k}$*.*

Proof. See Appendix 1. □

According to the turbo principle, the extrinsic output distribution of the LCMF SISO detector can be defined as

$$\Lambda_1(\mathbf{s}) = \prod_n \prod_k \lambda_{1,n,k}(s_n(k)), \tag{75}$$

where $\lambda_{1,n,k}(s_n(k)) \triangleq q_{n,k}/\lambda_{2,n,k}$. Thus, the soft output of the standard LCMF SISO detector can be interpreted as a naive MF approximation to the objective function $p_r(\mathbf{s})$. This novel formulation provides new insight into existing algorithms and suggests some enhancing adjustments for them as well. In particular, the following points are worth noting.

1. In the presence of perfect soft symbol feedback, i.e., $\overline{\mathbf{s}} = \mathbf{s}$, the ISI is perfectly removed and the receiver achieves the matched filter bound.
2. The soft symbol estimates, which are used in the soft interference cancellation, are computed by exploiting the full APP information on the channel symbols. The full APP information is the sum of the information induced by the coding, the estimate of which is provided by the channel decoder in the form of extrinsic information $\lambda_{2,n.k}$, and the ISI channel induced information, embodied by the soft output of the detector, i.e., $\ln q_{n,k} = \ln \lambda_{2,n,k} + \ln \lambda_{1,n,k}$. This contrasts with most of the earlier contributions dealing with the LCMF SISO detectors, where only the extrinsic output of the decoder is used for the soft interference cancellation. However, a similar remark concerning the use of full APP information in LCMF detectors was made in [248] by using ad hoc reasoning.
3. Since the soft interference cancellation exploits the full APP information, an improvement may be obtained whenever either the coding-induced information or the channel-induced information is updated. Consequently, the convergence rate of turbo receivers based on the linear SISO detectors can, at least in some cases, be accelerated by making also the SISO detector itself iterative, i.e., by updating the channel induced information $\lambda_{1,n,k}$ more often than the coding induced information $\lambda_{2,n,k}$. Another interesting variant of the linear SISO detector is obtained by dividing the feedback filter into two parts, one cancelling the precursor and the other cancelling the postcursor [249, 153]. The soft input to the postcursor cancellation filter can be updated on-line while the input to the precursor cancellation filter is computed using information from the previous iteration.

4. Because the minimization of the KL-divergence in $\mathcal{F}(Q)$ takes place over the first argument instead of the second one, the mean field optimization does not fall into the purview of standard information geometry [170, 7]. Moreover, the variational free energy $\mathcal{F}(Q)$ is not a convex function of $q_{n,k}$, entailing the possible existence of multiple local minima to which the mean field algorithm can converge to. However, the risk of being occasionally trapped into the bad local minimum can be reduced by applying the mean field annealing (MFA) technique [33]. A key idea behind this technique is to vary the value of the temperature T as the iterations evolve, starting from a high temperature value and then cooling the system down along with iterations. It has been successfully applied, for example, to the traveling salesman problem [22]. There the MFA theory was shown to bring out some of the essential features of the problem.
5. A major drawback of the naive mean field approximation is that it over-estimates the entropy of the variational distribution. This is due to fact that the actual probabilistic interactions between the neighboring symbols are approximated with their average influence (mean field) on any specific symbol, i.e., the symbols are assumed to be statistically independent under the variational distribution. This is demonstrated with a simple example model for which factor graph descriptions using exact and mean field factorizations are presented in Fig. 3. Thus, an improvement on the mean field solution may be obtained by defining a more sophisticated structure for the trial distribution $Q(\mathbf{s})$ or by modifying the variational free energy term $\mathcal{F}(Q)$ itself so that the probabilistic interactions between symbols can be taken into account in a more sophisticated way. Well-known examples of the latter approach are the Bethe method resulting in the BP algorithm [265] as well as the Kikuchi and other cluster variational approximation methods resulting in a variety of generalized BP algorithms [265, 196, 197].

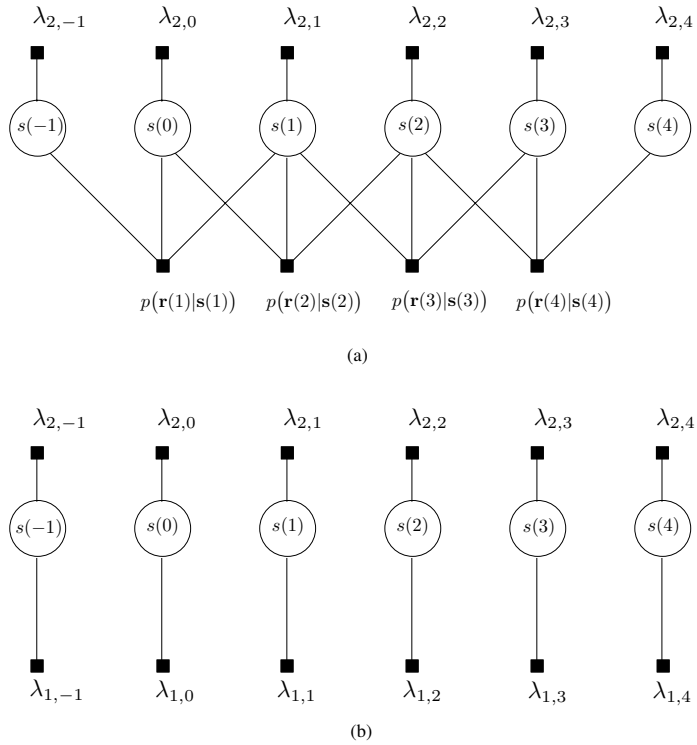

Fig 3. A factor graph description for an example model using (a) an exact factorization and (b) a mean field factorization. Only one transmitter antenna is assumed, and so the following notational convention is used: $\lambda_{i,1,k}(s(k)) \equiv \lambda_{i,k}\ \forall i,k$**.**

The above MF factorization is obtained under the requirement that the data symbols take values in a discrete domain, as they actually do in reality. However, let us consider next the case where this requirement is relaxed, with a hope of achieving more efficient detection at the receiver.

3.2.2 Linear MMSE SISO detectors

The linear MMSE SISO detectors have been proposed earlier by several authors [238, 239, 211, 221, 1]. In the following proposition, a statistical physics interpretation of these well-known detectors is provided by applying the VFEM framework. Towards

this end, the requirement that the data symbols take values only in a discrete space is relaxed. Instead, they are assumed to take on values in a continuous complex valued space, i.e., $\mathbf{s} \in \mathbb{C}^{NK}$. In addition, the data symbols are regarded as random variables with independent Gaussian *a priori* probability density function (pdf).[11] Accordingly, the postulated *a priori* pdf of $\mathbf{s}$ is given as $P(\mathbf{s}) = \mathcal{CN}(\mathbf{s}, \overline{\mathbf{s}}, \mathbf{\Sigma}_i)$, where $\overline{\mathbf{s}}$ is a mean vector and $\mathbf{\Sigma}_i = \text{diag}(\bar{\sigma}_i)$ is a diagonal covariance matrix where the vector of diagonal elements is given by $\bar{\sigma}_i = [\sigma^2_{i,1,1}, \cdots, \sigma^2_{i,1,K}, \cdots, \sigma^2_{i,N,K}]^T$ (the subscript i stands for the "input to the SISO detector").

Proposition 2 *Consider the minimization of* $\mathcal{F}(Q) = -\ln Z + \sum_{\mathbf{s}} Q(\mathbf{s}) \ln \frac{Q(\mathbf{s})}{p_r(\mathbf{s})}$ *subject to constraint that the variational distribution* $Q(\mathbf{s}) = \mathcal{CN}(\mathbf{s}, \hat{\mathbf{s}}, \mathbf{\Sigma}_o)$*, where the mean vector* $\hat{\mathbf{s}}$ *and the diagonal covariance matrix* $\mathbf{\Sigma}_o = diag(\bar{\sigma}_o)$ *are unknown parameters subject to minimization, with* $\bar{\sigma}_o$ *defined as*

$$\bar{\sigma}_o = [\sigma^2_{o,1,1}, \cdots, \sigma^2_{o,1,K}, \cdots, \sigma^2_{o,N,K}]^T.$$

This minimization results in the output distribution of the linear MMSE SISO detector [239], defined as $Q_{MMSE}(\mathbf{s}) = \mathcal{CN}(\mathbf{s}, \hat{\mathbf{s}}, \mathbf{\Sigma}_o)$*, where*

$$\hat{\mathbf{s}} = \overline{\mathbf{s}} + \mathbf{\Sigma}_i \mathbf{H}^H (\mathbf{H}\mathbf{\Sigma}_i \mathbf{H}^H + \sigma_v^2 \mathbf{I})^{-1} (\mathbf{r} - \mathbf{H}\overline{\mathbf{s}}) \tag{76}$$

$$\sigma^2_{o,n,k} = \frac{\sigma_v^2}{\mathbf{e}^H_{n,k} \big(\mathbf{H}^H \mathbf{H} + \sigma_v^2 \mathbf{\Sigma}_i^{-1}\big) \mathbf{e}_{n,k}}. \tag{77}$$

Proof. See Appendix 1. □

Since turbo processing entails the exchange of extrinsic information between the detector and decoder blocks, the *extrinsic* information corresponding to the symbol $s_n(k)$ at the output of the MMSE SISO detector has to be extracted from $Q_{\text{MMSE}}(\mathbf{s})$ by using some mapping. As proposed in [239], a viable mapping is obtained simply by setting $\overline{s}_n(k) = 0$ and $\sigma^2_{i,n,k} = 1$. In doing so and additionally utilizing the knowledge of the symbol constellation points, the *extrinsic* information concerning the symbol $s_n(k)$ can be expressed as

$$\lambda_{1,n,k}(\alpha_i) = \mathcal{CN}\Big(s_n(k) = \alpha_i, \hat{s}_n^{(\text{extr})}(k), \big(\sigma^2_{o,n,k}\big)^{(\text{extr})}\Big), \tag{78}$$

[11] A pdf of a complex Gaussian random variable vector $\mathbf{x}$ with K elements is given by $\mathcal{CN}(\mathbf{x}, \mathbf{m}_x, \mathbf{\Sigma}_x) = \frac{1}{\pi^K \det(\Sigma_x)} e^{-(\mathbf{x}-\mathbf{m}_x)^H \mathbf{\Sigma}_x^{-1} (\mathbf{x}-\mathbf{m}_x)}$, where $\mathbf{m}_x$ denotes the mean vector and $\mathbf{\Sigma}_x$ denotes the covariance matrix.

where

$$\hat{s}_n^{(\text{extr})}(k) = \mathbf{e}_{n,k}^H \boldsymbol{\Sigma}_i \mathbf{H}^H \big(\mathbf{H}\boldsymbol{\Sigma}_i\mathbf{H}^H + \sigma_v^2\mathbf{I} + (1-\sigma_{i,n,k}^2)\mathbf{H}\mathbf{e}_{n,k}\mathbf{e}_{n,k}^H\mathbf{H}\big)^{-1} \cdot \big(\mathbf{r} - \mathbf{H}\overline{\mathbf{s}} + \mathbf{H}\mathbf{e}_{n,k}\overline{s}_n(k)\big) \tag{79}$$

$$\big(\sigma_{o,n,k}^2\big)^{(\text{extr})} = \frac{\sigma_v^2}{\mathbf{e}_{n,k}^H\mathbf{H}^H\mathbf{H}\mathbf{e}_{n,k} + \sigma_v^2} = \frac{\sigma_v^2}{E_{n,k} + \sigma_v^2} = \frac{1}{\text{SNR}_{n,k}+1}, \tag{80}$$

and $\text{SNR}_{n,k}$ denotes the signal-to-noise ratio at the kth symbol interval at the receiver, i.e., $\text{SNR}_{n,k} = E_{n,k}/\sigma_v^2$. Another method for defining the extrinsic information in the context of linear SISO detectors has been proposed in [148].

As shown above, the linear MMSE SISO detector can be formulated as an instance of what is here referred to as the Bayesian mean field (BMF) approximation. This novel interpretation yields new insight into the existing algorithms, as stated by the following remarks.

1. The Hessians of $\mathcal{F}(Q)$ with respect to mean vector, i.e., $\partial^2\mathcal{F}(Q)/\big(\partial\hat{\mathbf{s}}\partial\hat{\mathbf{s}}^H\big)$, and with respect to variance vector $\bar{\sigma}_o$, i.e., $\partial^2\mathcal{F}(Q)/\big(\partial\bar{\sigma}_o\partial\bar{\sigma}_o^H\big)$, are positive semidefinite matrices implying that $\mathcal{F}(Q)$ is convex in terms of both the mean vector and the variance vector. Due to convexity, the variational free energy in this case has only one minimum, which is a global minimum, and it is obtained when $\hat{\mathbf{s}}$ and $\boldsymbol{\Sigma}_o$ are defined as in (76) and (77).
2. It can be easily verified that

$$\arg\min_{\hat{\mathbf{s}},\boldsymbol{\Sigma}_o} D\big(\mathcal{CN}(\mathbf{s},\hat{\mathbf{s}},\boldsymbol{\Sigma}_o)\|p_r(\mathbf{s})\big) = \arg\min_{\hat{\mathbf{s}},\boldsymbol{\Sigma}_o} D\big(p_r(\mathbf{s})\|\mathcal{CN}(\mathbf{s},\hat{\mathbf{s}},\boldsymbol{\Sigma}_o)\big). \tag{81}$$

   This relation is a direct consequence of the joint Gaussianity of $\mathbf{r}$ and $\mathbf{s}$. In fact, the conditional pdf $p_r(\mathbf{s})$ is also Gaussian with the mean vector given by (76) and the covariance matrix given by $\boldsymbol{\Sigma}_{\mathbf{s}|\mathbf{r}} = \sigma_v^2\big(\mathbf{H}^H\mathbf{H} + \sigma_v^2\boldsymbol{\Sigma}_i^{-1}\big)^{-1}$.
3. The output distribution $Q_{\text{MMSE}}(\mathbf{s})$ can be expressed in a factorized form as

$$Q_{\text{MMSE}}(\mathbf{s}) = \prod_n\prod_k \mathcal{CN}\big(s_n(k),\hat{s}_n(k),\sigma_{o,n,k}^2\big), \tag{82}$$

   where, based on (64) and (81), each factor distribution can alternatively be obtained as $\mathcal{CN}\big(s_n(k),\hat{s}_n(k),\sigma_{o,n,k}^2\big) = \int_{\mathbf{s}_{\setminus s_n(k)}} p_r(\mathbf{s})d\mathbf{s}_{\setminus s_n(k)} = p_r(s_n(k))$ (the multidimensional integration is computed over all symbols except the symbol $s_n(k)$). Consequently, analogously to the APP demodulator, the output pdf of the MMSE SISO detector can be expressed as a product of marginal functions of the target distribution $p_r(\mathbf{s})$.

4. The Bayesian mean field approach postulates the existence of the *a priori* pdf $P(\mathbf{s})$ of data vector $\mathbf{s}$. But it is not directly guided by the VFEM framework how this *a priori* pdf should be specified in different phases of turbo-processing. At the first iteration, a reasonable choice is the Gaussian pdf with zero mean and unit-variance, i.e., $P(\mathbf{s}) = \mathcal{CN}(\mathbf{s}, \mathbf{0}, \mathbf{I})$, leading to the standard MMSE equalizer with soft output. Alternatively the *a priori* pdf could be assumed to be flat over the values of $\mathbf{s}$ where the likelihood function $p(\mathbf{r}|\mathbf{s})$ has a nonvanishing support, essentially leading to the zero-forcing (ZF) equalizer. At the later iterations, an intuitively appealing and commonly made choice (see, e.g., [238, 239]) for the *a priori* pdf is the Gaussian pdf whose mean vector $\overline{\mathbf{s}}$ and covariance matrix $\mathbf{\Sigma}_i$ are computed by exploiting the extrinsic information obtained as a feedback from the channel decoder, i.e., $\overline{s}_n(k) = E_{\lambda_{2,n,k}}[s_n(k)]$ and $\sigma^2_{i,n,k} = E_{\lambda_{2,n,k}}\left[|s_n(k)|^2\right] - |\overline{s}_n(k)|^2$. Moreover, a computationally more attractive implementation is obtained by setting $\mathbf{\Sigma}_i = \frac{1}{KN}\sum_n \sum_k \sigma^2_{i,n,k} \times \mathbf{I}_{NK}$ and assuming that the channel matrix is cyclic (e.g., via the use of the cyclic prefixed transmission) [254, 253]. It may, however, be advantageous from the performance point of view to calculate the mean and covariance of $\mathcal{CN}(\mathbf{s}, \overline{\mathbf{s}}, \mathbf{\Sigma}_i)$ in terms of the full symbol APP information, instead of only exploiting the extrinsic information. This latter approach could be justified by the fact that the MMSE SISO detector approaches the LCMF SISO detector when the reliability of the soft symbol decisions increases.
5. Unlike what has previously been proposed in the literature (see, e.g., [238, 239]), the optimal variance of the *extrinsic* output of the MMSE SISO detector (from the VFEM point of view) is exclusively determined by the instantaneous SNR. Previously, the variance of the residual interference has also been taken into account [238, 239] in defining the extrinsic output of the SISO detector. However, the performance of the associated turbo receiver seems to be quite robust, according to numerical simulations, to the small variations in the output variance.

So far, a theoretical justification for various existing SISO blocks, such as the APP demodulator, the LCMF SISO detector, and the MMSE SISO detector, has been obtained by minimizing the KL divergence under different conditions and assumptions. Thus, the KL divergence can be regarded as a unifying element in deriving and characterizing various SISO blocks for turbo processing. Next, a step forward is taken by exploiting the properties of the KL divergence in order to gain new insights into the performance of these SISO blocks.

3.3 Comparison of SISO blocks in the light of information divergences

As shown in the previous sections, both the APP demodulators and the linear SISO detectors can be interpreted as devices that approximate the symbol sequence distribution with a fully factorized distribution such that an information loss is minimized in some sense. In this light, it seems that the basic difference between these devices is not so much the amount of structure they model, but rather the measure of information loss they minimize. Since different divergence measures tend to prefer different types of solutions, it is important to identify the conditions under which each of the approximations is most favorable. An aim of this section is to discuss, mostly at an intuitive level, the insight one can gain into the convergence properties of turbo receivers, simply by examining the properties of the information divergences that their associated SISO detectors try to minimize.

The MF approximation is characterized by the fact that it finds a single, most massive mode of the target distribution while excluding the less plausible modes.[12] This is because the KL divergence $D(Q(\mathbf{s})\|p_r(\mathbf{s}))$ is forced to zero in the regions where the target distribution $p_r(\mathbf{s})$ has vanishing density. In contrast, minimizing $D(p_r(\mathbf{s})\|Q(\mathbf{s}))$ will favor $Q(\mathbf{s})$ distributions which try to cover all modes of $p_r(\mathbf{s})$ although this involves assigning high probability to $Q(\mathbf{s})$ in areas of low probability under $p_r(\mathbf{s})$ [82]. Because of that, the KL divergence $D(p_r(\mathbf{s})\|Q(\mathbf{s}))$ is termed "inclusive" divergence in [82]. By analogy, the KL divergence $D(Q(\mathbf{s})\|p_r(\mathbf{s}))$ could be called an "exclusive" divergence since it tends to exclude all but the most massive mode of the target distribution. As an example, an arbitrary bimodal function $P(x)$ may be approximated by a Gaussian trial function $Q(x)$, as shown in Fig. 4(a), when the "exclusive" divergence $D(Q\|P)$ is used for optimization, while the result of the optimization may look like that of Fig. 4(b) when the "inclusive" divergence $D(P\|Q)$ is used as the optimization criterion.

[12] Strictly speaking, the mode of a probability distribution is the value of a random variable at which the probability density function attains its maximum value. When a pdf has multiple local maxima, it is, however, common to refer to all of such local maxima as modes of the distribution. That is also the case in this thesis. The massiveness of the (local) mode is defined as a probability mass within a suitable range of values around the (local) mode.

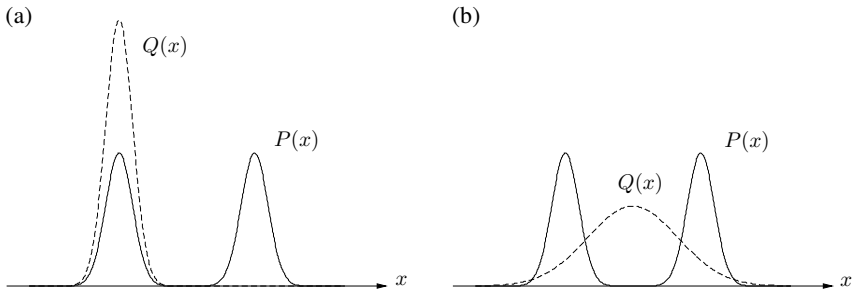

Fig 4. Approximating a bimodal function $P(x)$ by the Gaussian trial function $Q(x)$ with its mean and variance serving as the free parameters to be adjusted. The result of minimizing $D(Q\|P)$, as is suggested by the VFEM framework, is illustrated in part (a), while the result of minimizing $D(P\|Q)$ is illustrated in part (b).

The relevant question now is which one of these two different KL divergences is more favorable from the turbo processing point of view. Obviously, it is better to try to cover all modes of the target distribution, at least initially. This is because the LCMF SISO detectors can easily get trapped by a "wrong" mode of the target distribution, which is manifested by an error floor in their performance. This danger is naturally highest at the first iteration of the turbo processing, whereas at later iterations the existence of reliability information from the decoder typically makes the probability mass in $p_r(\mathbf{s})$ concentrate in one mode, rendering the LCMF SISO detector as a viable choice in these circumstances. These statements are backed up by EXIT chart analysis and numerical simulation results presented in [238]. Indeed, the convergence problems related to using the LCMF SISO detector as a building block of turbo receivers were clearly observed and brought out in [238], while turbo processing based on the APP demodulation was shown to represent a rather stable and fast-converging iterative detection strategy. Actually, these considerations may partly account for the fact that the APP demodulators are sometimes regarded as "optimal" SISO demodulators.

While the LCMF SISO detector seems to be in difficulties at early iterations, particularly in severe ISI channels, the MMSE SISO detector, which is based on the Bayesian

MF approximation, seems to work properly. This is because the postulated Gaussian *a priori* pdf of the data symbols makes the *a posteriori* pdf $p_r(\mathbf{s})$ also Gaussian. Hence, the target distribution in this Bayesian case is unimodal and convex, which greatly facilitates the search for the global minimum. In fact, the inclusive and exclusive KL divergences are equal in this Bayesian case.

Based on the above reasoning, it may be a good strategy to apply the exact APP or linear MMSE SISO detection at early iterations of the turbo processing and then, after initial convergence, to switch to the low-complexity LCMF detector. A related turbo receiver architecture was called a hybrid turbo detection scheme in [238]. In essence, the above considerations, together with the EXIT chart analysis of [238], provide a theoretical justification for the hybrid turbo detection schemes, for which some experimental results were presented in [238, 194]. In addition, a conceptually similar hybrid turbo detection scheme has been proposed for the CDMA system in [230].

3.4 Discussion on advanced variational methods with applications

As mentioned earlier, an improvement on the naive MF approximation may be obtained either by defining a more sophisticated structure for the trial distribution $Q(\mathbf{s})$ or by modifying the variational free energy term $\mathcal{F}(Q)$ itself. In this section, both of these approaches are elaborated to some extent. As an example of the former approach, a variational approximation technique called the structured mean field approximation [90] is considered first. It is based on the idea that the original probabilistic model is divided into substructures and the factorization within each substructure is pursued by using the inclusive KL divergence minimization as an optimization criterion (entailing exact marginalization), whereas the probabilistic interactions between substructures are modeled by appealing to the exclusive KL divergence.

As a concrete example, an advanced SISO detector for MIMO systems in the presence of frequency-selective channel is derived by applying the structured MF approximation. In effect, this system model fits perfectly into the class of models sometimes called *factorial* hidden Markov models (factorial HMMs) [90]. The HMMs are ubiquitous and often encountered also in the field of digital communications. Graphical modeling of HMMs, such as factor graph modeling, is closely related to modeling of HMMs by using a *trellis diagram*—a more commonly used tool in the area of com-

munications to describe the time-evolution of a finite-memory system. In essence, the information about the past is in both descriptions conveyed through a single discrete variable, called either the hidden state or the trellis state depending on the modeling concept.[13] As is well known, an exact algorithm for inferring the posterior probabilities of the hidden state variables is obtained via the forward-backward processing algorithm, known generally as the BCJR algorithm by the digital communications community.

The structured MF approximation in the context of factorial HMMs is based on a generalization where the hidden state of the model is factored into the multiple state variables and is therefore represented in a distributed manner. In the MIMO systems under consideration, a natural factorization is such that the dynamics of the independent data streams from different transmitter antennas are dealt with separately in the graphical description of the model. Specifically, the variational free energy $\mathcal{F}(Q)$ is minimized over the partially factorized distribution $Q(\mathbf{s}) = \prod_{n=1}^{N} Q^{(n)}(\mathbf{s}_n)$, where "per-branch" distributions $Q^{(n)}(\mathbf{s}_n)$ are defined in terms of the free parameters $q_{n,k}(\mathbf{s}_n(k))$ as follows:

$$Q^{(n)}(\mathbf{s}_n) = \prod_{k=1}^{K} T_k^{(n)}\big(\acute{\mathbf{s}}_n(k-1), s_n(k), \acute{\mathbf{s}}_n(k)\big) \prod_{k=1}^{K} q_{n,k}(\mathbf{s}_n(k)) \prod_{k=1}^{K} \lambda_{2,n,k}. \quad (83)$$

In (83), $T_k^{(n)}\big(\acute{\mathbf{s}}_n(k-1), s_n(k), \acute{\mathbf{s}}_n(k)\big)$ denotes the local trellis check function of the nth subsystem (Markov chain) constraining the possible combinations of $\acute{\mathbf{s}}_n(k-1)$, $s_n(k)$, and $\acute{\mathbf{s}}_n(k)$ [137], where

$$\acute{\mathbf{s}}_n(k) = [s_n(k), \cdots, s_n(k-L-1)]^T \in \mathcal{S}^L \quad (84)$$

denotes the trellis state vector at time k. A straightforward computation of the zero-gradient points of the free energy functional $\mathcal{F}(Q)$ with respect to the free parameters $q_{n,k}(\mathbf{s}_n(k))\ \forall n, k$ yields a set of fixed-point equations given by

$$q_{n,k}(\mathbf{s}_n(k)) = \frac{1}{\gamma_{n,k}} \times \exp\left[-\frac{1}{T}\sum_{m}\Big|r_m(k) - \sum_{n'\neq n} \hat{\mathbf{s}}_{n'}^T(k)\mathbf{h}_{m,n'}(k) - \mathbf{s}_n^T(k)\mathbf{h}_{m,n}(k)\Big|^2\right], \ \forall n, k, \quad (85)$$

where $\gamma_{n,k}$ is a scaling factor, and $\hat{\mathbf{s}}_n(k) = E_{Q^{(n)}}\big[\mathbf{s}_n(k)\big]$. The "per-branch" distribution $Q^{(n)}(\mathbf{s}_n)$ can be computed by operating the standard SP algorithm on the factor

[13] The terms hidden state and trellis state are used interchangeably later on in this thesis.

graph (trellis diagram) describing the dynamics of the nth transmitter antenna signal. A factor graph representation of this structured MF approximation is presented in Fig. 5, assuming that the transmitter employs two antennas.

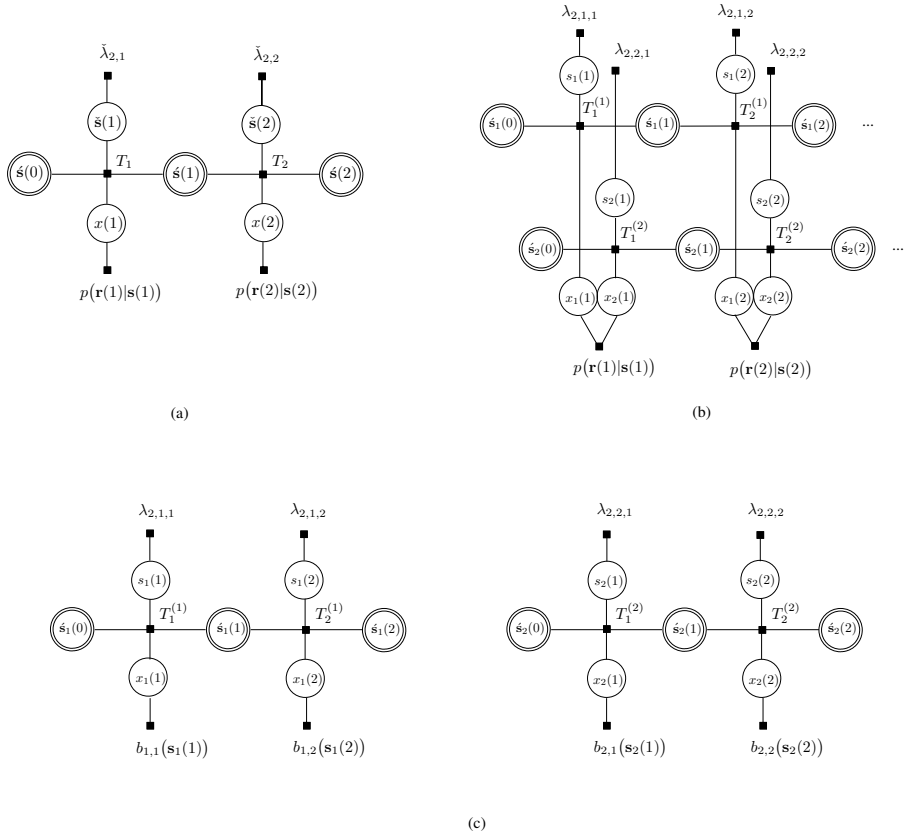

Fig 5. Factor graph representations using (a) a cycle-free graph, (b) a factorial hidden Markov model and (c) a structured mean-field approximation for an example MIMO system with two transmitter antennas. The following notations are used: $x(k) = \sum_m \mathbf{h}_m^T(k)\mathbf{s}(k)$**,** $x_n(k) = \sum_m \mathbf{h}_{m,n}^T(k)\mathbf{s}_n(k)$**,** $\acute{\mathbf{s}}(k) = \left[\acute{\mathbf{s}}_1^T(k), \acute{\mathbf{s}}_2^T(k)\right]^T$**,** $\check{\mathbf{s}}(k) = \left[s_1(k), s_2(k)\right]^T$**, and** $\check{\lambda}_{2,k} = \left[\lambda_{2,1,k}, \lambda_{2,2,k}\right]^T$**.**

Effectively, the structured MF approximation yields a bank of "per-branch" APP demodulators which are preceded by a soft interference canceller (SIC). Moreover, solv-

ing the fixed point equations (85) entails block-iterative processing where the soft output of each APP demodulator is fed back into the SIC. A block diagram of a related turbo receiver (assuming three transmitter antennas) is presented in Fig. 6, including separate SISO decoders for all transmitter antenna signals. Interestingly, a related per-branch SISO demodulator was proposed on an ad-hoc basis for frequency-selective MIMO channels in [95]. However, unlike the receiver proposed above, the MIMO receiver in [95] separates the different transmitter antenna signals prior to the demodulator bank by using a probabilistic separator, which, in effect, executes the updating rules of the SP algorithm. Hence, the MIMO receiver algorithm in [95], as an entity, is identical to the SP algorithm operating on the LFG of Fig. 5(b).

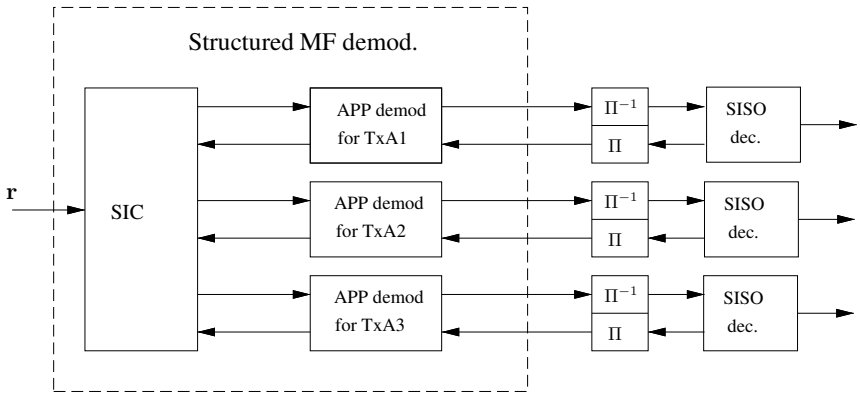

Fig 6. Block diagram of a turbo receiver based on the structured MF approximation. The blocks labeled as Π and Π^{-1} denote the pseudorandom interleaver and the corresponding deinterleaver, respectively.

Another approach to improving the naive mean field approximation is to modify the free energy $\mathcal{F}(Q)$ itself. One option, originally proposed in statistical physics, is to apply the so-called advanced mean field techniques where a second order self-coupling reaction term is added into $\mathcal{F}(Q)$. This leads to the free energy functional which was called an Onsager-Gibbs free energy in [258]. Related higher order correction terms were proposed by using a linear response theory in [131]. Another related technique known as a Bethe method [265] approximates the entropy term of the variational free

energy $\mathcal{F}(Q)$ with the following Bethe entropy term:

$$\begin{aligned}\mathcal{H}_{\text{Bethe}}(Q) = &-\sum_{k}\sum_{\mathbf{s}(k)} q_k(\mathbf{s}(k)) \ln q_k(\mathbf{s}(k)) \\ &+\sum_{n}\sum_{k}(d_{n,k}-1)\sum_{s_n(k)} q_{n,k}(s_n(k)) \ln q_{n,k}(s_n(k)),\end{aligned} \tag{86}$$

where $d_{n,k}$ denotes the degree of the symbol node of $s_n(k)$ (see [265] for details). The Bethe free energy $\mathcal{F}_{\text{Bethe}}(Q) = \mathcal{U}(Q) - \mathcal{H}_{\text{Bethe}}(Q)$ is then minimized subject to the normalization constraints and the local consistency constraints $\sum_{\mathbf{s}(k)\setminus s_n(k)} q_k(\mathbf{s}(k)) = q_{n,k}(s_n(k))$.

Interestingly enough, as shown in [265], the set of fixed point equations obtained by the constrained minimization of $\mathcal{F}_{\text{Bethe}}$ are equal to the update equations of the standard SP algorithm operating on the associated factor graph. In fact, the fixed points of the SP algorithm correspond to the local *minima* (not only the zero-gradient point) of the Bethe free energy [110]. Importantly, the local consistency requirements do not, however, automatically imply global consistency in that the beliefs $q_k(\mathbf{s}(k))$ and $q_{n,k}(s_n(k))$ do not necessarily correspond to the marginal distributions of any single global distribution $Q(\mathbf{s})$, unless the graph is cycle-free. Moreover, in contrast to the mean field approximation, none of the above cited variational approximation methods that are based on the modified free energy functional can be interpreted as an upper bound to the Helmholtz free energy.

Despite the improved entropy, experiments have shown [139, 51] that the Bethe method is not able to produce satisfactory solutions unless the girth of the factor graph associated with the detector is at least six. Since the advanced mean field approximations can generally be regarded as inferior to the Bethe method [258], their use for inference is beset by similar difficulties. Instead, an entire array of even more sophisticated approximations, including the Kikuchi and cluster variational methods [265, 197], have been proposed as potential tools for obtaining the desired balance between complexity and performance.

Moreover, in [137, 51], clustering and stretching techniques were introduced as effective means to modify the factor graph with short cycles so that those nodes of the graph that are causing problems are eliminated. A nice feature of these techniques is that the resulting modified graph still gives an exact graphical representation of the underlying system. In fact, any graph with cycles can be transformed, through a process of clustering its nodes together so as to form aggregated nodes, into a structure known

as a junction tree [141]. The standard SP algorithm can then be applied to perform exact inference on this junction tree. Of course, the performance advantages obtained via clustering are not accrued without increased computational complexity.

Another effective modification to the graphical model of the system may be obtained by eliminating only trouble-causing *edges* in the graph via the application of the MF approximation, instead of using an impoverished assumption of complete factorizability. The key point here is that a proper mixture of different variational approximations, like the Bethe or Kikuchi and (structured) MF approximations, could provide a system designer with an effective tool to obtain enhancements of standard turbo receivers in a more systematic way.

Algorithms minimizing the variational free energy functions directly by using the double-loop optimization technique are described in [257, 111, 234, 268]. In contrast to the SP and MF algorithms, they guarantee convergence but are typically an order of magnitude slower than the algorithms solving the fixed points of the free energy function. In addition, they are more complex to implement.

3.5 Complexity considerations

Various SISO detectors, assuming perfect CSI, were discussed, reinterpreted, and genaralized in the previous sections. In this section, these are further considered by paying special attention to relative complexities of these detection schemes. While important, the problem of assessing the complexity pertaining to any detection algorithm is difficult, and the result of such an assessment may be somewhat ambiguous. A detailed complexity analysis would imply that all the arithmetic operations required by the algorithm are exactly counted, and, even then, the "true" complexity will depend on the computing platform and the selected implementation architecture. Therefore, discussion here concerning the complexities is kept at a very general level, and only approximate time complexities of the algorithms are considered.[14]

Let us consider a SISO detection in a MIMO system with N transmitter antennas and M receiver antennas. As is often cited in the literature, computation of exact symbol APPs, i.e., an exact inference algorithm, is computationally intractable in such systems,[15] and particularly so if the block length K is large. This statement becomes con-

[14]The time complexity of an algorithm is defined as a number of computational steps that it takes for the algorithm to solve the problem at hand.

[15]Problems that are solvable in theory, but cannot be solved in practise, are often called *intractable* problems.

crete by considering a *naive* exact inference algorithm which consists of modeling the MIMO system as HMM with J^{NL} states and operating a (forward-backward processing) SP algorithm on the corresponding factor graph (see Fig. 5(a)). It has approximate time complexity $\mathcal{O}(KJ^{2NL})$, where $\mathcal{O}(\cdot)$ should be read as "on the order of." However, by exploiting the conditional independency structure of factorial HMMs encoded by the missing edges in the corresponding factor graph (see Fig. 5(b)), the exact inference can alternatively be implemented with a time complexity of $\mathcal{O}(KNLJ^{NL+1})$ (see [90]) by using the junction tree algorithm by Lauritzen and Spiegelhalter [142]. The junction tree algorithm is obtained as a result of moralizing and triangulating the factor graph of Fig. 5(b) so that a new factor graph with no loops is attained and then operating the standard SP algorithm on this transformed factor graph. Despite a significant reduction in complexity compared to the naive inference algorithm, the junction tree algorithm is miserably complex for many practical applications.

On the other hand, the distributed computation of factorized state probabilities, inherently done in structured MF-based SISO detectors, enables a substantial reduction in detector complexity. The SIC front-end decouples the received signal into conditionally independent signals which can, therefore, be demodulated separately, as depicted in Fig. 5(c). Hence, a time complexity of the structured MF receiver is $\mathcal{O}(KNJ^{2L})$. The same time complexity pertains to the SP algorithm operating on the factor graph of Fig. 5(b) (i.e., the SISO detector proposed in [95]). However, the probabilistic separator assumed by the scheme in [95] may be slightly more complex than the SIC front-end in the structured MF solution. It is important to note that in both of these cases the complexity reduction is obtained at the cost of decreased performance.

Another class of approximate inference algorithms is obtained by breaking the state variable of the HMM into separate symbol variables which are, thus, represented as separate nodes in the corresponding FG. Then, the approximate symbol APPs can be computed by operating a standard SP algorithm on the resulting LFG, but the approximation may not be acceptable unless the channel is sparse enough (see, e.g., [139]). Another option is to transform the LFG into one where all the conditional dependencies are removed via the MF approximation (see Fig. 3). This approach results in the iterative linear SISO detector whose time complexity is $\mathcal{O}(KN(L+1)J^2)$ per iteration.

The linear MMSE SISO detector is more complex than the LCMF SISO detector since the inverse of the square matrix of size $M(K+L)\times M(K+L)$ has to be computed per each transmitted data block (see (79) for details). The requirement of inverting a potentially large matrix is avoided, however, if the frequency-domain processing [254,

253] (facilitated by the insertion of a CP into the transmission format) is assumed, in which case the complexity of the linear MMSE detector is dictated by the size of the fast Fourier transform (FFT). It is important to note that the size of the FFT in the context of frequency-domain equalizers is, in general, a design parameter and may be selected independently from the data block size [115]. It is, however, very hard to pin down precisely when the frequency-domain implementation should be used instead of the time-domain implementation. The choice is in any case governed by tradeoffs in computation time and data storage requirements, by the block size parameters, by the issues related to the channel dynamics, and so on.

3.6 Numerical results

In this section, the performance of various turbo detection schemes in two different system architectures is evaluated by computer simulations in terms of either the bit error rate (BER) or frame error rate (FER) versus E_b/N_0 (E_b denotes the average signal energy per information bit and N_0 denotes the spectral density of the receiver thermal noise). It should be noted that E_b/N_0 values were defined by taking into account the loss due to the training symbols (i.e., known preamble and tail symbols). In most cases, a matched filter bound (MFB), which is obtained via numerical simulation, is provided as a lower bound to the performance of the investigated receivers. The MFB gives the performance of an "ideal" receiver, which assumes perfect channel state information (CSI) and which cancels the effects of ISI perfectly (see also the Remark 1 in Section 3.2.1). In fixed (time invariant) channels, the MFB coincides with the performance curve obtained in the AWGN channel (in brief, the MFB equals the AWGN bound in the case of fixed channels).

In the first system architecture considered in this section, each transmitter antenna transmits independent data which is first encoded with a 64-state convolutional code with code rate 1/2 (code generators in octal form are (133,171)),[16] then interleaved with a pseudo-random bit-interleaver,[17] and, finally, modulated into QPSK or 16-QAM symbols by using a Gray-like mapping of information bits into channel symbols. Before transmission, the data symbols are collected into frames with the following frame struc-

[16]There may be more than one transmitter antenna entailing the MIMO feature.

[17]Interleaving is done by a pseudo-random permutation scheme where the permutation pattern changes from the coding block to another. This applies, in fact, to all pseudo-random interleaving schemes considered in this thesis.

ture: each frame is composed of 6 cyclic-prefixed data blocks with block size K and two cyclic-prefixed training blocks (one training block at each end of the frame) with each one having K_{pilot} symbols. The length of the cyclic prefix is denoted as K_{prefix}. The extension of the data and training blocks with the cyclic prefix effectively separates the blocks in time and, thus, enables the exploitation of the frequency domain equalization.

In Fig. 7 and Fig. 8, the performance of a single ($N = 1$ and $M = 1$) antenna system was simulated in terms of BER and FER versus E_b/N_0, respectively. The main aim of these simulations was to study how the two different ways to form the soft input of the MMSE SISO detector affect the performance. The frame parameters in this study were: $K = 512$, $K_{\text{pilot}} = 64$, and $K_{\text{prefix}} = 40$. The channel was a fixed three-tap channel with the impulse response $\mathbf{h} = [1/\sqrt{6}\ 2/\sqrt{6}\ 1/\sqrt{6}]^T$ (referred to as Proakis-B channel later in the text), and the receiver assumed perfect channel state information (CSI). The abbreviation FDE stands for the frequency domain equalizer, and TFDE($N_{outerIter} \times N_{innerIter}$) refers to the turbo detection scheme where the frequency domain MMSE SISO detector (with the covariance matrix $\mathbf{\Sigma}_i$ set at $\mathbf{\Sigma}_i = \frac{1}{KN}\sum_n \sum_k \sigma^2_{i,n,k} \times \mathbf{I}_{NK}$) and SISO decoder are iteratively interconnected in such a way that there are two iteration loops: an outer iteration loop, performing $N_{outerIter}$ iterations, where the decoder is invoked and an inner iteration loop, performing $N_{innerIter}$ iterations per each outer iteration, where only the channel-induced information is updated (see Remarks 2 and 3 of Section 3.2.1 for the theoretical basis of this scheme). Furthermore, the abbreviation fullAPP refers to the turbo detection scheme where the *a priori* distribution $P(\mathbf{s}) = \mathcal{CN}(\mathbf{s}, \overline{\mathbf{s}}, \mathbf{\Sigma}_i)$ is defined in terms of full APP information about the symbols, whereas the abbreviation extrINFO refers to the case where $P(\mathbf{s})$ is solely defined in terms of the extrinsic information supplied by the decoder (see Remark 2 in Section 3.2.1 and Remark 4 in Section 3.2.2 for further details). Finally, TurboMAP refers to the turbo detection scheme where the exact APP demodulator, implemented with the BCJR algorithm, is iteratively interconnected with the SISO decoder.

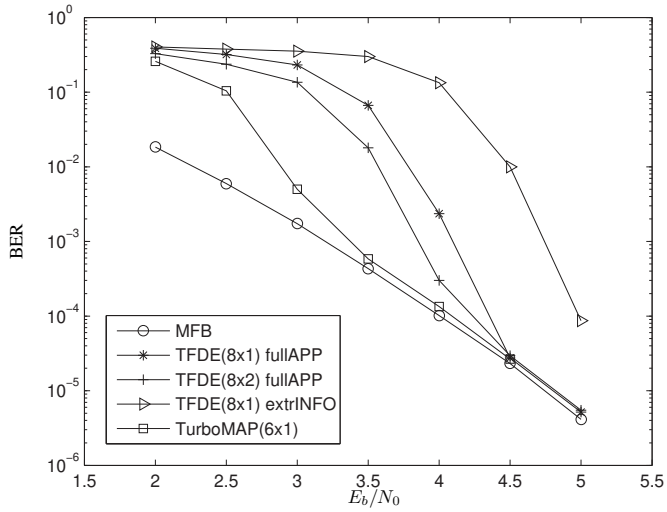

Fig 7. BER versus E_b/N_0 for a single antenna QPSK-modulated system in a fixed Proakis-B channel.

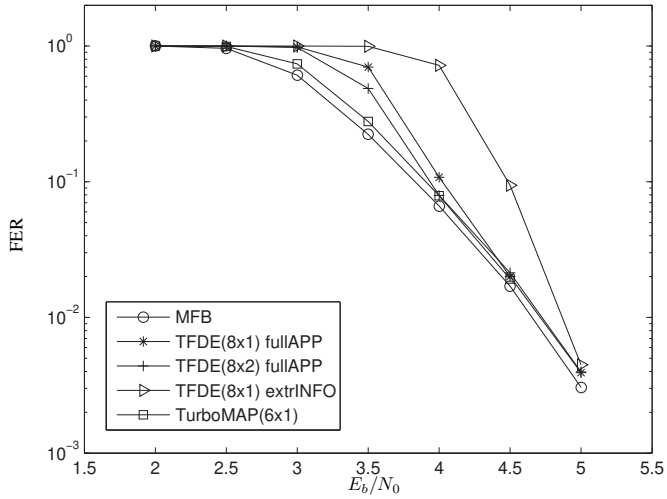

Fig 8. FER versus E_b/N_0 for a single antenna QPSK-modulated system in a fixed Proakis-B channel.

Both of these performance figures back up the theory in that the full APP information should be exploited in soft interference cancellation, even when the MMSE SISO detector is employed (in the case of the LCMF detector, similar results have been simulated earlier in [248]). It may also be noticed from the figures that the performance of the linear MMSE turbo detectors can be slightly improved if more than one inner iteration is carried out per each outer iteration.

In Fig. 9, the performance of a 3×2 ($N = 3$ and $M = 2$) antenna system in the frequency-selective Rayleigh block-fading channel was simulated. Each subchannel (between different transmitter and receiver antenna pairs) was assumed to have three independently fading taps with an exponentially decaying delay profile, implying that the averaged powers of the subchannel taps were set at

$$\left[\sigma^2_{h_{m,n,l}}\right]_{l=0}^{2} = \left[0.412 \;\; 0.327 \;\; 0.260)\right]^T.$$

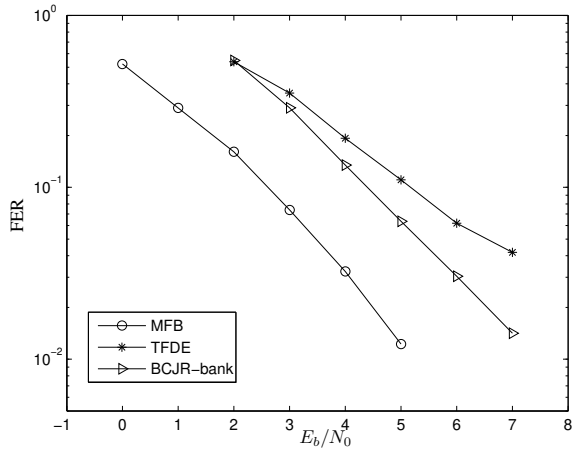

Fig 9. FER versus E_b/N_0 for a 3×2 antenna system in a frequency-selective Rayleigh fading channel (3 channel taps). The block size was $K = 256$ symbols.

The phrase "block-fading" implies that the channel was assumed to be constant during a frame while changing value randomly from frame to frame. Moreover, the receiver was assumed to have perfect CSI. Provided that each transmitter antenna transmits independent data and the channels between different transmitter and receiver antenna pairs are uncorrelated, it doesn't matter, from the receiver performance point of view, whether the transmitter antennas belong to one physical terminal or to different terminals transmitting at the same frequency band (sometimes referred to as the multiuser MIMO system). The frame parameters in this simulation were as follows: $K = 256$, $K_{\text{pilot}} = 64$, and $K_{\text{prefix}} = 20$.

A main purpose of this simulation case is to demonstrate that the link performance of the MIMO systems, where the number of transmitter antennas exceeds the number of receiver antennas, can be improved compared to the performance obtained with the linear MMSE SISO detector. This happens at the cost of increasing complexity, of course. In particular, the improvement, which may be noticed from Fig. 9, was achieved by using a hybrid turbo detection scheme (referred to as the BCJR-bank in the figure) where the first iteration was carried out by performing the MMSE SISO detection together with the SISO decoding, and, then, at the subsequent iterations, the MMSE SISO detector was replaced by the SISO demodulator that was based on the structured MF approximation. As shown in Fig. 6, the structured MF scheme is composed of the SIC followed by the bank of APP demodulators where each APP demodulator infers probabilistic values for data symbols of one of transmitted independent data streams (see (83) and (85) for details). The structured MF approximation was not considered at the first iteration due to high risk of getting frequently trapped by a wrong mode of the target distribution. When that occurs, large bursts of detection errors are typically produced, and they can eventually cause an error floor to suitable performance curves.

In the second simulation exercise, MIMO turbo receivers, which are based either on the structured MF or BP algorithms, are compared in an experimental setup that is characterized by the following arrangements. First, a 64 state convolutional encoder (code generators in octal form are (133,171)) with rate 1/2 was applied as an outer encoder to encode the information bits. The starting and terminating states of the encoder were assumed to be known in order to facilitate the decoding algorithm. After the QPSK mapper, the information symbols were interleaved by a pseudorandom interleaver. Thereafter, the interleaved symbols were provided as input to the inner space-time encoder, which in this test case was assumed to be a simple delayed repetition scheme. Specifically, the input data stream to the space-time encoder is fed directly to

the first transmitter antenna, whereas its delayed versions are fed to additional transmitter antennas (delay is T for the second antenna, $2T$ for the third antenna, and so forth). The resulting serial concatenated coding scheme contributes to the performance of the communication link by providing both *diversity* gain and *coding* gain, though no attempt was made to optimize the scheme in terms of these gain factors (see, e.g., [231] for more optimal space-time coding schemes). The parallel output symbol streams of the space-time encoder were further processed separately. This includes interleaving by a $K \times K$ block interleaver (BI) in conjunction with the assumption of a burst-to-burst independent channel and organizing the interleaved symbols into fixed size bursts with K information symbols preceded by $K_{pre} = 2$ known preamble symbols and followed by $K_{tail} = 2$ known tail symbols, rendering the starting and terminating states of the demodulator trellis to be known. A block diagram of the transmitter is presented in Fig. 10.

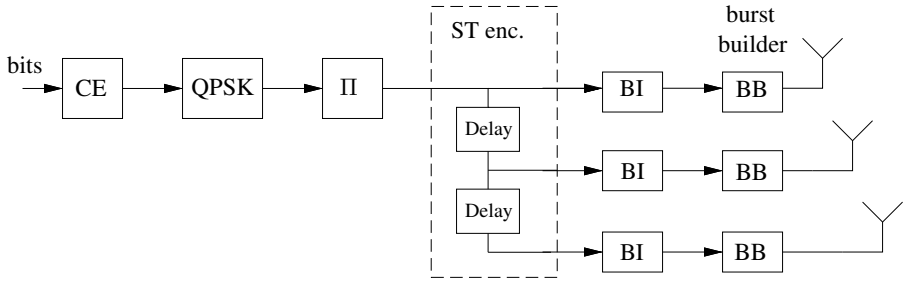

Fig 10. Block diagram of a transmitter for a space-time coded transmission system.

The channel between each pair of transmitter and receiver antennas was assumed to be frequency-selective with three independently fading taps whose standard deviations were set at ($\sigma_{\text{tap1}} = 1/\sqrt{6}$, $\sigma_{\text{tap2}} = 2/\sqrt{6}$ and $\sigma_{\text{tap3}} = 1/\sqrt{6}$). Each Rayleigh fading channel tap assumed Clarke's power spectrum with normalized Doppler spread $B_d T_s = 0.001$. In addition, the channels between different transmitter-receiver antenna pairs are assumed to be uncorrelated.

At the receiver, "inverse" operations to those at the transmitter are performed. Moreover, the receiver assumes that perfect CSI is available. A block diagram of a structured MF-based turbo receiver in the case of three transmitter antennas is presented in Fig. 11.

The structured MF demodulator is equivalent to the one described in Section 3.4, while the block labeled as "SISO ST dec." infers the posterior probability values for data symbols pertaining to the input and output of the (inner) space-time encoder. The BP-based turbo receiver is similar to the one described in Fig. 11, except that the SIC is replaced by the SP-based probabilistic separator described in [95].

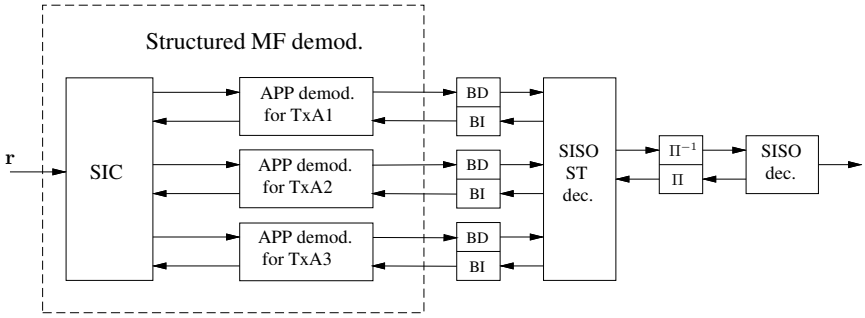

Fig 11. Block diagram of a MF inference-based turbo receiver structure in a space-time coded system with three transmitter antennas. Note that BD stands for the block deinterleaver.

The performance results in terms of the BER as a function of E_b/N_0 is presented in Fig. 12, when $K = 30$. With the 2×1 and 2×2 antenna configurations, the performance of the variational inference-based turbo receivers is compared to the performance of the full-trellis-based "optimal" turbo receivers (referred to as BCJR turbo). It is readily noticed that, under these system assumptions, the MF- and BP-based turbo receivers are able to achieve practically the same performance as the full-trellis-based turbo receivers, despite significant computational savings. Another interesting finding from this figure is that the addition of a third transmitter antenna does not give any performance gain in the considered channel. On the contrary, it gives some performance loss if only one receiver antenna is in use. This is because the loss due to increased cross-talk interference exceeds the obtained diversity gain.

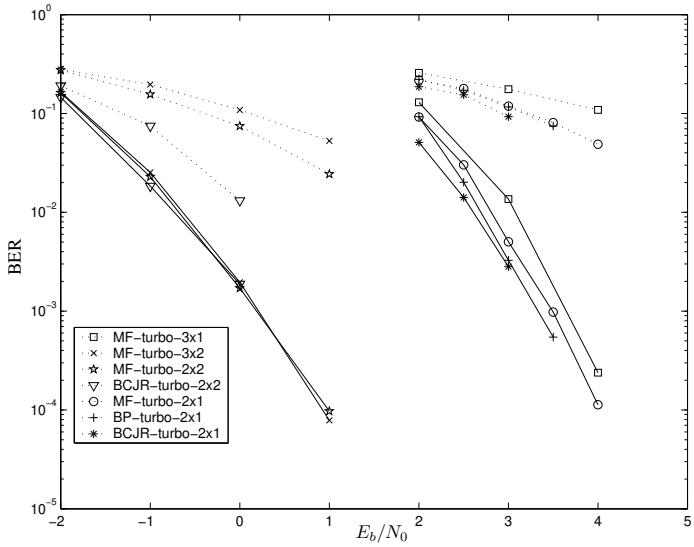

Fig 12. BER versus E_b/N_0 for space-time coded MIMO systems. The dotted lines denote the performance after first iteration, whereas the straight lines denotes the performance when up to 7 iterations have been run.

In Fig. 13, the performance results are presented for the 2×1 antenna configuration with and without the outer channel coding, using the otherwise same simulation setup as previously. In the uncoded case, the differences between various inference algorithms are naturally bigger than in the coded case. In fact, the MF-turbo receiver with a fixed temperature ($T = N_0$) seems to saturate at quite a high BER value, whereas the mean field annealing (MFA)-based turbo receiver is able to avoid such an early saturation, even though no attempt was made to optimize the annealing scheduling.

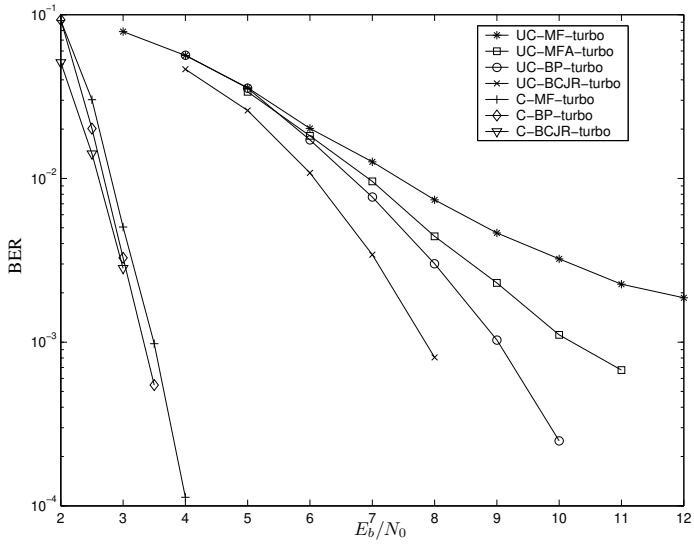

Fig 13. BER versus E_b/N_0 for a space-time coded 2×1 antenna system with and without outer channel coding (UC refers to an uncoded and C refers to a coded system).

3.7 Summary and conclusions

In this chapter, the familiar problem of turbo-processing was revisited, looking at the problem from the statistical physics perspective. First, a unifying framework for approximate inference, based on the statistical physics concept called the variational free energy minimization, was introduced. Then, as an example of its utility, it was shown that the existing linear SISO detectors can be formulated as instances of the VFEM algorithm. The obtained new formulation not only provides new insights into the structural properties and performance of those algorithms but may also give an impetus for totally new designs. A glimpse of this was seen by obtaining a feasible approximation to the APP demodulator in the presence of the frequency-selective MIMO channel via the structured MF technique. Finally, a time complexity of various SISO detection algorithms was assessed and their performance was evaluated via computer simulations.

4 Adaptive SISO detectors in frequency-selective fading channels

In this chapter, a class of adaptive SISO demodulators with embedded recursive channel estimators are derived using the VFEM framework. In particular, the classical Baum-Welch algorithm [206] and its Bayesian counterpart, the BEM algorithm [38], are formulated into recursive forms which are computationally much simpler than the equivalent block-processing algorithms and, above all, allow their integration with the well-known SISO algorithms [27, 144] originally derived for known channels. A common feature of the obtained recursive channel estimators is that they exploit the *a posteriori* probabilities (APPs) of the *trellis branches*, inherently obtainable from the forward processing part of the trellis-based SISO algorithm. In consequence, they are called *soft decision-directed* (SDD) channel estimators.

Also, a design of lower complexity linear SISO detectors, incorporating the SDD channel estimators, is addressed later in this chapter. This chapter starts, however, with a discussion on design criteria for SISO demodulators in the presence of an unknown channel. In particular, building upon the VFEM framework presented in Chapter 2, the EM and Bayesian EM-based demodulators are formulated as instances of the VFEM algorithm. For notational simplicity, only single transmitter and single receiver antenna systems are considered in this chapter. Furthermore, it is assumed that all frequency instabilities are perfectly corrected.

This chapter is organized so that in Section 4.1, a design of feasible SISO demodulators in the presence of an unknown channel is considered from the optimality point of view. Then, in Section 4.2, SDD least mean square and SDD recursive least square estimators are derived and their embedding into a forward-backward processing SISO algorithm is accomplished. In Section 4.3, an SDD Kalman filter and its reduced complexity version in the presence of the Rayleigh fading multipath channel are derived. Furthermore, in Section 4.4, joint iterative linear SISO detectors and SDD channel estimators are derived by appealing to the VFEM framework. The complexity issues are considered in Section 4.5, and the numerical performance results for uncoded and coded turbo-processing systems are presented in Section 4.6. Conclusions are finally provided in Section 4.7.

4.1 Design criteria for APP detectors in the presence of unknown channel

In the design of SISO demodulators in the presence of unknown channels, a distinction will be made between deterministic channels, which can be either time-varying or constant in time, and random channels. In the latter case, the receiver is assumed to have *a priori* knowledge of the statistics of the communication channel. As usual, the statistics of the radio channel are assumed to be fully characterized by a complex Gaussian distribution whose first and second order statistics are known *a priori*. In practise, however, the fading statistics have be estimated.

When dealing with the unknown channel and data, the issue of identifiability deserves to be considered. While there may be many definitions of identifiability, here this term is used to point out the problem of distinguishing between data sequences when the channel is unknown or only its statistical properties are known. The problem arises when there are two or more *hypothetical* data sequences which, due to the unknown channel, produce identical likelihood values at the receiver or, equivalently, when the joint likelihood of the channel and data for two or more hypothetical data sequences has the same value. Hence, there is no means to distinguish between these sequences at the receiver. Likewise, the channel can be said to be unidentifiable when there are more than one hypothesis on the channel impulse response which, due to unknown data, produce the same likelihood or joint likelihood (with the data) at the receiver.

The identifiability problem is particularly severe if the receiver has no knowledge of or no control over the signaling format, which, therefore, is simply assumed to be a random stream of modulated digital data by the receiver [41]. In general, the channel can be made identifiable and the data sequences distinguishable by imposing some code structure on the transmitted signal. Note that the insertion of pilot symbols into the signaling format is also a special type of coding method. Next, however, different design criteria for SISO demodulators are discussed without considering the issue of identifiability any further.

4.1.1 Deterministic time-invariant channels

A problem of designing a SISO demodulator in an environment where the communication channel is assumed to be unknown, deterministic, and constant during the period of data burst is addressed first. In this case, the impulse response of the channel is de-

noted simply as $\mathbf{h}$. As already pointed out earlier, the optimal detection in the sense of minimizing the BER is an ill-posed problem in the presence of deterministic channels. Instead, a reasonable choice for the soft output of the SISO demodulator is obtained via the soft-output generalized likelihood principle, in which case the symbol APP is given by

$$\mathrm{APP_d}(s(k)) = c \sum_{\mathbf{s}\setminus s(k)} \max_{\mathbf{h}} p(\mathbf{r}, \mathbf{s}|\mathbf{h}), \tag{87}$$

where c denotes a normalizing constant and the subscript "d" refers to the deterministic channel. In particular, the formula (87) implies that the CIR is estimated separately for each possible data sequence using the ML criterion, and the obtained sequence metrics are then combined so as to produce the approximate symbol APPs. Although limited by high computational complexity, this design criterion is a good starting point for practical receiver designs [8]. For example, practical SISO demodulators for deterministic time-varying channels can be obtained by applying simplified ML-based CIR estimators with channel tracking capabilities (e.g., the LMS algorithm) and sequence pruning techniques (e.g., via the PSP-technique). In principle, the order of summation and maximization in (87) could be reversed, yielding another potential definition for the soft-output of the SISO demodulator in the deterministic channel. However, it is not clear (let alone obvious) whether this criterion could be used as a basis for practical SISO demodulators.

An alternative design criterion is obtained by defining the approximate symbol APPs as follows

$$\mathrm{APP_d}(s(k)) = p(s(k)|\mathbf{r}, \hat{\mathbf{h}}_{\mathrm{ML}}), \tag{88}$$

where the (blind) ML estimate of the CIR is given as

$$\hat{\mathbf{h}}_{\mathrm{ML}} = \arg\max_{\mathbf{h}} p(\mathbf{r}|\mathbf{h}) = \arg\max_{\mathbf{h}} \sum_{\mathbf{s}} p(\mathbf{r}, \mathbf{s}|\mathbf{h}). \tag{89}$$

While the direct computation of (89) appears unattractive, the EM algorithm [62] provides a low-complexity iterative procedure for computing (88) and (89) jointly. The EM algorithm starts with some initial CIR estimate $\hat{\mathbf{h}}^{(0)}$, and then proceeds to iteratively generate successive estimates by repeatedly applying the following two steps, given at the ith iteration as

$$\mathtt{E-step}: \qquad \text{Compute } \; Q^{(i)}(\mathbf{s}) = p(\mathbf{s}|\mathbf{r}, \mathbf{h}^{(i-1)}) \tag{90}$$

$$\mathtt{M-step}: \qquad \hat{\mathbf{h}}^{(i)} = \arg\max_{\mathbf{h}} E_{Q^{(i)}}\left[\ln p(\mathbf{s}, \mathbf{r}|\mathbf{h})\right]. \tag{91}$$

Thus, in the E-step, the unknown data sequence is effectively represented by the *a posteriori* probability distribution over the values of $\mathbf{s}$, while in the M-step these probabilities are used to maximize the expected likelihood of the received signal samples over the unknown CIR. It is important to note that the *a posteriori* distribution $p(\mathbf{s}|\mathbf{r},\mathbf{h})$ has a natural decomposition due to the Markovian property of the frequency-selective channel, facilitating the computations pertaining to the M-step. Iterations between the E- and M-steps are continued until the desired degree of accuracy has been achieved. Although convergence to the exact ML solution can not be guaranteed, the EM algorithm has an appealing property of monotonically increasing the log likelihood $\mathcal{L}(\mathbf{h}) = \ln p(\mathbf{r}|\mathbf{h})$ as a function of iterations [62].

Interestingly, the EM algorithm can also be interpreted as the VFEM algorithm, as demonstrated by the example 3 of Section 2.3. This was also formally proven in [176]. Fitting the EM algorithm into the variational optimization framework is, nevertheless, summarized here as well. This is done in order to properly emphasize its importance in providing a natural framework for further approximations.

First, let us define the functional $\mathcal{G}(Q,\mathbf{h})$ as

$$\mathcal{G}(Q,\mathbf{h}) \triangleq -D\big(Q(\mathbf{s})\|p(\mathbf{s},\mathbf{r}|\mathbf{h})\big) = E_Q\big[\ln p(\mathbf{s},\mathbf{r}|\mathbf{h})\big] + \mathcal{H}(Q), \tag{92}$$

where $\mathcal{H}(Q)$ denotes the entropy of $Q(\mathbf{s})$. The functional $\mathcal{G}(Q,\mathbf{h})$ can be regarded as the inverse free energy (see (53) for the definition of the free energy) and, by applying the Bayes rule to (92), it can be written alternatively as

$$\begin{aligned}\mathcal{G}(Q,\mathbf{h}) &= \sum_{\mathbf{s}} Q(\mathbf{s}) \ln \frac{p(\mathbf{s}|\mathbf{r},\mathbf{h})p(\mathbf{r}|\mathbf{h})}{Q(\mathbf{s})} \\ &= \mathcal{L}(\mathbf{h}) - D\big(Q(\mathbf{s})\|p(\mathbf{s}|\mathbf{r},\mathbf{h})\big).\end{aligned} \tag{93}$$

Thus, the inverse free energy $\mathcal{G}(Q,\mathbf{h})$ defines a lower bound on the likelihood of $\mathbf{h}$ with equality if and only if $Q(\mathbf{s}) = p(\mathbf{s}|\mathbf{r},\mathbf{h})$. Consequently, based on (92) and (93), the EM algorithm can be formulated as an alternating maximization algorithm, where the ith iteration cycle is defined as[18]

$$\mathtt{E-step}: \qquad Q^{(i)}(\mathbf{s}) = \arg\max_{Q(\mathbf{s})} \mathcal{G}\big(Q(\mathbf{s}), \hat{\mathbf{h}}^{(i-1)}\big) \tag{94}$$

$$\mathtt{M-step}: \qquad \hat{\mathbf{h}}^{(i)} = \arg\max_{\mathbf{h}} \mathcal{G}\big(Q^{(i)}(\mathbf{s}), \mathbf{h}\big). \tag{95}$$

[18] A similar interpretation of the EM algorithm has been presented also in [54].

Now, the approximate EM algorithms, dubbed here the variational EM algorithms, are obtained from (94) and (95) by performing maximization in the E-step over a restricted class of distributions. In particular, the maximization over the fully factorized distributions results in what is often referred to as the mean-field expectation-maximization (MF-EM) algorithm. In fact, the MF-EM algorithm has earlier been applied to solving various estimation problems in a wide variety of applications [90, 270, 272, 271, 273]. These algorithms are suboptimal in the sense that the likelihood is not guaranteed to increase at each iteration any more. However, the lower bound on the likelihood is guaranteed not to decrease at any iteration, since the inequality

$$\mathcal{L}(\mathbf{h}) \geq \mathcal{G}(Q, \mathbf{h}) \tag{96}$$

always holds (follows directly from (93)). Therefore, the convergence of the variational EM algorithms can be defined rather in terms of the bound than in terms of the likelihood. In Section 4.2, the design criterion (88) and (89) is used as a starting point in deriving new adaptive SISO demodulators. In Section 4.4, adaptive linear SISO detectors, incorporating the SDD channel estimators, are obtained by making use of the variational EM algorithms.

4.1.2 Random Gaussian channels

A design criterion for the SISO demodulation in the presence of a Rayleigh fading channel is considered next. The fading statistics are assumed to be known by the receiver. The soft output of the SISO demodulator in this case can be defined in terms of exact symbol APPs as follows

$$\mathrm{APP}_{\mathrm{p}}(s(k)) = \sum_{\mathbf{s} \setminus s(k)} \int_{\mathbf{\Theta}} p(\mathbf{s}, \mathbf{\Theta}|\mathbf{r}) d\mathbf{\Theta}, \tag{97}$$

where $\mathbf{\Theta}$ denotes the set of channel snapshots given as $\mathbf{\Theta} = \{\mathbf{h}(1), \ldots, \mathbf{h}(K)\}$ and the subscript "p" refers to probabilistic modeling. In general, the complexity of computing exact symbol APPs scales exponentially in the sequence length K, since no trellis-processing applies. However, as an exception to this rule, in the case of transmission over frequency-nonselective Rayleigh fading channels with equal-energy signal constellations, the exact symbol-by-symbol soft-decision metrics implied by the min-sum algorithm can be evaluated with *polynomial* worst case complexity in the sequence

length regardless of the operating SNR [10, 175]. As in deterministic channels, practical approximate SISO detectors in the presence of a random Gaussian channel can be obtained by pruning of the sequence tree, e.g., via the PSP principle. In flat-fading channels, a family of approximate SISO detectors were obtained in [175], by simplifying some key steps in the polynomial-complexity algorithm.

An alternative design criterion for the soft output of the SISO demodulator in the presence of the unknown Rayleigh fading channel is given by

$$\mathrm{APP}_{\mathrm{p}}(s(k)) = p\big(s(k)|\mathbf{r}, \hat{\boldsymbol{\Theta}}_{\mathrm{MAP}}\big) = \sum_{\mathbf{s}\setminus s(k)} p\big(\mathbf{s}|\mathbf{r}, \hat{\boldsymbol{\Theta}}_{\mathrm{MAP}}\big), \tag{98}$$

where the (blind) sequence MAP estimate for the channel is defined as

$$\hat{\boldsymbol{\Theta}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{\Theta}} p\big(\boldsymbol{\Theta}|\mathbf{r}\big). \tag{99}$$

A low-complexity iterative algorithm for computing jointly (98) and (99) can be obtained via a Bayesian EM algorithm, where the data symbols are regarded as the hidden data, and the data symbols and received samples jointly form the complete data specification [38, 152]. The Bayesian EM algorithm can also be formulated as an instance of the VFEM algorithm as shown by the following proposition.

Proposition 3 *The Bayesian EM algorithm for computing (98) can be interpreted as an instance of the VFEM algorithm, where* $\mathcal{F}(Q) \propto D\big(Q(\mathbf{s},\boldsymbol{\Theta})\|p(\mathbf{r},\mathbf{s},\boldsymbol{\Theta})\big)$ *($\propto$ denotes proportionality) and the trial distribution* $Q(\mathbf{s},\boldsymbol{\Theta})$ *is assumed to take a factorized form* $Q(\mathbf{s},\boldsymbol{\Theta}) = Q(\mathbf{s})Q(\boldsymbol{\Theta})$. *In addition,* $Q(\boldsymbol{\Theta})$ *is assumed to be of form* $Q(\boldsymbol{\Theta}) = \prod_k \delta\big(\mathbf{h}(k) - \hat{\mathbf{h}}(k)\big)$, *where* $\delta\big(\mathbf{h}(k) - \hat{\mathbf{h}}(k)\big)$ *denotes a vector Dirac delta function with the following properties:* $\int_{\mathbf{h}(k)} \delta\big(\mathbf{h}(k) - \hat{\mathbf{h}}(k)\big) f\big(\mathbf{h}(k)\big) d\mathbf{h}(k) = f\big(\hat{\mathbf{h}}(k)\big)$ *and* $\delta(\mathbf{0}) = 1$.

Proof. The variational free energy $\mathcal{F}(Q)$ can be expanded as follows:

$$\begin{aligned}
\mathcal{F}(Q) \propto & -\sum_{\mathbf{s}} \int_{\boldsymbol{\Theta}} Q(\mathbf{s}) \prod_k \delta\big(\mathbf{h}(k) - \hat{\mathbf{h}}(k)\big) \log p(\mathbf{r},\mathbf{s}|\boldsymbol{\Theta}) d\boldsymbol{\Theta} \\
& - \int_{\boldsymbol{\Theta}} \prod_k \delta\big(\mathbf{h}(k) - \hat{\mathbf{h}}(k)\big) \log p(\boldsymbol{\Theta}) d\boldsymbol{\Theta} \\
& + \sum_{\mathbf{s}} Q(\mathbf{s}) \log Q(\mathbf{s}) + \sum_k \int_{\mathbf{h}(k)} \delta\big(\mathbf{h}(k) - \hat{\mathbf{h}}(k)\big) \log \delta\big(\mathbf{h}(k) - \hat{\mathbf{h}}(k)\big) d\mathbf{h}(k) \\
= & -\sum_{\mathbf{s}} Q(\mathbf{s}) \sum_k \log p\big(r(k), \mathbf{s}(k)|\hat{\mathbf{h}}(k)\big) - \sum_k \log p(\hat{\boldsymbol{\Theta}}) + \sum_{\mathbf{s}} Q(\mathbf{s}) \log Q(\mathbf{s}).
\end{aligned} \tag{100}$$

Minimizing $\mathcal{F}(Q)$ alternatively with respect to $Q(\mathbf{s})$ and $Q(\mathbf{\Theta})$ while keeping the other fixed yields the Bayesian EM algorithm given at the ith iteration as

$$\texttt{E}-\texttt{step}: \qquad \text{Compute } \; p\big(\mathbf{s}|\mathbf{r}, \hat{\mathbf{\Theta}}^{(i-1)}\big) \tag{101}$$

$$\texttt{M}-\texttt{step}: \qquad \hat{\mathbf{\Theta}}^{(i)} = \arg\max_{\mathbf{\Theta}} E\Big[\log p(\mathbf{r}, \mathbf{s}, \mathbf{\Theta})\Big|\mathbf{r}, \hat{\mathbf{\Theta}}^{(i-1)}\Big], \tag{102}$$

where $E[a|b]$ denotes a conditional expectation of a given b. □

The VFEM framework provides a nice link between the exact symbol APPs and the approximate symbol APPs obtained via the Bayesian EM algorithm. On the other hand, given that $p(\mathbf{s}, \mathbf{\Theta}|\mathbf{r}) = p(\mathbf{s}|\mathbf{r}, \mathbf{\Theta})p(\mathbf{\Theta}|\mathbf{r})$, the BEM algorithm can be interpreted also so that the full description of our knowledge about $\mathbf{\Theta}$, $p(\mathbf{\Theta}|\mathbf{r})$, is being approximated by a delta-function, a pdf that is concentrated on $\hat{\mathbf{\Theta}}$. From this perspective, any other approximating distribution $Q(\mathbf{\Theta})$, no matter how inaccurate it may be, can be expected to be an improvement on the train of spikes produced by the standard Bayesian EM algorithm. For example, a simple Gaussian approximation could be a better choice. While detailed analysis and comparison of different choices for the family of trial distributions would certainly be worth examining, that issue is only slightly touched later in Chapter 6 of this thesis, in the context of iterative detection in the presence of phase uncertainty. In addition, some new results along these lines have recently been presented in [40]. In Section 4.3, it will be shown that the Bayesian EM algorithm (101) and (102), in the case where the fading statistics are described by an AR description, can be implemented by iteratively interconnecting the SDD Kalman smoother and the BCJR algorithm.

4.2 EM-based SISO demodulators for deterministic channels

Two versions of recursive SDD channel estimators for deterministic channels are next derived by using the well-known block-processing Baum-Welch algorithm as a basis. The Baum-Welch algorithm (BWA) is a special case of the EM algorithm, and it has been extensively used in the processing of HMMs [206]. The data symbols are treated by the BWA as nuisance parameters while it aims to find iteratively the ML estimate of the CIR. Specifically, assuming first that the channel is constant over the processing block (i.e., $\mathbf{h}(k) = \mathbf{h}(k-1) = \mathbf{h}$), the estimated CIR at the ith iteration of the BWA

can be formulated as a solution to the following optimization problem:

$$\begin{aligned}\hat{\mathbf{h}}^{(i)} &= \arg\max_{\mathbf{h}} E_{p(\mathbf{s}|\mathbf{r},\hat{\mathbf{h}}^{(i-1)})}\big[\ln p(\mathbf{s},\mathbf{r}|\mathbf{h})\big] \\ &= \arg\min_{\mathbf{h}} E_{p(\mathbf{s}|\mathbf{r},\hat{\mathbf{h}}^{(i-1)})}\left[\sum_k \big|r(k)-\mathbf{s}^T(k)\mathbf{h}\big|^2\right].\end{aligned} \tag{103}$$

Solving the minimization in (103), the updated estimate for the CIR can be expressed explicitly as

$$\hat{\mathbf{h}}^{(i)} = \left(\sum_k E\big[\mathbf{s}^*(k)\mathbf{s}^T(k)\big|\mathbf{r},\hat{\mathbf{h}}^{(i-1)}\big]\right)^{-1}\sum_k E\big[\mathbf{s}^*(k)\big|\mathbf{r},\hat{\mathbf{h}}^{(i-1)}\big]r(k), \tag{104}$$

where $E\big[\mathbf{s}^*(k)\mathbf{s}^T(k)|\mathbf{r},\hat{\mathbf{h}}^{(i-1)}\big]$ and $E\big[\mathbf{s}(k)|\mathbf{r},\hat{\mathbf{h}}^{(i-1)}\big]$ are obtained by means of the computationally efficient forward-backward (BCJR) algorithm [129]. In fact, in order to compute these expectations, the receiver should know the value of receiver noise variance σ_v^2 which, therefore, should also be estimated. In this section, however, it is assumed to be known.

The BWA as such is applicable only to unknown time-invariant channels. In addition, a computationally demanding matrix inversion has to be calculated at each iteration. These shortcomings providing an impetus, two different approaches will next be pursued in order to find more widely applicable, practical receiver structures.

First, the unknown CIR can be estimated iteratively by using the steepest descent method. In particular, the estimated CIR at the ith iteration can be expressed as

$$\begin{aligned}\hat{\mathbf{h}}^{(i)} &= \hat{\mathbf{h}}^{(i-1)} + \mu\left.\frac{\partial \ln p(\mathbf{r}|\mathbf{h})}{\partial \mathbf{h}}\right|_{\mathbf{h}=\hat{\mathbf{h}}^{(i-1)}} \\ &= \hat{\mathbf{h}}^{(i-1)} + \mu\sum_k \left.\frac{\partial E\Big[\ln p\big(r(k),\mathbf{s}(k)|\mathbf{h}\big)\big|\mathbf{r},\hat{\mathbf{h}}^{(i-1)}\Big]}{\partial \mathbf{h}}\right|_{\mathbf{h}=\hat{\mathbf{h}}^{(i-1)}},\end{aligned} \tag{105}$$

where the derivatives with respect to the complex valued channel impulse response are obtained by using Wirtinger calculus.[19] It should also be noted that the latter equality in (105) follows from an application of the Fisher's identity (see [62] and [161]), stating

[19] Since the real valued function $\log p(\mathbf{r}|\mathbf{h})$ is not *analytic* in complex valued $\mathbf{h}$, the ordinary complex derivative can not be used in this case. Instead, the notation of derivative must be tied in with the concept of a differential. So, as commonly done in digital commmunications and digital signal processing literature, Wirtinger calculus (see, e.g., Appendix A of [75] for details) is used here for computing the derivatives of this kind. This is also done later in all equivalent cases if not noted otherwise.

that

$$\left.\frac{\partial \ln p(\mathbf{r}|\mathbf{h})}{\partial \mathbf{h}}\right|_{\mathbf{h}=\hat{\mathbf{h}}^{(i-1)}} = \left.\frac{\partial E\Big[\ln p\big(\mathbf{r},\mathbf{s}|\mathbf{h}\big)\big|\mathbf{r},\hat{\mathbf{h}}^{(i-1)}\Big]}{\partial \mathbf{h}}\right|_{\mathbf{h}=\hat{\mathbf{h}}^{(i-1)}}. \quad (106)$$

Assuming that the step size parameter μ is small enough to ensure stability and that the initialization is done properly, the steepest descent algorithm will eventually converge to the exact ML solution [106]. Assuming additionally that μ is small enough to ensure an over-damped response, the likelihood function $p(\mathbf{r}|\mathbf{h})$ is monotonically increased at each iteration and, hence, the iterative algorithm specified by (105) can be viewed as a *generalized* EM algorithm [161] under these assumptions.

While the *generalized* Baum-Welch algorithm avoids the computation of the matrix inversion, it is still applicable only to time-invariant channels. An adaptive version of the BWA algorithm can be obtained by using the standard tool called *stochastic gradient approximation* [106]. Specifically, modeling the channel as a *time-variant* unknown deterministic process, the CIR estimate at time instance k and iteration i can be written as

$$\hat{\mathbf{h}}^{(i)}(k) = \hat{\mathbf{h}}^{(i)}(k-1) + \mu \times E_{p(\mathbf{s}|\mathbf{r},\hat{\mathbf{H}}^{(i-1)})}\Bigg[\Big(r(k) - \mathbf{s}^T(k)\hat{\mathbf{h}}^{(i)}(k)\Big)\mathbf{s}^*(k)\Bigg], \quad (107)$$

where $\hat{\mathbf{H}}^{(i-1)}$ denotes the estimate of the channel convolutional matrix obtained at the $(i-1)$th iteration. By exploiting the fact that the symbol vector $\mathbf{s}(k)$ takes discrete values in a finite space, the CIR estimate can be further written as

$$\hat{\mathbf{h}}^{(i)}(k) = \hat{\mathbf{h}}^{(i)}(k-1) + \mu \times \sum_{j=1}^{J^{L+1}} \Big(r(k) - \boldsymbol{\xi}_j^T \hat{\mathbf{h}}^{(i)}(k-1)\Big)\boldsymbol{\xi}_j^* \varphi^{(i-1)}(k,j), \quad (108)$$

where $\boldsymbol{\xi}_j$ denotes the jth element in the J^{L+1}-dimensional vector space spanned by the symbol space $\mathcal{S}$ and the conditional (branch) APP $\varphi^{(i-1)}(k,j)$ is defined as

$$\varphi^{(i-1)}(k,j) \triangleq P\big(\mathbf{s}(k) = \boldsymbol{\xi}_j|\mathbf{r},\hat{\mathbf{H}}^{(i-1)}\big). \quad (109)$$

The matrix of the conditional APPs

$$\big\{\varphi^{(i)}(k,j)\big\}_{j=1,\cdots,J^{L+1}}^{k=1,\cdots,K}$$

can be computed efficiently by the forward-backward processing SISO algorithms. After the last iteration, the branch APPs can be combined to produce the symbol APPs which can then be further processed by the iterative turbo detector or can be quantized

to produce MAP symbol decisions. The CIR estimator of (108) will be hereafter referred to as a soft-decision directed LMS (SDD-LMS) estimator.

In [13], an essentially similar SDD-LMS algorithm was proposed using a slightly different reasoning. However, its relation to the *generalized* EM algorithm was not given. Moreover, the embedding of the SDD-LMS estimator to the known SISO algorithms was also missing. Therefore, the rigorous derivation of the SDD-LMS estimator via the recursive EM algorithm gives new insight into the optimality of this estimator.

Secondly, guided again by the EM formalism, a novel SDD-RLS estimator based on a double sequential updating technique is derived. Towards this end, the objective function of the standard EM algorithm is modified in order to afford the tracking of the time-varying CIR. Specifically, the new objective function $Q\big(\mathbf{h}(k)|\hat{\mathbf{H}}^{(i-1)}\big)$, to be minimized at the kth recursion of the ith iteration with respect to $\mathbf{h}(k)$, is defined as a sum of weighted *averaged* error squares as follows:

$$Q\Big(\mathbf{h}(k)\big|\hat{\mathbf{H}}^{(i-1)}\Big) \triangleq \sum_{l=1}^{k} \lambda^{k-l} E_{p(\mathbf{s}|\mathbf{r},\hat{\mathbf{H}}^{(i-1)})}\Big[\big|r(l)-\mathbf{s}^T(l)\mathbf{h}(k)\big|^2\Big], \tag{110}$$

where λ is a forgetting factor. Solving the zero-gradient point of $Q\Big(\mathbf{h}(k)\big|\hat{\mathbf{H}}^{(i-1)}\Big)$ with respect to $\mathbf{h}(k)$ yields the following expression:[20]

$$\Big(\sum_{l=1}^{k}\lambda^{k-l}\sum_{j=1}^{J^{L+1}}\boldsymbol{\xi}_j^*\boldsymbol{\xi}_j^T\varphi(l,j)\Big)\hat{\mathbf{h}}(k) = \sum_{l=1}^{k}\lambda^{k-l}\sum_{j=1}^{J^{L+1}} r(l)\boldsymbol{\xi}_j^*\varphi(l,j), \tag{111}$$

where $\varphi(l,j)$ is defined in (109).

Now, let a running auto-correlation matrix be defined as

$$\boldsymbol{\Xi}(k,j) \triangleq \sum_{l=1}^{k}\lambda^{k-l}\sum_{i=1}^{j}\boldsymbol{\xi}_i^*\boldsymbol{\xi}_i^T\varphi(l,i) \tag{112}$$

and a running cross-correlation vector as

$$\mathbf{z}(k,j) \triangleq \sum_{l=1}^{k}\lambda^{k-l}\sum_{i=1}^{j} r(l)\boldsymbol{\xi}_i^*\varphi(l,i). \tag{113}$$

The matrix $\boldsymbol{\Xi}(k,j)$ can equivalently be written as

$$\boldsymbol{\Xi}(k,j) = \boldsymbol{\Xi}(k,j-1) + \boldsymbol{\xi}_j^*\boldsymbol{\xi}_j^T\varphi(k,j), \tag{114}$$

[20]The iteration indices are suppressed for notational convenience.

and $\boldsymbol{\Xi}(k,0) = \lambda\boldsymbol{\Xi}(k-1, J^{L+1})$. By applying the *matrix inversion lemma* [106], the running inverse auto-correlation matrix is obtained as

$$\mathbf{P}(k,j) = \boldsymbol{\Xi}^{-1}(k,j) = \mathbf{P}(k,j-1) - \mathbf{k}(k,j)\boldsymbol{\xi}_j^T\mathbf{P}(k,j-1), \tag{115}$$

where

$$\mathbf{k}(k,j) = \frac{\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^*\varphi(k,j)}{1+\boldsymbol{\xi}_j^T\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^*\varphi(k,j)} \tag{116}$$

and $\mathbf{P}(k,0) = \lambda^{-1}\mathbf{P}(k-1, J^{L+1})$. After some manipulation, the *branch* update for the CIR estimate is obtained as

$$\begin{aligned}\hat{\mathbf{h}}(k,j) &= \boldsymbol{\Xi}^{-1}(k,j)\mathbf{z}(k,j)\\ &= \hat{\mathbf{h}}(k,j-1) + \mathbf{k}(k,j)\big(r(k) - \boldsymbol{\xi}_j^T\hat{\mathbf{h}}(k,j-1)\big).\end{aligned} \tag{117}$$

And finally, the *time* update for the CIR estimate is obtained simply as $\hat{\mathbf{h}}(k+1,0) = \hat{\mathbf{h}}(k, J^{L+1}) \equiv \hat{\mathbf{h}}(k)$. Putting these together, the soft decision directed RLS (SDD-RLS) estimator can now be summarized by the following pseudocode: For each instant of time, $k = 1, \ldots, K$, compute

$$\begin{aligned}&\mathbf{P}(k,0) = \lambda^{-1}\mathbf{P}(k-1, J^{L+1})\\ &\hat{\mathbf{h}}(k,0) = \hat{\mathbf{h}}(k-1)\\ &\texttt{for}\quad j = 1, \ldots, J^{L+1}\\ &\qquad \mathbf{k} = \frac{\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^*\varphi(k,j)}{1+\boldsymbol{\xi}_j^T\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^*\varphi(k,j)}\\ &\qquad \hat{\mathbf{h}}(k,j) = \hat{\mathbf{h}}(k,j-1) + \mathbf{k}\left(r(k) - \boldsymbol{\xi}_j^T\hat{\mathbf{h}}(k,j-1)\right)\\ &\qquad \mathbf{P}(k,j) = \mathbf{P}(k,j-1) - \mathbf{k}\,\boldsymbol{\xi}_j^T\,\mathbf{P}(k,j-1)\\ &\texttt{end}\\ &\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k, J^{L+1}).\end{aligned}$$

It may be readily noticed that the recursions in the SDD-RLS estimator are running over the *time* and the *branch* indices. Hence, each new update of the CIR estimate in *time* requires an exponentially increasing number (with respect to the channel memory) of recursions over the branch indices. Despite this drawback, the SDD-RLS estimator as well as the SDD-LMS estimator can be regarded as a viable solution for high-performance turbo-like channel tracking in all but extremely frequency dispersive channels (implying a very large value of L). Importantly, the soft data statistics supplied by the channel decoder are efficiently exploited in the channel estimation. The

newly estimated channel convolution matrix $\hat{\mathbf{H}}$ can subsequently be used to update the matrix of conditional (branch) APPs which, on the other hand, also embody the soft information about the channel symbols supplied by the SISO decoder. Like all iterative algorithms, the adaptive BWAs have to be initialized properly. In practice, the initialization is obtained with the aid of pilot symbols and, therefore, the SDD-LMS/RLS estimators should not be regarded as blind estimators as is sometimes done in the literature (see, e.g., [13]).

Aside from the iterative formulation, the SDD-LMS and SDD-RLS estimators can also be easily embedded into the fixed interval (FI) or fixed lag (FL) SISO algorithms which have been developed for known channels. In effect, only the branch APPs $\varphi(k,j)$ as defined in (109) have to be replaced by the branch APPs $\varphi'(k,j)$ defined as

$$\varphi'(k,j) \stackrel{\triangle}{=} P\big(\mathbf{s}(k) = \boldsymbol{\xi}_j | \mathbf{r}_1^k, \hat{\mathbf{h}}_1^{k-1}\big), \tag{118}$$

where $\mathbf{r}_1^k \stackrel{\triangle}{=} [r(1), \ldots, r(k)]^T$ and $\hat{\mathbf{h}}_1^{k-1} \stackrel{\triangle}{=} \big\{\hat{\mathbf{h}}(1), \ldots, \hat{\mathbf{h}}(k-1)\big\}$ with the latter equation denoting the sequence of estimated CIR snapshots. The forward processing of the adaptive BCJR algorithm (or the A-SISO algorithm) with an embedded SDD-LMS/RLS estimator (referred to here as APP-SDD-LMS/RLS algorithm) can be carried out recursively for $k = 1, \ldots, K$ as follows :

$$\gamma(k,l,m) = P\big(\acute{\mathbf{s}}(k) = S_l \big| r(k), \acute{\mathbf{s}}(k-1) = S_m, \hat{\mathbf{h}}(k-1)\big) \tag{119}$$

$$\varphi'(k,j) = \alpha(k-1,m)\gamma(k,l,m), \quad \{l,m\} \in \mathcal{C}_k(j) \tag{120}$$

$$\alpha(k,l) = \sum_{m \in \mathcal{F}_k(l)} \alpha(k-1,m)\gamma(k,l,m), \tag{121}$$

where the CIR estimate $\hat{\mathbf{h}}(k-1)$ is updated in the SDD manner using $\varphi'(k-1,j)$ and, as shown in Fig. 14, the initial channel estimate $\hat{\mathbf{h}}_{\text{init}} = \hat{\mathbf{h}}(0)$ is obtained by using a pilot symbol aided estimator (PSAE). In addtion, $\acute{\mathbf{s}}(k)$ denotes the trellis state vector at time k (see also (84)) and S_l denotes the value of the lth state of the trellis.

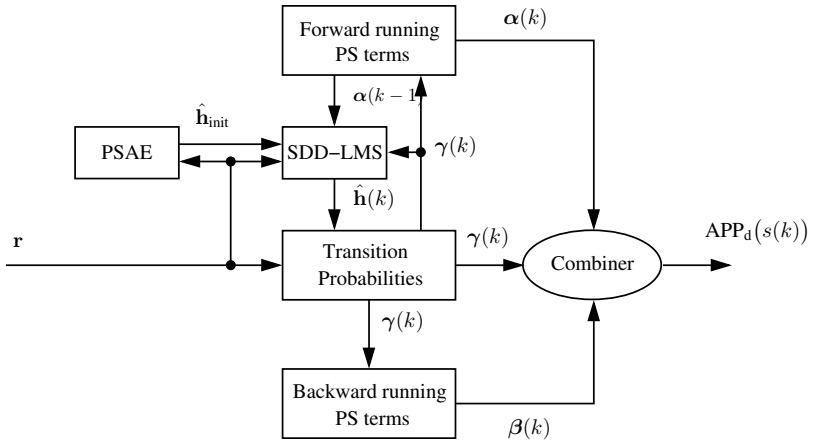

Fig 14. Block diagram of the adaptive trellis-based SISO demodulator with an embedded SDD-LMS channel estimator (cf. Eq. (119)-(123)).

The backward running product sum variables $\beta(k,l)$ for $k = K-1,\ldots,1$ are computed using the stored values of $\gamma(k,l,m)$ as

$$\beta(k,l) = \sum_{m\in\mathcal{B}_k(l)} \beta(k+1,m)\gamma(k,m,l) \tag{122}$$

and, as depicted in Fig. 14, the symbol APPs for $k = 1,\cdots,K$ are finally computed as

$$P\big(s(k)=\alpha_j\big|\mathbf{r},\hat{\mathbf{h}}_1^K\big) = \sum_{\{l,m\}\in\mathcal{D}_k(j)} \alpha(k-1,m)\gamma(k,m,l)\beta(k,l). \tag{123}$$

The following definitions are used for the sets:

$$\begin{aligned}
\mathcal{F}_k(l) &= \{m\in\mathcal{I}_s : (\acute{\mathbf{s}}(k)=S_l, \acute{\mathbf{s}}(k-1)=S_m) \text{ is allowable forward transition}\}\\
\mathcal{B}_k(l) &= \{m\in\mathcal{I}_s : (\acute{\mathbf{s}}(k)=S_m, \acute{\mathbf{s}}(k-1)=S_l) \text{ is allowable backward transition}\}\\
\mathcal{C}_k(j) &= \{l,m\in\mathcal{I}_s : (\acute{\mathbf{s}}(k)=S_l, \acute{\mathbf{s}}(k-1)=S_m) \text{ is consistent with } \mathbf{s}(k)=\boldsymbol{\xi}_j\}\\
\mathcal{D}_k(j) &= \{l,m\in\mathcal{I}_s : (\acute{\mathbf{s}}(k)=S_l, \acute{\mathbf{s}}(k-1)=S_m)) \text{ is consistent with } s(k)=\alpha_j\},
\end{aligned}$$

where $\mathcal{I}_s = \{1,2,\ldots,J^L\}$ denotes the index set for the trellis states. The computations of the APP-SDD-RLS algorithm are summarized in Table 3. Likewise, the FL SISO algorithms, e.g., the SISO algorithm by Li, Vucetic and Sato [144], can be modified to allow built-in adaptation to unknown time-varying channels.

Table 3. Update equations for the APP-SDD-RLS algorithm.

APP-SDD-LMS

Compute $\hat{\mathbf{h}}(0)$ using pilot symbols.

Initialization:

$\alpha(0,1)=1,\ \alpha(0,j)=0$ for $j\neq 1$

$\beta(K+1,1)=1,\ \beta(K+1,j)=0$ for $j\neq 1$

`For` $k=1,\ldots,K$

$\gamma(k,l,m)=P\big(\acute{\mathbf{s}}(k)=S_l\big|r(k),\acute{\mathbf{s}}(k-1)=S_m,\hat{\mathbf{h}}(k-1)\big)$ $\forall l,m$

$\varphi'(k,j)=\alpha(k-1,m)\gamma(k,l,m),\ \{l,m\}\in\mathcal{C}_k(j),\ \forall j$

$\alpha(k,l)=\sum_{m\in\mathcal{F}_k(l)}\alpha(k-1,m)\gamma(k,l,m),\ \forall l$

Forward processing:

$\mathbf{P}(k,0)=\lambda^{-1}\mathbf{P}(k-1,J^{L+1})$

$\hat{\mathbf{h}}(k,0)=\hat{\mathbf{h}}(k-1)$

`For` $j=1,\ldots,J^{L+1}$

$\mathbf{k}=\frac{\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^*\varphi'(k,j)}{1+\boldsymbol{\xi}_j^T\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^*\varphi'(k,j)}$

$\hat{\mathbf{h}}(k,j)=\hat{\mathbf{h}}(k,j-1)+\mathbf{k}\Big(r(k)-\boldsymbol{\xi}_j^T\hat{\mathbf{h}}(k,j-1)\Big)$

$\mathbf{P}(k,j)=\mathbf{P}(k,j-1)-\mathbf{k}\boldsymbol{\xi}_j^T\mathbf{P}(k,j-1)$

`end`

$\hat{\mathbf{h}}(k)=\hat{\mathbf{h}}(k,J^{L+1})$

`end`

Backward processing:

`For` $k=K,\ldots,1$

$\beta(k,l)=\sum_{m\in\mathcal{B}_k(l)}\beta(k+1,m)\gamma(k,m,l),\ \forall l$

`end`

Combining:

$P\big(s(k)=\alpha_j\big|\mathbf{r},\widehat{\mathbf{H}}\big)=$ $\sum_{\{l,m\}\in\mathcal{D}_k(j)}\alpha(k-1,m)\gamma(k,m,l)\beta(k,l),\ \forall j$

4.3 BEM-based SISO demodulators for random Gaussian channels

The SDD channel estimator and its reduced complexity version are next derived for the frequency-selective Rayleigh fading channel when the dynamics of the fading channel are described by the state space model given in (42) and (43). While noting that in the absence of the frequency instabilities $\boldsymbol{f}(k) = \boldsymbol{h}(k)$, the linear Gaussian state space model is repeated here for easy reference as

$$\begin{aligned} \boldsymbol{h}(k) &= \boldsymbol{A}\boldsymbol{h}(k-1) + \boldsymbol{G}\mathbf{w}(k-1) \\ r(k) &= \boldsymbol{s}^T(k)\boldsymbol{h}(k) + v(k). \end{aligned}$$

The Bayesian EM algorithm, in its exact and approximate forms, is used as a basis for the estimator formulations below.

4.3.1 Demodulators via exact BEM algorithm

The BEM algorithm aims to find the MAP estimate of the Rayleigh fading channel in the presence of unknown data symbols. In fact, under the above assumptions on the dynamics of the channel, it can be realized exactly by iteratively cross-coupling the BCJR algorithm with the fixed interval Kalman smoother (KS) operating on an "averaged" state space model where the averaging is performed over the states of the demodulator trellis [152]. Computing the model parameters of this "averaged" state space model is, however, very tedious, including forward-backward processing of the whole set of symbol APPs and received signal samples. In addition, the computation of matrix inversion is required at every backward processing step. And yet, the embedding of the SDD channel estimator, obtained by operating the KS on the "averaged" state-space model, into the existing SISO algorithms does not seem to be feasible since the computation of the state-space model itself requires the existence of the symbol APPs.

Motivated by these shortcomings of the "averaged" state space model, a novel formulation of the BEM algorithm that avoids these problems is presented next. In particular, as proved by the following theorem, the BEM algorithm can alternatively be implemented by iteratively cross-coupling the BCJR algorithm with a vector Kalman smoother operating on a suitably defined vector state space model.

Theorem 1 *Consider the following linear Gaussian state-space model (at the ith iteration of the BEM algorithm):*

$$\boldsymbol{h}(k) = \boldsymbol{A}\boldsymbol{h}(k-1) + \boldsymbol{G}\boldsymbol{w}(k-1) \tag{124}$$

$$\tilde{\mathbf{r}}^{(i)}(k) = \tilde{\mathbf{D}}^{(i)}(k)\boldsymbol{h}(k) + \tilde{\mathbf{v}}(k), \tag{125}$$

where

$$\tilde{\mathbf{r}}^{(i)}(k) = \left[r(k)\sqrt{\varphi^{(i)}(k,1)}, \ldots, r(k)\sqrt{\varphi^{(i)}(k,J^{L+1})} \right]^T \in \mathbb{C}^{J^{L+1}} \tag{126}$$

and

$$\tilde{\mathbf{D}}^{(i)}(k) = \begin{bmatrix} \boldsymbol{\xi}_1^T\sqrt{\varphi^{(i)}(k,1)} & \mathbf{0}_{1\times(p-1)(L+1)} \\ \boldsymbol{\xi}_2^T\sqrt{\varphi^{(i)}(k,2)} & \mathbf{0}_{1\times(p-1)(L+1)} \\ \vdots & \vdots \\ \boldsymbol{\xi}_{J^{L+1}}^T\sqrt{\varphi^{(i)}(k,J^{L+1})} & \mathbf{0}_{1\times(p-1)(L+1)} \end{bmatrix} \in \mathbb{C}^{J^{L+1}\times p(L+1)}, \tag{127}$$

and where the pdf of the noise vector $\tilde{\mathbf{v}}(k)$ is given by

$$\tilde{\mathbf{v}}(k) \sim \mathcal{CN}\big(\tilde{\mathbf{v}}(k), \mathbf{0}_{J^{L+1}\times 1}, \sigma_v^2\mathbf{I}_{J^{L+1}}\big) \tag{128}$$

and is assumed to be independent of the noise vector $\boldsymbol{w}(k)$.[21] *Then, assigning the complete data set as $\mathcal{X} = \{\mathbf{r}, \mathbf{s}\}$ and the incomplete data set as $\mathcal{Y} = \{\mathbf{r}\}$, the maximum a posteriori (MAP) sequence estimator of $\boldsymbol{\Theta} = \big\{\boldsymbol{h}(1), \ldots, \boldsymbol{h}(K)\big\}$ at the ith iteration of the BEM algorithm, defined as*

$$\hat{\boldsymbol{\Theta}}^{(i)} = \Big\{\hat{\boldsymbol{h}}^{(i)}(1), \ldots, \hat{\boldsymbol{h}}^{(i)}(K)\Big\}, \tag{129}$$

where $\hat{\boldsymbol{h}}^{(i)}(k) = E\Big[\boldsymbol{h}(k)\big|\big\{\tilde{\mathbf{r}}^{(i)}\big\}_1^K, \big\{\tilde{\mathbf{D}}^{(i)}\big\}_1^K\Big]$ with the following definitions used for the vector and matrix sequences:

$$\big\{\tilde{\mathbf{r}}^{(i)}\big\}_1^K = \big\{\tilde{\mathbf{r}}^{(i)}(1), \ldots, \tilde{\mathbf{r}}^{(i)}(K)\big\} \tag{130}$$

and

$$\big\{\tilde{\mathbf{D}}^{(i)}\big\}_1^K = \big\{\tilde{\mathbf{D}}^{(i)}(1), \ldots, \tilde{\mathbf{D}}^{(i)}(K)\big\}, \tag{131}$$

is computed by a fixed interval Kalman smoother operating on the state space model (124) and (125).

[21] The notation $\sim$ should be read as "is distributed according to."

Proof. See Appendix 2. □

Contrary to the "averaged" state space model, building of the state space model (124) and (125) does not require any actual processing at all. On the other hand, the vector Kalman filter operating on this state space model seems to require an inversion of the $J^{L+1} \times J^{L+1}$ matrix at each time step. However, by virtue of the diagonality of the covariance matrix of $\tilde{\mathbf{v}}_k$, the matrix inversions can be avoided by applying the sequential processing technique outlined in [11]. As a result, a computationally efficient CIR estimator, hereafter referred to as the SDD-KS estimator, is obtained.[22] The cross-coupling of this sequentially updated SDD-KS estimator iteratively with the BCJR algorithm realizes exactly the BEM algorithm and, hence, is subject to highly favorable convergence properties of the BEM algorithm, proved in [152].

Departing from the strict BEM formalism, further complexity reduction can be obtained by replacing the KS with the KF in the above formulation, leading to what is hereafter referred to as SDD-KF estimator. Computations required by the SDD-KF estimator to update the CIR estimate over one time step is summarized by the following corollary.

Corollary 1 *The computations involved in the SDD-KF estimator can be described with the following pseudocode:*[23] *For each time step,* $k = 1, \ldots, K$*, compute*

$$
\begin{aligned}
&\hat{\mathbf{h}}(k,0) = \boldsymbol{A}\hat{\mathbf{h}}(k-1)\\
&\mathbf{P}(k,0) = \boldsymbol{A}\mathbf{P}(k-1, J^{L+1})\boldsymbol{A}^H + \boldsymbol{G}\boldsymbol{G}^H\\
&\texttt{for} \quad j = 1, \cdots, J^{L+1}\\
&\qquad \mathbf{k} = \frac{\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^* \varphi(k,j)}{\sigma_v^2 + \boldsymbol{\xi}_j^T \mathbf{P}(k,j-1)\boldsymbol{\xi}_j^* \varphi(k,j)}\\
&\qquad \hat{\mathbf{h}}(k,j) = \hat{\mathbf{h}}(k,j-1) + \mathbf{k}\left(r(k) - \boldsymbol{\xi}_j^T \hat{\mathbf{h}}(k,j-1)\right)\\
&\qquad \mathbf{P}(k,j) = \mathbf{P}(k,j-1) - \mathbf{k}\,\boldsymbol{\xi}_j^T \mathbf{P}(k,j-1)\\
&\texttt{end}\\
&\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k, J^{L+1}).
\end{aligned}
$$

[22] See [163] for detailed analysis of the complexity pertaining to the sequential Kalman processing, as opposed to the direct vector Kalman processing.

[23] The first order ($p = 1$) AR channel model is assumed here. Generalization to any value of p can be obtained by replacing $\boldsymbol{\xi}_j^T$ with $[\boldsymbol{\xi}_j^T, \mathbf{0}_{1\times(p-1)(L+1)}]$ and $\hat{\mathbf{h}}$ with $\hat{\boldsymbol{h}}$ (see (7) and (38) for suitable definitions). In addition, the iteration indices have been suppressed for notational convenience.

The matrix $\mathbf{P}$ *denotes the error covariance matrix of the channel estimate, and* $\mathbf{k}$ *denotes the Kalman gain vector.*

Proof. This corollary follows directly by applying the sequential processing technique outlined in [11] to the vector Kalman filter operating on the state space model (124) and (125). However, for the sake of clarity and easy access, the proof is also briefly sketched below.

The sequential processing refers to the computational procedure in which the measurement vector (125) is processed one component at a time. Since the covariance matrix of the noise vector $\tilde{\mathbf{v}}(k)$ is diagonal, it makes sense to partition the measurement equation (125) into components as follows:

$$r(k)\sqrt{\varphi(k,j)} = \boldsymbol{\xi}_j^T \sqrt{\varphi(k,j)}\mathbf{h}(k) + \tilde{v}_j(k) \tag{132}$$

for $j = 1, 2, \ldots, J^{L+1}$, where $\tilde{v}_j(k)$ denotes the jth element of $\tilde{\mathbf{v}}(k)$ and its variance is given by $E[\tilde{v}_j(k)\tilde{v}_j^*(k)] = \sigma_v^2$. Sequential processing simply implies that, instead of processing $\tilde{\mathbf{r}}(k)$ as a single observation vector, the elements of $\tilde{\mathbf{r}}(k)$ are processed one at a time. Thus, instead of calculating

$$\hat{\mathbf{h}}(k|k) \triangleq E\Big[\mathbf{h}(k)\big|\{\tilde{\mathbf{r}}\}_1^k\Big] = E\Big[\mathbf{h}(k)\big|\{\tilde{\mathbf{r}}\}_1^{k-1}, \tilde{\mathbf{r}}(k)\Big] \tag{133}$$

in terms of $\hat{\mathbf{h}}(k|k-1)$ and $\tilde{\mathbf{r}}(k)$ as usually, first the quantity

$$\hat{\mathbf{h}}(k,1) = E\Big[\mathbf{h}(k)\big|\{\tilde{\mathbf{r}}\}_1^{k-1}, r(k)\sqrt{\varphi(k,1)}\Big] \tag{134}$$

is calculated in terms of $\hat{\mathbf{h}}(k|k-1)$ and $r(k)\sqrt{\varphi(k,1)}$, then

$$\hat{\mathbf{h}}(k,2) = E\Big[\mathbf{h}(k)\big|\{\tilde{\mathbf{r}}\}_1^{k-1}, r(k)\sqrt{\varphi(k,1)}, r(k)\sqrt{\varphi(k,2)}\Big] \tag{135}$$

is calculated in terms of $\hat{\mathbf{h}}(k,1)$ and $r(k)\sqrt{\varphi(k,2)}$, and so on until $\hat{\mathbf{h}}(k|k) = \hat{\mathbf{h}}(k, J^{L+1}) \doteq \hat{\mathbf{h}}(k)$. The intermediate estimates $\hat{\mathbf{h}}(k,j)$ for $j = 1, 2, \ldots, J^{L+1}$ are achieved by a direct application of the Kalman filtering equations to the measurement equation (125), regarding j as a running variable. Hence, the SDD-KF estimator can be realized by the computational procedure given by the above pseudocode. □

In essence, the SDD-KF is updated sequentially over the *time* and the *branch* indices in a similar manner as was done in the case of the SDD-RLS estimator. Thus, analogous to the data-aided mode in which case the close relationship between the RLS and KF estimators have been identified earlier [106, 219], the SDD-RLS and SDD-KF

estimators structurally resemble each other to an amazing extent. However, despite the structural similarities, the derivations leading to these two estimators required essentially different type of reasoning. On the other hand, these two estimators also differ in two important aspects. First, unlike the SDD-RLS estimator, the SDD-KF estimator can exploit the knowledge of the channel second order statistics and, secondly, it can also take into account the possible correlation between the channel taps. These two factors potentially give the SDD-KF estimator some performance advantage over the SDD-RLS estimator, particularly if the channel taps are correlated and correlation is known to the receiver.

Finally, the SDD-KF estimator can be embedded into the known FI and FL SISO algorithms as easily as the SDD-LMS/RLS estimators, resulting in what will hereafter be referred to as an APP-SDD-KF algorithm. Next, however, computationally even simpler SDD estimators for random Gaussian channels are pursued by making use of the notion of "per-channel-path-based estimation."

4.3.2 Demodulators via modified BEM algorithm

A reduced complexity per-survivor processing (PSP) Kalman filter bank for estimating a fast-fading frequency-selective channel was proposed in [215]. Motivated by the good performance of that PSP-based estimator, a novel soft decision directed reduced complexity Kalman smoother/filter (SDD-RCKS/KF) will, in this subsection, be derived for equal-energy signal constellations. The reduced complexity PSP Kalman processing, in the form proposed in [215], relies on the assumption of a wide sense stationary uncorrelated scattering (WSSUS) channel model, which inherently implies that the cross correlation terms of the matrices $\boldsymbol{A}$ and $\boldsymbol{G}$ are zero. Consequently, the channel process $\boldsymbol{h}(k)$ in (35) can be decomposed into $L+1$ independently fading channel tap gains. As will be discussed shortly, this requirement is, however, unnecessarily strict for the reduced complexity SDD tracker to be derived below. In fact, a significant complexity reduction can be obtained without resorting to such an assumption.

The complexity reduction in the SDD Kalman processing is based on the idea of decomposing the received signal samples into independent multipath components whose estimated values are called pseudo observations, i.e.,

$$r(k) = \boldsymbol{s}^T(k)\boldsymbol{h}(k) + v(k) = \sum_{l=1}^{L+1} x_l(k), \tag{136}$$

where the multipath component $x_l(k)$ is defined as

$$x_l(k) \triangleq s(k-l+1)h_{l-1}(k) + \ddot{v}_l(k). \tag{137}$$

Guided by the ideas presented in [72], the receiver noise sample $v(k)$ is arbitrarily decomposed into the $L+1$ components $\ddot{v}_l(k)$ so that $\sum_{l=1}^{L+1} \ddot{v}_l(k) = v(k)$. The noise components $\ddot{v}_l(k)\ \forall l$ are *designed* to be statistically independent, zero-mean, and Gaussian with the variance $\sigma^2_{\ddot{v}_l} = \eta\sigma_v^2$, where $\eta = 1/(L+1)$.

Since there is no direct access to the multipath components $x_l(k)\ \forall l,k$, they can be regarded as a "missing" data from the BEM algorithm point of view. Based on this, a foundation for the SDD-RCKS/KF is set up by the following lemma. This actually provides another implementation for the BEM algorithm by using a more informative specification for the complete data set than was assumed by the BEM algorithm in Theorem 1.

Lemma 2 *The MAP sequence estimate of $\boldsymbol{\Theta}$, as defined in (129), at the ith iteration of the BEM algorithm, when the complete data set is defined as $\mathcal{X} = \left\{\{x_l(k)\}_{l=1,\dots,L+1}^{k=1,\dots,K}, \mathbf{s}\right\}$ and the incomplete data set as $\mathcal{Y} = \{\mathbf{r}\}$, is computed by a fixed interval vector KS operating on the state space model where the state equation is given by (124) and the measurement equation is given by*[24]

$$\tilde{\mathbf{x}}^{(i)}(k) = \tilde{\mathbf{D}}_e^{(i)}(k)\mathbf{h}(k) + \tilde{\mathbf{v}}_e(k). \tag{138}$$

In (138), $\tilde{\mathbf{x}}^{(i)}(k)$ is given as

$$\tilde{\mathbf{x}}^{(i)}(k) = \left[\left(\tilde{\mathbf{x}}_1^{(i)}\right)^T(k), \dots, \left(\tilde{\mathbf{x}}_{L+1}^{(i)}\right)^T(k)\right]^T \in \mathbb{C}^{(L+1)J^{L+1}}, \tag{139}$$

$$\tilde{\mathbf{x}}_l^{(i)}(k) = \left[\tilde{x}_{l,1}^{(i)}(k), \dots, \tilde{x}_{l,J^{L+1}}^{(i)}(k)\right]^T \in \mathbb{C}^{J^{L+1}}, \tag{140}$$

and

$$\tilde{x}_{l,j}^{(i)}(k) = \hat{x}_{l,j}^{(i)}(k)\sqrt{\varphi^{(i)}(k,j)}, \tag{141}$$

where

$$\hat{x}_{l,j}^{(i)}(k) = \left[\boldsymbol{\xi}_j\right]_l \hat{h}_{l-1}^{(i-1)}(k) + \eta\Big(r(k) - \boldsymbol{\xi}_j^T\hat{\mathbf{h}}^{(i-1)}(k)\Big) \tag{142}$$

[24]For notational simplicity, it is assumed that $p = 1$.

and $\left[\boldsymbol{\xi}_j\right]_l$ *denotes the* l*th element of* $\boldsymbol{\xi}_j$*. Moreover,*

$$\tilde{\mathbf{D}}_e^{(i)}(k) = diag\Big(\left[\tilde{\mathbf{D}}^{(i)}(k)\right]_1, \ldots, \left[\tilde{\mathbf{D}}^{(i)}(k)\right]_{L+1}\Big) \in \mathbb{C}^{(L+1)J^{L+1}\times(L+1)}, \tag{143}$$

where $\left[\tilde{\mathbf{D}}^{(i)}(k)\right]_l$ *denotes the* l*th column of* $\tilde{\mathbf{D}}^{(i)}(k)$ *and*

$$\tilde{\mathbf{v}}_e(k) \sim \mathcal{CN}\big(\tilde{\mathbf{v}}_e(k), \mathbf{0}_{(L+1)J^{L+1}\times 1}, \eta\sigma_v^2 \times \mathbf{I}_{(L+1)J^{L+1}}\big). \tag{144}$$

Proof. See Appendix 2. □

The convergence rates of the two BEM algorithms, presented by Theorem 1 and Lemma 2, are not equal in general since the convergence rate of the EM and its Bayesian counterpart depends on the specification of a complete data set [161]. Specifically, the BEM algorithm given by Lemma 2 can be expected to have a slower convergence speed because its complete data set includes an augmented set of the hidden variables, i.e., the complete data set is more informative. In addition, there is no guarantee that these two BEM algorithms convergence to the same stationary point of the posterior distribution. Also, the latter BEM algorithm is notably more complex than the first one, unless a WSSUS channel model is assumed. Under the WSSUS assumption, a bank of scalar KS will produce the desired channel estimates.

In the case of a constant envelope PSK modulation, the BEM algorithm given by Lemma 2 will always reduce to the low-complexity format provided by the following theorem (for the special case of $p = 1$), regardless of the WSSUS channel assumption.

Theorem 2 *Suppose that the complete and incomplete data sets are defined as* $\mathcal{X} = \big\{\{x_l(k)\}_{l=1,\ldots,L+1}^{k=1,\ldots,K}, \mathbf{s}\big\}$ *and* $\mathcal{Y} = \{\mathbf{r}\}$*, respectively, and that* $|s(k)|^2 = 1,\ \forall k$*. Then, the MAP sequence estimate of* $\boldsymbol{\Theta}$ *is computed, at the* i*th iteration of the BEM algorithm, by a fixed interval vector KS operating on the state space model, which consists of the state equation given by (124) and the measurement equation expressed as*

$$\hat{\mathbf{x}}^{(i)}(k) = \mathbf{h}(k) + \tilde{\mathbf{v}}_e(k), \tag{145}$$

where the vector of pseudo-observations $\hat{\mathbf{x}}^{(i)}(k)$ *is defined as*

$$\hat{\mathbf{x}}^{(i)}(k) = \sum_{j=1}^{J^{L+1}} diag\big(\boldsymbol{\xi}_j\big)\big[\hat{x}_{1,j}^{(i)}(k), \ldots, \hat{x}_{L+1,j}^{(i)}(k)\big]^T \varphi^{(i)}(k,j)\ \in \mathbb{C}^{L+1}, \tag{146}$$

and $\tilde{\mathbf{v}}_e(k) \sim \mathcal{CN}\big(\tilde{\mathbf{v}}_e(k), \mathbf{0}_{(L+1)\times 1}, \eta\sigma_v^2 \times \mathbf{I}_{(L+1)}\big)$.

Proof. See Appendix 2. □

At the E-step, the branch APPs $\varphi^{(i)}(k,j)$ are updated by the standard BCJR algorithm by using the most recent channel estimate. A block diagram of this reduced complexity BEM algorithm is shown in Fig. 15.

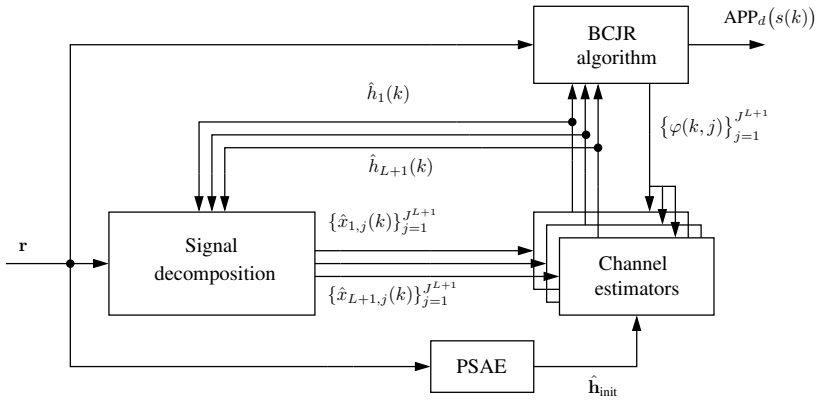

Fig 15. Block diagram of the reduced complexity BEM algorithm based on decomposing the received signal into independent multipath components.

The sequential MAP channel estimates by Theorem 2 can be embedded into the known FI and FL SISO algorithms if the KS is replaced with the KF and the per-trellis-branch-based pseudo-observations given by (142) are defined in terms of the predicted channel estimates, instead of using the smoothed channel estimates from the previous iteration. Thus, the pseudo-observations $\hat{x}_{l,j}^{(i)}(k)$ are replaced by the predicted pseudo-observations $\hat{x}_{l,j}^{(i)}(k|k-1)$ defined as

$$\hat{x}_{l,j}^{(i)}(k|k-1) \triangleq \big[\boldsymbol{\xi}_j\big]_l \hat{h}_{l-1}^{(i)}(k|k-1) + \eta\Big(r(k) - \boldsymbol{\xi}_j^T \hat{\mathbf{h}}^{(i)}(k|k-1)\Big), \tag{147}$$

where $\hat{\mathbf{h}}^{(i)}(k|k-1) = E\Big[\mathbf{h}(k)\big|\big\{\tilde{\mathbf{x}}^{(i)}\big\}_1^{k-1}, \big\{\tilde{\mathbf{D}}_{\mathrm{e}}^{(i)}\big\}_1^{k-1}\Big]$. Interestingly, the pseudo-observations in [215] are obtained from (147) by setting $\eta = 1$ and replacing the hypothesized data vector $\boldsymbol{\xi}_j$ with the tentative data-decisions along with the survivor-path of the trellis.

By inserting (147) into (146) and applying the sequential Kalman processing onto the resulting state space model, an A-SISO algorithm referred to as the APP-SDD-RCKF algorithm is found. It contains an embedded SDD Kalman filter which can be

described at time k by the following pseudocode (assuming a first order AR channel model):

$$\hat{\mathbf{h}}(k,0) = \boldsymbol{A}\hat{\mathbf{h}}(k-1)$$

$$\mathbf{P}(k,0) = \boldsymbol{A}\mathbf{P}(k-1,L+1)\boldsymbol{A}^H + \boldsymbol{G}\boldsymbol{G}^H$$

`for` $l = 1, \cdots, L+1$

$$\mathbf{k} = \frac{[\mathbf{P}(k,l-1)]_l}{\eta\sigma_v^2 + [\mathbf{P}(k,l-1)]_{l,l}}$$

$$\hat{\mathbf{h}}(k,l) = \hat{\mathbf{h}}(k,l-1) + \eta\,\mathbf{k}\sum_{j=1}^{J^{L+1}}\Big(r(k) - \boldsymbol{\xi}_j^T\hat{\mathbf{h}}(k,l-1)\Big)\big[\boldsymbol{\xi}_j^*\big]_l\varphi(k,j)$$

$$\mathbf{P}(k,l) = \mathbf{P}(k,l-1) - \mathbf{k}\big[\mathbf{P}^T(k,l-1)\big]_l^T$$

`end`

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k,L+1).$$

In a Rayleigh flat fading channel, the SDD-RCKF estimator essentially coincides with the SDD-KF algorithm presented in [262]. If the off-diagonal elements of the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ and $\mathbf{G}$ are zero, implying uncorrelated channel tap gains, the SDD-RCKF estimator is effectively implemented by running only one recursion per time step. In that case, it closely resembles, both structurally and from the computational requirements point of view, a *normalized* SDD-LMS estimator incorporating an adaptive step size parameter. Computations related to the SDD-KF and SDD-RCKF estimators at time k are summarized in Table 4.

Table 4. Update equations for SDD-KF and SDD-RCKF estimators at time k.

SDD-KF

$$\mathbf{P}(k,0) = \boldsymbol{A}\mathbf{P}\big(k-1, J^{L+1}\big)\boldsymbol{A}^H + \boldsymbol{G}\boldsymbol{G}^H$$
$$\hat{\mathbf{h}}(k,0) = \boldsymbol{A}\hat{\mathbf{h}}(k-1)$$
For $\quad j = 1, \ldots, J^{L+1}$

$$\mathbf{k} = \frac{\mathbf{P}(k,j-1)\boldsymbol{\xi}_j^* \varphi'(k,j)}{\sigma_v^2 + \boldsymbol{\xi}_j^T \mathbf{P}(k,j-1)\boldsymbol{\xi}_j^* \varphi'(k,j)}$$
$$\hat{\mathbf{h}}(k,j) = \hat{\mathbf{h}}(k,j-1) + \mathbf{k}\Big(r(k) - \boldsymbol{\xi}_j^T \hat{\mathbf{h}}(k,j-1)\Big)$$
$$\mathbf{P}(k,j) = \mathbf{P}(k,j-1) - \mathbf{k}\boldsymbol{\xi}_j^T \mathbf{P}(k,j-1)$$

end

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k, J^{L+1})$$

SDD-RCKF

$$\mathbf{P}(k,0) = \boldsymbol{A}\mathbf{P}(k-1, L+1)\boldsymbol{A}^H + \boldsymbol{G}\boldsymbol{G}^H$$
$$\hat{\mathbf{h}}(k,0) = \boldsymbol{A}\hat{\mathbf{h}}(k-1)$$
For $\quad l = 1, \ldots, L+1$

$$\mathbf{k} = \frac{\big[\mathbf{P}(k,l-1)\big]_l}{\eta\sigma_v^2 + \big[\mathbf{P}(k,l-1)\big]_{l,l}}$$
$$\hat{\mathbf{h}}(k,l) = \hat{\mathbf{h}}(k,l-1) + \eta\,\mathbf{k}\sum_{j=1}^{J^{L+1}}\Big(r(k) - \boldsymbol{\xi}_j^T \hat{\mathbf{h}}(k,l-1)\Big)\big[\boldsymbol{\xi}_j^*\big]_l \varphi'(k,j)$$
$$\mathbf{P}(k,l) = \mathbf{P}(k,l-1) - \mathbf{k}\Big(\big[\mathbf{P}^T(k,l-1)\big]_l\Big)^T$$

end

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k, L+1)$$

A decoupled channel estimator bank, based on a similar decomposition of the received signal into multipath components, has also been proposed in an independent work reported in [29]. As opposed to the SDD-RCKF estimator described above, the

estimator in [29], however, has some disadvantages. Specifically, it is applicable only to deterministic time-invariant channels and, since the channel estimation is based on a block-processing, the associated demodulator must iterate between estimation and inference.

4.4 Linear adaptive SISO detectors via MF-EM algorithm

The EM- or BEM-based adaptive demodulators, in their exact forms, may not be computationally feasible for broadband applications, where the CIR typically spans many symbol intervals. Hence, lower complexity alternatives are of great interest. The complexity of the EM- and BEM-based demodulators can be reduced by performing either the E-step or the M-step, or both, approximately. Approximate M-step may produce reduced-complexity channel estimators, as was shwon in the previous section. Indeed, the SDD-LMS and SDD-RCKF estimators can be regarded as realizations of this strategy.

An approximate E-step can be obtained by performing a maximization of $\mathcal{G}\big(Q(\mathbf{s}), \hat{\mathbf{h}}\big)$ (see (94)) or a minimization of $\mathcal{F}(Q)$ over a restricted class of trial distributions $Q(\mathbf{s})$. The resulting algorithm is dubbed a variational EM algorithm.[25] The mean-field EM (or BEM) algorithm is obtained by restricting $Q(\mathbf{s})$ to be fully factorized, in which case the inference is performed by the linear SISO detectors.

Assuming that the channel is deterministic and time-invariant, it follows from (92) and (95) that the channel estimate can be obtained as (iteration indices are suppressed)

$$\hat{\mathbf{h}} = \arg\max_{\mathbf{h}} E_Q \left\{ \ln p(\mathbf{s}, \mathbf{r}|\mathbf{h}) \right\}, \tag{148}$$

where the expectation may be taken under the MF distribution $Q = \prod_k q_k(s(k))$ or under the Bayesian MF distribution $Q = \prod_k \mathcal{CN}\big(s(k), \hat{s}(k), \sigma^2_{o,k}\big)$, implying that the linear SISO detector is applied for detection. A straightforward formulation yields

$$\hat{\mathbf{h}} = \left(\sum_k \overline{\mathbf{s}}^*(k) \overline{\mathbf{s}}^T(k) + \mathbf{\Sigma}_k \right)^{-1} \left(\sum_k \overline{\mathbf{s}}^*(k) r(k) \right), \tag{149}$$

where $\overline{\mathbf{s}}(k) = \big[\overline{s}(k), \ldots, \overline{s}(k-L)\big]^T$ and $\mathbf{\Sigma}_k = \text{diag}(\sigma^2_{s_k}, \ldots, \sigma^2_{s_{k-L}})$. The soft symbol value at time k is given as $\overline{s}(k) = E_{q_k^{(i)}}\big[s(k)\big]$ and its variance $\sigma^2_{s_k}$ is given as $\sigma^2_{s_k} = E_{q_k^{(i)}}\big[|s(k)|^2\big] - |\overline{s}(k)|^2$.

[25] With a broader definition, the variational EM algorithms can be regarded as a subclass of approximate EM algorithms, where the E-step is computed by using any variational inference method.

Recursive SDD channel estimators derived in the previous section can be easily modified so as to allow the lower complexity linear SISO detector to execute an approximate E-step. Specifically, the branch APPs are approximated as follows:

$$\varphi(k,j) \approx \prod_{l=1}^{L+1} \breve{\varphi}(k-l+1,i), \quad i \in \mathcal{A}_{k,l}(j), \tag{150}$$

where $\mathcal{A}_{k,l}(j) = \{i \in \{1,\ldots,J\} : \ s(k-l+1) = \alpha_i \text{ is consistent with } \mathbf{s}(k) = \boldsymbol{\xi}_j\}$. Furthermore, in the case of the LCMF SISO detector $\breve{\varphi}(k,i) = q_k(s(k) = \alpha_i)$, whereas in the case of the MMSE SISO detector $\breve{\varphi}(k,i) = \mathcal{CN}\big(s(k) = \alpha_i, \hat{s}(k), \sigma^2_{o,k}\big)$. Unfortunately, the approximate E-step via (150) does not automatically entail any complexity reduction for the SDD channel estimation. An exception is the SDD-LMS estimator, which under the approximation (150) takes the following low-complexity format:

$$\hat{\mathbf{h}}^T(k) = \hat{\mathbf{h}}^T(k-1) + \mu\left[\Big(r(k) - \hat{\mathbf{h}}^T(k-1)\overline{\mathbf{s}}(k)\Big)\overline{\mathbf{s}}^H(k) - \hat{\mathbf{h}}^T(k-1)\boldsymbol{\Sigma}_k\right]. \tag{151}$$

It is readily noticed that if the probabilistic information on the data symbols is perfect, the estimator (151) reduces to the standard LMS algorithm, whereas if no information on the channel symbols is available, it follows from (151) that $\hat{\mathbf{h}}(k) = (1-\mu)\hat{\mathbf{h}}(k-1)$, i.e., the channel estimate "dies out" in a time period determined by the parameter μ. On the other hand, assuming that soft data decisions are reliable enough so that the covariance matrix $\boldsymbol{\Sigma}_k$ can be approximated to be a zero matrix, all the SDD channel estimators proposed in the previous section assume a low-complexity format. Specifically, they are obtained directly from the corresponding data-aided estimators by simply replacing *a priori* known data symbols by their averaged values where the averaging is taken under the MF distribution. For example, the SDD-LMS estimator assumes a format

$$\hat{\mathbf{h}}^T(k) = \hat{\mathbf{h}}^T(k-1) + \mu\left[\Big(r(k) - \hat{\mathbf{h}}^T(k-1)\overline{\mathbf{s}}(k)\Big)\overline{\mathbf{s}}^H(k)\right]. \tag{152}$$

The low-complexity SDD channel estimator similar to (151) as well as some related estimators, which were based upon the assumption that the variance matrix $\boldsymbol{\Sigma}_k$ is zero, have actually been proposed earlier in [195] by using heuristic reasoning. However, a new formulation of these known SDD estimators via the MF-EM algorithm provides a solid theoretical foundation for them. Also, it gives new insight into the optimality of these estimators, revealing, for example, that their convergence can be characterized in terms of the lower bound rather than in terms of the likelihood, as discussed in Section 4.1. Other variants of SDD channel estimators in the context of linear SISO detectors are presented in [225, 143].

4.5 Complexity considerations

Evaluating the complexity of the adaptive SISO detectors in the presence of an unknown channel is at least as difficult as it was in the case of perfect CSI. Of course, the added computational requirements related to the adaptive detectors are due to the unknown channel, which has to be estimated either implicitely or explicitely. However, the complexities pertaining to detection and channel estimation cannot, in general, be quantified separately in complexity analysis since the interplay between these operations affects the overall receiver architecture significantly.

When the CSI is not available at the receiver, the *exact* APP symbol detector in most cases has an exponential complexity in the channel coherence time. Thus, for channels having coherence time on the order of the sequence length, the complexity appears to be exponential in the sequence length, irrespective of the frequency-selectivity of the channel [8]. An exception for this general rule is obtained in the case of equal energy constellations transmitted over flat Rayleigh fading channels. Then, the exact symbol-by-symbol soft-decision metrics, under the min-sum criterion, can be evaluated with polynomial worst case complexity, as shown in [175].

The exact APP symbol detector can be formulated so that it consists of likelihood updates on a forward and backward tree, assisted by per-path channel estimators, and binding of the correspondig metrics of the past and future [8]. So, in the light of this fact, suboptimal SISO detectors can be obtained as the result of applying one or more of the following simplifications: 1) nonexhausive tree search, 2) suboptimal channel estimators, and 3) suboptimal binding of the past and future metrics. Each of these approaches was explored to some extent in [8]. Regarding the first of the listed simplifications, there exist many ways to prune the sequence tree, the earliest of which have actually been considered for hard-decision devices [12, 209]. Especially the PSP-principle [209] has in recent years gained popularity in research communities and industry.

The fixed complexity APP SISO detector, based on the PSP-principle, can be characterized by the following operations: the sum-product terms are computed for the forward and backward running metrics over the entire sequence of received samples, and a separate channel estimate is kept for every state of the trellis while the estimates are updated in a PSP fashion. It is important to notice at this point that the trellis on which the PSP algorithm operates does not need to be tightly related to the trellis of the known channel SISO algorithm. In contrast, its size is, in general, a design parameter that determines the amount of pruning in the forward and backward trees, and, eventually, the

complexity of the PSP detector. On the other hand, for a fixed number of trellis states, the number of channel estimators updated at every recursion may be a compromise between the two extremes of tentative decisions, implying only one channel estimator at one end and the PSP at the other [208]. In addition, the channel estimation, using, e.g., PSP, can be performed along with forward recursions only or in both forward and backward directions [108]. So, roughly speaking, the complexity of the PSP-based SISO detector is determined by the size of the associated trellis and the number of channel estimators updated at every recursion, together.

The EM- and BEM-based SISO detectors, derived earlier in this chapter, operate on the fixed-size trellis, which is exactly the same as the one associated with the known channel SISO detector. Since there is only one channel estimator to be updated at every recursion, there is essentially no performance advantage to be gained by expanding the trellis beyond the size determined by the number of channel tap gains, i.e., J^L. Effectively, the proposed SDD channel estimators, whether of the LMS, RLS, or KF type, are updated at every time recursion by *averaging* the innovation part of the estimator over the trellis transition probabilities. Thus, from the complexity point of view, the SDD channel estimators can be considered to be comprised of J^{L+1} separate channel estimators, which are to be updated separately at every time step. So, summing up, the APP-SDD-LMS/RLS/KF detectors operate on the fixed-size trellis, but the SDD channel estimators have complexity which is on the order of combined complexity of J^{L+1} traditional channel estimators.

The SDD-RCKF estimator, being based on the fully factorized trellis transition probabilities,[26] has a complexity which is comprable with the combined complexity of $L+1$ traditional channel estimators. For example, in the case of QPSK modulated symbols transmitted over a three-tap fading channel, a reduction from 64 to 3 recursions over the trellis branches is achieved when compared to the SDD-KF estimator. Thus, the overall complexity reduction obtained by the RC Kalman processing can indeed be significant. For comparison, the PSP-based SISO detector proposed in [215] updates J^L Kalman filters at every time recursion and, therefore, its complexity even for moderate channel lengths ($L > 1$) is much higher than the complexity of the APP-SDD-RCKF algorithm. However, as in the known channel case, the adaptive linear SISO detectors have much lower complexity than any of the aforementioned trellis-based SISO detectors.

[26]Note that the channel estimation is here based on the fully factorized distribution, not the inference as in the case of liner SISO detectors.

4.6 Numerical results

The performance of the proposed EM- and BEM-based adaptive SISO algorithms was evaluated by computer simulations in a single-antenna system. For comparison, the performance of the conventional Viterbi detector with delayed (the delay being 10 symbol intervals) LMS estimator (VD-LMS) and of various PSP-based detectors as well as the performance of the APP detectors with perfect CSI were also simulated. It should be noted that the PSP-based SISO detectors considered in this section assume a forward-only channel estimation while the number of trellis states is tightly related to the number of channel tap gains, i.e., the number of states is J^L.

Specifically, the performance of uncoded and coded transmission of QPSK symbols over a discrete-time symbol-spaced frequency-selective Rayleigh fading channel was investigated by using Clarke's channel model [46] with normalized Doppler spread B_dT_s varying between values $B_dT_s = 0.001$ and $B_dT_s = 0.05$. The channel had three independently fading tap gains whose standard deviations were set at ($\sigma_{\text{tap1}} = 1/\sqrt{6}$, $\sigma_{\text{tap2}} = 2/\sqrt{6}$ and $\sigma_{\text{tap3}} = 1/\sqrt{6}$). The fading channel tap gains were realized by using the method described in [121], while their first order AR descriptions were obtained by extracting the AR(1) parameters directly from the simulated channel tap gains using the standard Yule-Walker method.

The transmitted symbols were organized into fixed-size bursts with K information symbols preceded by K_{pre} known preamble symbols and followed by K_{tail} known tail symbols. Thus, the starting and terminating states of the demodulator trellis were known. In addition, the known preamble symbols were used in obtaining the initial CIR estimate.

In adaptive SISO demodulators, the step size parameter of all LMS estimators was set at $\mu = 0.1$, while the forgetting factor of all RLS estimators were set at $\lambda = 0.9$, irrespective of the rate of variation of the channel. This is not, of course, an optimal choice in all channel dynamics. Moreover, the state-space models for Kalman-type channel estimators were based on the estimated AR(1) parameters. Such modeling also has its own demerits. First, the estimation of the AR(1) parameters may be very difficult due to fact the signal bandwidth is typically much larger than the Doppler bandwidth of the channel, causing serious oversampling of the channel, and, secondly, the symbol rate AR(1) model itself may not properly represent the time-continuous channel. Regardless of the above distracting arguments, the performance results provided below do, however, give some general ideas and hints about the relative goodness of the proposed

adaptive detectors, as compared to some of the existing ones. It is, however, important to note that, due to inadequacy in modeling, the comparisons made between different detectors may not be fully conclusive.

The block diagram of the communication system for coded transmission is shown in Fig. 16. A 64 state convolutional encoder (CE) with rate 1/2 (code generators in octal form are (133,171)) was used to encode the information bits. The starting and terminating states of the encoder were assumed to be known in order to facilitate the decoding algorithms. After the QPSK mapper, the symbols were interleaved using a $K \times K$ block interleaver (BI) in conjunction with the assumption of a burst-to-burst independent channel. Specifically, the data symbols were read into the BI row-wise while read out column-wise so that each data burst in the channel consists of data symbols of one column of BI, preceded by the known preamble symbols and followed by the known tail symbols. Each such data burst was then assumed to be subject to uncorrelated fading in the channel. At the receiver side, the adaptive SISO demodulator and the SISO decoder were iteratively connected through the block interleaver and the block deinterleaver (BD). Finally, the soft output information about the information bits obtained from the SISO decoder was quantized to obtain the estimated sequence of transmitted information bits.

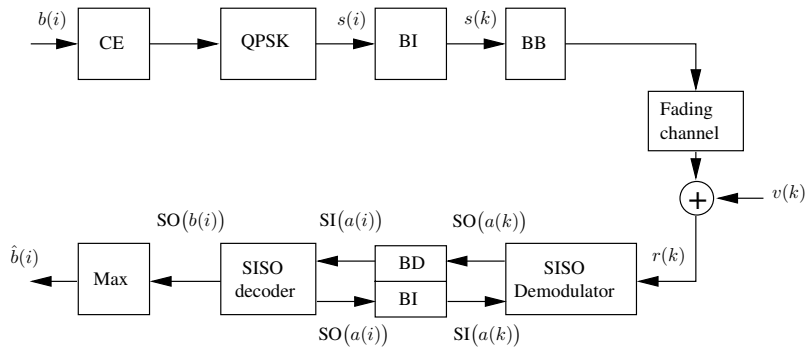

Fig 16. Block diagram of the coded communication system with the SISO demodulator and the SISO decoder iteratively connected by using the turbo principle. A burst builder (BB) is used to add the known preamble and tail symbols to each data burst. The acronyms SI and SO stand for soft input and soft output, respectively.

The bit error rate of various uncoded and coded QPSK transmission systems was evaluated as a function of E_b/N_0, where E_b denotes the transmitted signal energy per information bit. In all simulations, E_b/N_0 values were defined by taking into account the loss due to the training symbols (i.e., known preamble and tail symbols). The BER results for several uncoded QPSK transmission systems with different LMS-based detectors are presented in Fig. 17. The fading rate was $B_dT_s = 0.001$, and the burst parameters were as follows: $K_{\text{pre}} = 5$, $K = 30$ and $K_{\text{tail}} = 2$. The performance of an iterative detector, combining the SDD-LMS estimator and the standard BCJR algorithm together iteratively (cf. Eq. (108) and (109)) with six iterations, is shown along with the performance of the noniterative APP-SDD-LMS detector. The former detection algorithm is referred to as the LMS-based adaptive Baum-Welch (ABW-LMS) algorithm in the figure. While the APP-SDD-LMS detector behaves similarly to the conventional LMS-based Viterbi detector, the performance of the ABW-LMS detector with six iterations approaches the performance of the PSP-LMS detector, where the PSP technique is applied to the trellis with $J^L = 16$ states in a forward-only manner. A possible explanation for the relatively poor performance of the ABW-LMS and APP-SDD-LMS detectors is the slow convergence speed due to the averaging over all branches of the detector trellis [13]. Since the transmitted bursts as well as the preamble were fairly short, the convergence speed can have a significant impact on the performance. Short preamble causes large initial uncertainty in the channel estimates, which is also demonstrated by a relatively large gap in performance between the PSP detector and the Viterbi detector with perfect CSI.

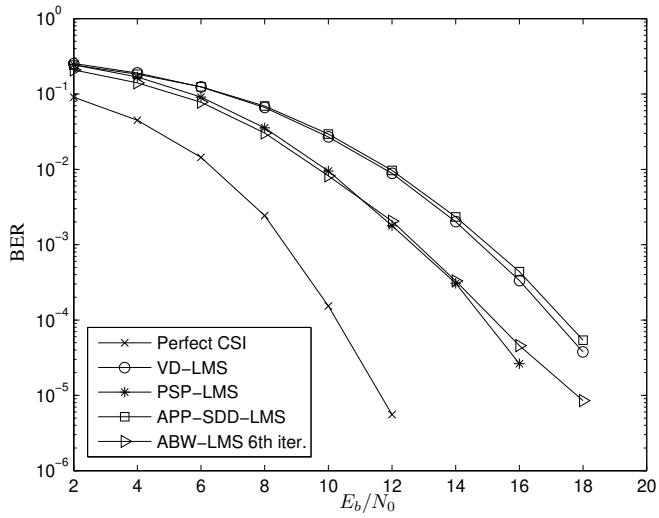

Fig 17. BER versus E_b/N_0 for an uncoded QPSK system with a fading rate of $B_dT_s = 0.001$.

The performance results for various APP algorithms with different embedded SDD channel estimators are presented for an uncoded transmission system in Fig. 18. Here, the same channel conditions and the same burst structure is used as for the results in Fig. 17, except that the size of the preamble is now extended to ten pilot symbols ($K_{\text{pre}} = 10$). The simulated performance is also shown for a demodulator which acquires the CIR estimate by using only the pilot symbols at the beginning of each burst (referred to as the APP-PSAE algorithm). Although the initial channel uncertainty was reduced, all the APP-SDD detectors exhibited only a moderate performance gain compared to the performance of the APP-PSAE detector. A conclusion drawn up from these two performance figures is that the demodulators exploiting the soft statistics of the data symbols do not pay back the increased computational effort, at least not in uncoded transmission systems.

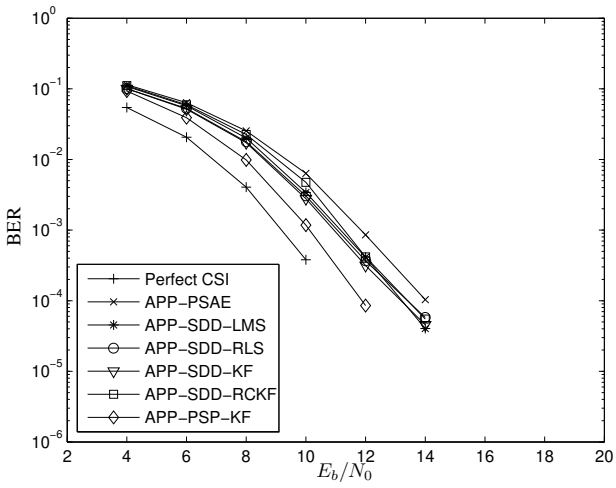

Fig 18. BER versus E_b/N_0 for an uncoded QPSK system when noniterative A-SISO demodulators are used.

In contrast to uncoded systems, the performance of coded transmission systems seems to be significantly improved when SDD channel estimators are employed instead of the hard decision-directed CIR estimators. In Fig. 19, the BER as a function of E_b/N_0 is presented for coded systems employing an iterative turbo receiver in which noniterative adaptive SISO demodulators of various kinds and a SISO decoder are iteratively interconnected as shown in Fig. 16. The system parameters were as follows: $B_dT_s = 0.001$, $K_{\text{pre}} = 5$, $K = 30$, and $K_{\text{tail}} = 2$. For comparison, the performance of the serial concatenation of the PSP-LMS detector and the hard-decision Viterbi decoder (PSP-VD) is presented. Due to the soft decoding gain, the performance of the iterative turbo receiver with the APP-SDD-LMS demodulator was better than the performance of the PSP-VD, even after the first iteration. When the number of iterations is increased, the superiority of the turbo receivers is clearly seen. It can also be noticed that the APP-SDD-RLS demodulator exhibits about a 1 dB performance gain with respect to the APP-SDD-LMS demodulator. Since the initial channel uncertainty was high, the better performance of the SDD-RLS estimator is mainly due to its faster convergence speed. The high initial channel uncertainty also caused a large performance loss for all the studied adaptive SISOs as compared to the SISO algorithm with perfect CSI.

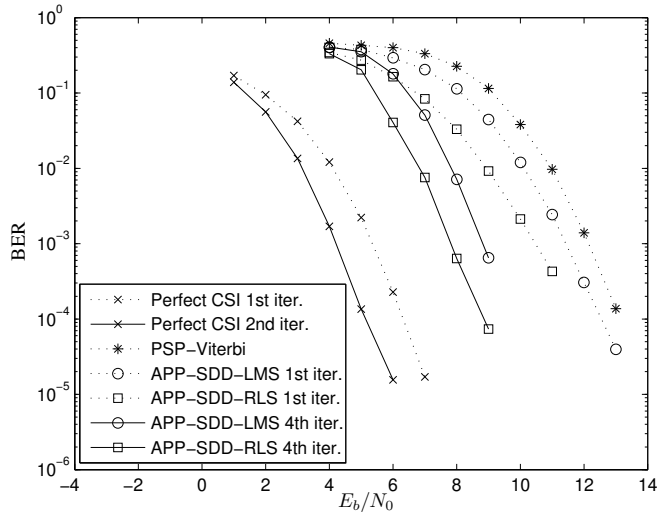

Fig 19. BER versus E_b/N_0 for a coded QPSK system with fading rate $B_dT_s = 0.001$ and $K = 30$.

The performance gap between the adaptive SISO demodulators and the SISO demodulators with the perfect CSI can be reduced by increasing the size of the preamble to 10 symbols, as demonstrated in Fig. 20. Also, the advantage of the faster convergence speed of the SDD-RLS estimator is almost lost when the quality of the initial CIR estimate is improved. The APP-PSP-KF is a forward-only version of the PSP-based adaptive SISO detector (assuming 16 states) presented in [8]. Interestingly, the APP-SDD-RCKF demodulator achieved the same performance as the APP-PSP-KF demodulator, while its computational complexity is significantly smaller, as discussed earlier. Somewhat surprisingly, however, the SDD-RCKF estimator performed slightly worse than the other SDD estimators, although it is structurally equal to the SDD-LMS estimator with an adaptive step-size parameter. In any case, even at the relatively small fading rate $B_dT_s = 0.001$, all these adaptive turbo receivers exhibited a remarkable performance gain, as compared with the turbo receiver including the PSAE demodulator.

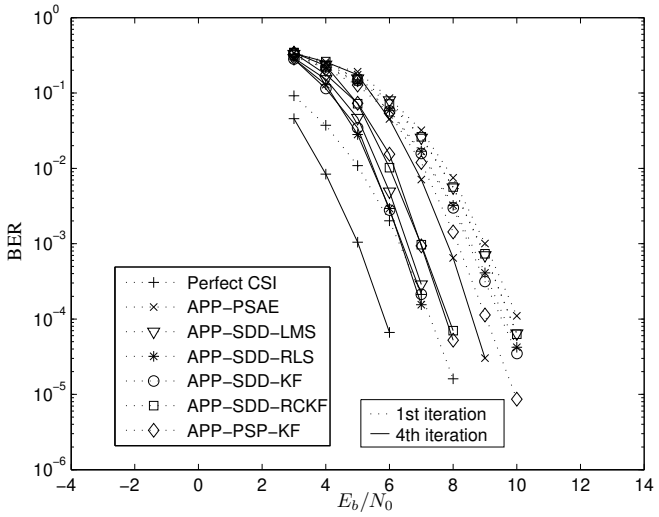

Fig 20. Comparison of various A-SISO algorithms in a coded QPSK system with fading rate $B_dT_s = 0.001$ and $K = 30$.

The performance of the adaptive turbo receivers was also simulated with the following system parameters: $B_dT_s = 0.01$, $K_{\text{pre}} = 10$, $K = 15$, and $K_{\text{tail}} = 2$. The BER as a function of E_b/N_0 for several turbo receivers with different noniterative APP demodulators is presented in Fig. 21. In this relatively fast fading channel, all turbo receivers exhibited, after the first iteration, an error floor at an intolerable high BER value. However, the APP-SDD demodulators were able to efficiently use the *a priori* information about the data symbols obtained from the SISO decoder. This was demonstrated by a remarkable performance gain achieved when the number of iterations was increased. Finally, the performance results are given for a very fast fading ($B_dT_s = 0.05$) frequency selective channel and for short block length ($K = 5$) in Fig. 22. The excellent channel tracking capability of the SDD Kalman filter is particularly exhibited in this fast fading channel condition.

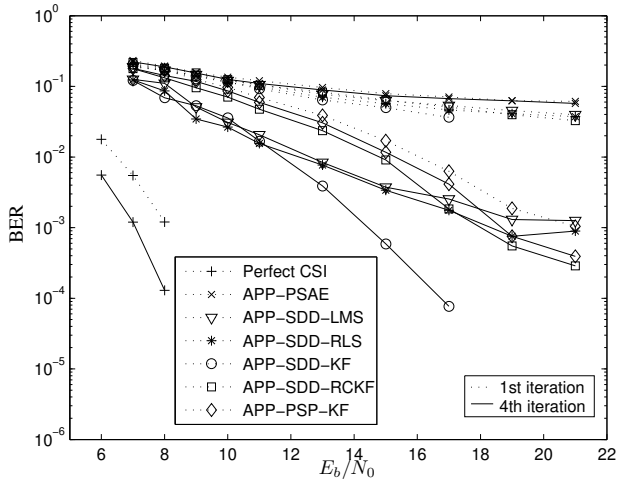

Fig 21. Comparison of various A-SISO algorithms in a coded QPSK system with fading rate $B_dT_s = 0.01$ and $K = 15$.

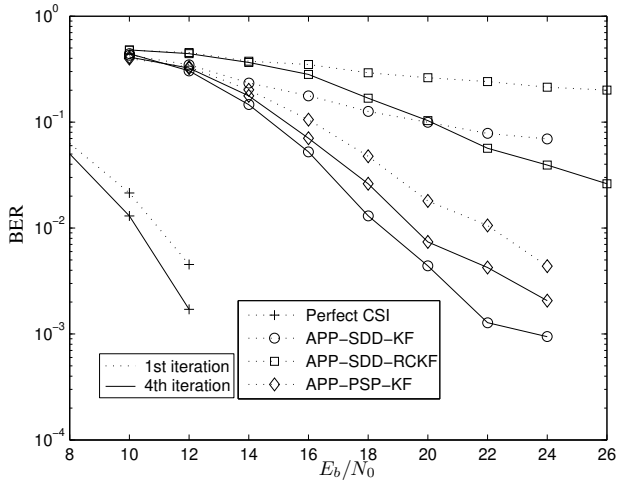

Fig 22. Comparison of various A-SISO algorithms in a coded QPSK system with fading rate $B_dT_s = 0.05$ and $K = 5$.

4.7 Summary and conclusions

In this chapter, two recursive versions of the Baum-Welch algorithm as well as two versions of the BEM-based APP demodulation algorithms were introduced and their performance was simulated in uncoded and coded QPSK transmission systems over a frequency-selective Rayleigh fading channel. Prior to that, different criteria for SISO demodulation in the presence of an unknown channel were shortly reviewed, by emphasizing the VFEM interpretation of the EM- and BEM-based demodulators in particular. Importantly, soft decision-directed LMS, RLS, KF, and RCKF channel estimators were derived in a unified way, and their embedding into the well-known forward-backward processing SISO algorithms was illustrated. Moreover, linear SISO detectors incorporating the SDD channel estimators were derived by appealing to the mean-field approximation.

The simulation results showed that an incorporation of the proposed APP demodulators into the iterative turbo-processing receivers can provide significant performance gain, particularly as compared to the serial concatenation of the PSP-based detector and the hard-decision Viterbi-decoder. As expected, the proposed APP-SDD algorithms achieved remarkable iteration gains compared to the APP algorithms where the channel estimation was solely based on the use of pilot symbols. While desirable, a complete characterization of the whole array of different adaptive iterative detectors in terms of performance and complexity is a complicated problem, and the discussions and simulation results presented in this chapter provide only some partial answers, if any.

5 Joint estimation of carrier frequency offset and statistical parameters of the multipath fading channel

Impressive headway has lately been made in many areas of advanced communication algorithms, including iterative turbo receiver algorithms, space-time signal processing algorithms, and soft decision aided channel estimation algorithms. At the same time, the problem of estimating the channel's statistical properties has been of less interest, in general. The focus of this chapter is a joint estimation of the statistical properties of the radio channel, the channel impulse response, and the carrier frequency offset.

From the transceiver adaptation point of view, important parameters describing the radio channel are the normalized bandwidth of the channel spectrum, referred to as the Doppler spread or fading rate, and the relative power levels of the different multipath components. In addition, the center frequency of the spectrum of the received signal, referred to as the carrier frequency offset (CFO), is an important parameter, which in a broad sense can be understood as a property of the communication channel.

The frequency offset is defined as a total difference between the transmitted carrier frequency and the local reference frequency at the receiver, and it is cumulatively caused by the Doppler shift of the channel, if one exists, and the transmitter and receiver oscillator instabilities. It should be noted that the oscillator instabilities may cause a frequency offset only if they are stable with time, otherwise they cause phase noise. The former case is of interest in this chapter, whereas the latter modeling assumption is considered in the next chapter. In general, the frequency offset and other channel parameters can be different for different multipath signal components due to different scattering environments or due to nonsynchronous transmitters in the systems where multiple transmitter antennas are employed.

In this chapter, the problem of estimating the CFO and the statistical parameters of the multipath fading channel jointly and reliably in various channel conditions is addressed. In particular, based on the complex bandpass channel modeling concept introduced in Chapter 2 and the EM algorithm, provable efficient joint estimation algorithms are derived. Moreover, building upon the previously established variational optimization framework, an approximate algorithm that is especially conducive to learning the

statistical parameters of the wideband radio channel is proposed. Also, new schemes for low-complexity recursive joint synchronization and channel estimation are provided.

Compared to numerous earlier contributions on estimation of the CFO, a fundamental advantage of the frequency offset estimator derived in this chapter is its generality. While the estimation range of the existing estimators is either strictly limited or forced to be traded off for accuracy, the estimator presented here has a large estimation range and still works equally well in a wide variety of scattering environments, encompassing those where the scattering environment is different for different multipath signal components. Also, a novel Kalman smoother-based frequency error detector (FED) which can be used as a building block in feedback frequency recovery schemes and which possesses the same favorable properties as the corresponding block-based estimator will be derived.

The outline of the chapter is as follows. In Section 5.1, a block-based estimation scheme for the CFO and for the statistical parameters of the channel is introduced. In Section 5.2, recursive feedforward and feedback estimation schemes are discussed, and it is shown that they can be easily integrated into various receiver structures. In Section 5.3, the CRLB analysis in the presence of time-varying fading channel is revisited. In Section 5.4, numerical results are presented, and Section 5.5 finally concludes the chapter.

5.1 Estimation of channel statistical parameters

In this section, the problem of estimating the statistical parameters of mobile radio channels, encompassing, thereby, the CFO also, is addressed. Actually, in the presence of CFO, only the complex AR parameters and receiver noise variance must be estimated in order to estimate the (frequecy-modulated) CIR and to perform coherent detection. It is, however, useful and sometimes even necessary to estimate the CFO explicitly. For example, large initial CFO has to be compensated at the receiver front-end so that the received signal is not distorted by the FE filters.

5.1.1 Block-based estimation of state-space model parameters

Let us consider the state-space signal model given by (42) and (43). The channel model given by (42) was earlier called the complex bandpass AR channel model. Such a notion was empowered by the incorporation of the CFO into CIR, effectively shifting the frequency response of the channel in frequency by the amount dictated by the CFO. A set of unknown parameters of the state-space model is given as $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, where $\boldsymbol{\theta}_1 = \{\sigma_v^2\}$ and $\boldsymbol{\theta}_2 = \{\mathbf{G}, \mathbf{A}\}$ include the parameters of the measurement equation and the parameters of the state equation, respectively. As is well-known, a direct data-aided ML estimator for the parameters $\boldsymbol{\Theta}$ is given by

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} p(\mathbf{r}|\mathbf{s}, \boldsymbol{\Theta}) = \arg\max_{\boldsymbol{\Theta}} \int_{\mathbf{F}} p(\mathbf{r}, \mathbf{F}|\mathbf{s}, \boldsymbol{\Theta}) d\mathbf{F}, \tag{153}$$

where $\mathbf{F}$ denotes a set of instantaneous CIR values, i.e., $\mathbf{F} = \left\{\mathbf{f}_1, \ldots, \mathbf{f}_K\right\}$. The conditional distribution $p(\mathbf{r}, \mathbf{F}|\mathbf{s}, \boldsymbol{\Theta})$ can be expressed as

$$\begin{aligned} p(\mathbf{r}, \mathbf{F}|\mathbf{s}, \boldsymbol{\Theta}) &= p(\mathbf{r}|\mathbf{F}, \mathbf{s}, \boldsymbol{\theta}_1) p(\mathbf{F}|\boldsymbol{\theta}_2) \\ &\propto \frac{1}{(\sigma_v^2)^K} \exp\left[-\frac{1}{\sigma_v^2} \sum_{k=1}^{K} \left|r(k) - \boldsymbol{s}^T(k)\boldsymbol{f}(k)\right|^2\right] \\ &\quad \cdot \prod_{l=1}^{L+1} \frac{1}{g_l^K} \exp\left[-\frac{1}{g_l} \sum_{k=1}^{K} \left|f_l(k) - \mathbf{a}_l^T \boldsymbol{f}(k-1)\right|^2\right], \end{aligned} \tag{154}$$

where $\mathbf{a}_l$ denotes the lth column of $\mathbf{A}^T$. Unfortunately, direct ML estimation of the state-space model parameters is not computationally feasible, forcing us to pursue other, computationally more attractive approaches. Thus, the EM approach is here adopted as a basis for deriving an iterative pilot-symbol-aided estimator for the state-space model parameters. In fact, the EM-based estimator for the generic state-space model has earlier been presented in [223], and, hence, the derivation given here may be regarded merely as an application of a similar approach to the specific system model at hand. But, on top of that, it will be shown that in the cases where the channel tap gains are virtually uncorrelated, the complexity of the EM-based parameter estimator can be significantly reduced by making use of the structured MF technique.

By assigning the set of the CIR snapshots $\mathbf{F}$ as a missing data, the objective function

of the EM algorithm at the ith iteration can be written as

$$Q\Big(\boldsymbol{\Theta}\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\Big) = E\Big[\ln p(\mathbf{r}|\mathbf{s},\boldsymbol{\theta}_1,\mathbf{F})\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big] + E\Big[\ln p(\mathbf{F}|\boldsymbol{\theta}_2)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]. \tag{155}$$

As is readily noticed, the maximization of the Q-function (the M-step) over the parameters of the measurement equation is decoupled from the maximization over the parameters of the state equation. The estimators $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are, however, coupled through computation of the conditional expectations. The Q-function (155) can be further expanded as[27]

$$Q\Big(\boldsymbol{\Theta}\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\Big) = \sum_{l=1}^{L+1} Q_l\Big(\boldsymbol{\Theta}\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\Big), \tag{156}$$

where Q_l for $l = 1,\ldots,L$ is given as

$$\begin{aligned} Q_l\Big(\boldsymbol{\Theta}\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\Big) = -K\ln g_l - \frac{1}{g_m}\sum_{k=1}^{K}\Bigg(& E\Big[|f_m(k)|^2\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big] \\ & - 2\Re\Big\{\mathbf{a}_l^T E\Big[\mathbf{f}(k-1)f_l^*(k)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]\Big\} \\ & + \mathbf{a}_l^T E\Big[\mathbf{f}(k-1)\mathbf{f}^H(k-1)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]\mathbf{a}_l^*\Bigg), \end{aligned} \tag{157}$$

and

$$\begin{aligned} Q_{L+1}\Big(\boldsymbol{\Theta}\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\Big) = & -K\ln\sigma_v^2 \\ & - \frac{1}{\sigma_v^2}\sum_{k=1}^{K}\Bigg(|r(k)|^2 - 2\Re\Big\{r^*(k)\mathbf{s}^T(k)E\Big[\mathbf{f}(k)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]\Big\} \\ & + \mathbf{s}^T(k)E\Big[\mathbf{f}(k)\mathbf{f}^H(k)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]\mathbf{s}^*(k)\Bigg). \end{aligned} \tag{158}$$

Above, $\Re\{\cdot\}$ denotes the real part of the argument. At the ith iteration, the EM algorithm can be written as

[27] For notational simplicity, it is assumed hereafter that $p = 1$ if not otherwise noted.

$$\mathtt{E}-\mathtt{step}:\ \text{For } k=0,1,\ldots,K \text{ compute}$$

$$E\Big[\mathbf{f}(k)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]=\hat{\mathbf{f}}^{(i)}(k|K)$$

$$E\Big[\mathbf{f}(k)\mathbf{f}^H(k)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]=\hat{\mathbf{f}}^{(i)}(k|K)\big(\hat{\mathbf{f}}^{(i)}(k|K)\big)^H+\mathbf{P}^{(i)}(k|K)\triangleq\boldsymbol{\Phi}_k^{(i)}$$

$$E\Big[\mathbf{f}(k-1)\mathbf{f}^H(k)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]=\hat{\mathbf{f}}^{(i)}(k-1|K)\big(\hat{\mathbf{f}}^{(i)}(k|K)\big)^H+\mathbf{P}^{(i)}(k-1,k|K)\triangleq\boldsymbol{\Psi}_k^{(i)}$$

$\mathtt{M}-\mathtt{step}:$ Compute

$$\hat{\mathbf{A}}^{(i)}=\sum_{k=1}^{K}\big(\boldsymbol{\Psi}_k^{(i)}\big)^H\Big(\sum_{k=1}^{K}\boldsymbol{\Phi}_{k-1}^{(i)}\Big)^{-1} \tag{159}$$

$$\hat{g}_l^{(i)}=\frac{1}{K}\sum_{k=1}^{K}\Big(\big[\boldsymbol{\Phi}_k^{(i)}\big]_{l,l}-\big(\hat{\mathbf{a}}_l^{(i)}\big)^T\boldsymbol{\psi}_{l,k}^{(i)}\Big) \tag{160}$$

$$\widehat{\sigma_v^2}^{(i)}=\frac{1}{K}\sum_{k=1}^{K}\Big(|r(k)|^2-2\Re\big\{r^*(k)\mathbf{s}^T(k)\hat{\mathbf{f}}^{(i)}(k|K)\big\}+\mathbf{s}^T(k)\boldsymbol{\Phi}_k^{(i)}\mathbf{s}^*(k)\Big), \tag{161}$$

where $\boldsymbol{\psi}_{l,k}^{(i)}$ denotes the lth column of the matrix $\boldsymbol{\Psi}_k^{(i)}$. The smoothed channel estimates $\hat{\mathbf{f}}^{(i)}(k|K)$ and the error covariance matrices $\mathbf{P}^{(i)}(k|N)$ and $\mathbf{P}^{(i)}(k,k-1|N)$ are directly obtained from the Kalman smoothing equations (see, e.g., [223]). It is noticed that (159) essentially gives a *Yule-Walker* solution for the AR parameters where the true autocorrelation matrix of the fading channel is replaced by its current estimate. A modification of the state-space model (42) and (43) in order to support an autoregressive moving average (ARMA) channel hypermodel, as well as an EM-based iterative algorithm for estimating the corresponding model parameters, is presented in Appendix 3.

The above EM-based AR parameter estimator turns out to be computationally unattractive if the number of the channel taps is high. In particular, the numerical stability problems and high computational demand of the Kalman smoother can make the practical implementation prohibitive in this case. A significant reduction in complexity and an improvement in stability can, however, be obtained under the assumption that the off-diagonal elements of the submatrices $\mathbf{A}_i$ are zero, reflecting the situation where

the channel tap gains are uncorrelated. Fortunately, also in practice the correlation between the channel taps can often be assumed to be very low, particularly if the rate of decay of the impulse response of the pulse-shaping filter is large [215].

If the correlation between the channel tap gains can be ignored, the model of the channel dynamics described by the state equation (42) naturally reduces to the $L+1$ parallel dynamic models, each one corresponding to one channel tap. In this case, the distribution $p(\mathbf{F}|\boldsymbol{\theta}_2)$ can be decomposed as

$$p(\mathbf{F}|\boldsymbol{\theta}_2) = \prod_{l=1}^{L+1} p\big(\tilde{\mathbf{F}}_l|\tilde{\mathbf{A}}_l, g_l\big), \tag{162}$$

where $\tilde{\mathbf{F}}_l = \big\{\tilde{\mathbf{f}}_l(1), \ldots, \tilde{\mathbf{f}}_l(K)\big\}$, $\tilde{\mathbf{f}}_l(k) = [f_l(k), \ldots, f_l(k-p+1)]^T$, and the model matrix $\tilde{\mathbf{A}}_l$ contains the AR coefficients of the lth channel tap. In addition, if a direct access to the output of each sub-channel, denoted as $y_l(k) = s(k-l+1)f_{l-1}(k)$, would be available, a parallel filtering structure could be used to estimate the parameters of each sub-channel independently. Since this is not available, the subchannel output signals $y_l(k)$ have to be estimated somehow. The following derivation makes use of the mean field approximation technique.

As pointed out in Chapter 2, the core of the MF approximation technique is to define a factorized trial distribution, called *variational* distribution, which in this case is postulated to be

$$Q(\mathbf{F}) = \prod_{l=1}^{L+1} q_l\big(\tilde{\mathbf{F}}_l\big) \prod_{l=1}^{L+1} p\big(\tilde{\mathbf{F}}_l|\tilde{\mathbf{A}}_l, g_l\big), \tag{163}$$

where $q_l\big(\tilde{\mathbf{F}}_l\big)$ for $l = 1, \ldots, L+1$ form a set of variational free parameters to be optimized. This is achieved by minimizing the KL divergence between the variational distribution and the target distribution $p(\mathbf{r}, \mathbf{F}|\mathbf{s}, \boldsymbol{\Theta})$, given by

$$D\Big(Q(\mathbf{F})\big\|p(\mathbf{r}, \mathbf{F}|\mathbf{s}, \boldsymbol{\Theta})\Big) = \int_{\mathbf{F}} Q(\mathbf{F}) \ln \frac{Q(\mathbf{F})}{p\big(\mathbf{r}, \mathbf{F}|\mathbf{s}, \boldsymbol{\Theta}\big)} d\mathbf{F}, \tag{164}$$

with respect to the free parameters $q_l\big(\tilde{\mathbf{F}}_l\big)$ $\forall l$. Inserting (163) into (164) and solving the zero-gradient points of (164) with respect to the free parameters yields, after straightforward computation, a set of fixed-point equations given by

$$q_l^{\text{MF}}\big(\tilde{\mathbf{F}}_l\big) = C \cdot \exp\left[-\frac{1}{\sigma_v^2} \sum_{k=1}^{K} \big|\hat{y}_l(k) - s(k-l+1)f_{l-1}(k)\big|^2\right] \tag{165}$$

for $l = 1, \ldots, L+1$, where C is a constant, and

$$\hat{y}_l(k) \triangleq r(k) - \sum_{\substack{j=1 \\ j \neq l}}^{L+1} s(k-j+1)\hat{f}_{j-1}(k|K) \tag{166}$$

defines the estimated output signal of the lth sub-channel. Furthermore, $\hat{f}_{j-1}(k|K)$ denotes the MMSE estimate of the jth channel tap gain at time k given that the whole received signal sequence $\mathbf{r}$ is received. Notably, the MF solution (165) and (166) can be interpreted as a parallel interference cancellation scheme for the multipath signal components.

The sequential update for $q_l^{\mathrm{MF}}(\tilde{\mathbf{F}}_l)$ $\forall l$ according to (165) and (166) essentially implies an iterative processing where the "pseudo-observations" $\hat{y}_l(k)$ at each iteration are obtained with the aid of the channel estimates computed at the previous iteration. The MMSE channel estimates $\hat{f}_{l-1}(k|K)$ for all channel taps can be computed in parallel by the per-branch Kalman smoothers operating on the per-branch state-space models. A state-space model for the lth channel tap, when assuming the first order AR(1) channel model (in this case $\tilde{\mathbf{A}}_l$ is a scalar denoted as a_l), is given by

$$f_{l-1}(k) = a_l\, f_{l-1}(k-1) + \sqrt{g_l}\, w_l(k-1) \tag{167}$$

$$\hat{y}_l(k) = s(k-l+1) f_{l-1}(k) + \tilde{n}_l(k). \tag{168}$$

From (165) follows that the variance $\sigma^2_{\tilde{n}_l}$ of the noise $\tilde{n}_l(k)$ should be nominally set at $\sigma^2_{\tilde{n}_l} = \sigma_v^2$ but, in general, it is preferable, from the convergence point of view, to keep it as a design parameter that can change value from iteration to iteration.

A reduced complexity AR (RC-AR) parameter estimator can now be obtained by using the factorized MF distribution $Q^{\mathrm{MF}}(\mathbf{F}) = \prod_l q_l^{\mathrm{MF}}(\tilde{\mathbf{F}}_l)$ and the associated bank of state-space models ((167)-(168)) as a basis for the EM algorithm. In fact, the iterative RC-AR parameter estimator is realized by operating the scalar Kalman smoothers, one corresponding to each channel tap, on the per-branch state-space models and estimating the statistical parameters of each channel tap separately. In the first iteration, when the smoothed channel estimates are not yet available, the pseudo-observations $\hat{y}_l(k)$ can be defined in terms of the KF channel predictions $\hat{f}_l(k|k-1)$. This kind of prediction-feedback mechanism has earlier been applied successfully to noniterative vector Kalman estimation [215]. Moreover, the convergence properties of the MF-EM algorithm can be improved by applying the mean field annealing technique [33], i.e., the variance $\sigma^2_{\tilde{n}_l}$ is initially set at some high value and then gradually decreased along

with the iterations. This corresponds to the hypothetical situation where the receiver is first heated up and then gradually cooled down.

5.1.2 Estimates for frequency offset, Doppler spread, and delay-power profile

The AR parameters of the state equation (42) pertaining to any particular channel tap gain carry information about the carrier frequency offset, Doppler spread, and the averaged power, characterizing this channel tap. For the lth channel tap, they are denoted by ν_l, $\Omega_D^{(l)}$, and ϱ_l^2, respectively. Typically, all channel taps experience the same frequency offset and Doppler spread, but the averaged power level varies from tap gain to tap gain. Under this assumption, $\nu_l \equiv \nu \; \forall l$ and $\Omega_D^{(l)} \equiv \Omega_D \; \forall l$.

Assuming the first order AR modeling, estimates for the frequency offset ν_l, Doppler spread $\Omega_D^{(l)}$, and averaged power level ϱ_l^2 of the lth channel tap gain can be simply obtained by

$$\hat{\nu}_l = \frac{1}{2\pi} \arg(\hat{a}_l) \tag{169}$$

$$\widehat{\Omega_D^{(l)}} = \arccos\left(\frac{1 + |\hat{a}_l|^2 - \frac{1}{\xi}(1 - |\hat{a}_l|)^2}{2|\hat{a}_l|} \right) \tag{170}$$

$$\hat{\varrho}_l^2 = \frac{g_l}{1 - |\hat{a}_l|^2} \, , \tag{171}$$

where the function $\arg(\cdot)$ returns the phase of its argument, $a_l \stackrel{\triangle}{=} [\mathbf{A}]_{l,l}$, and ξ is an attenuation of the AR-spectrum at the estimated frequency with respect to zero frequency (e.g., $\xi = 0.5$ corresponds to the -3dB bandwidth). The frequency offset estimator (169) is based on the fact that the diagonal elements of the model matrix $\bar{\mathbf{A}}$ are positive real numbers, and, therefore, the nonzero phase angle of any diagonal element of the matrix $\mathbf{A}$ must be solely due to the frequency offset. Typically, all the channel tap gains experience the same frequency offset, in which case an improved estimate for the carrier frequency offset can be obtained as

$$\hat{\nu} = \sum_{l=1}^{L+1} \hat{\varrho}_l \, \hat{\nu}_l. \tag{172}$$

The estimator (172) is hereafter called the weighted frequency offset estimator, while an unweighted estimator is obtained simply by $\hat{\nu} = 1/(L+1) \sum_{l=1}^{L+1} \hat{\nu}_l$. The Doppler

rate estimate (170) and the average power estimate (171) follow directly from the AR modeling assumptions (see (42)) and from the standard AR spectral estimation methods [158]. Similarly to (172), improved estimate for the Doppler spread can be obtained, by averaging, in the cases where all channel tap gains experience the same scattering environment.

For the pth order AR model, the estimate for the frequency offset of the lth channel tap, under the restriction that the normalized frequency offset $\nu_l < 1/4$, can be expressed as

$$\hat{\nu}_l = \frac{1}{p}\frac{1}{2\pi}\sum_{i=1}^{p}\frac{1}{i}\arg\Big(\text{sign}\big(\Re\{\hat{a}_{i,l}\}\big)\hat{a}_{i,l}\Big), \tag{173}$$

where $a_{i,l} \triangleq \big[\mathbf{A}_i\big]_{l,l}$ and the function sign$(\cdot)$ returns the sign of its argument. The estimate (173) is based on the finding that the diagonal elements of the sub-matrix $\bar{\mathbf{A}}_i$ are (positive or negative) real numbers, and, therefore, a nonzero angle of the element $a_{i,l}$, when reduced to the interval $[0\ \pi/2]$, must be directly related to the frequency offset, as long as the range of the frequency offset is limited as $\nu_l < 1/4$. Meanwhile, the explicit values for the Doppler spread and for the delay-power profile are somewhat difficult to obtain from the estimated state-space model parameters if higher order AR modeling is used.

5.2 Fully adaptive receiver structures

Fully adaptive receiver structures in the presence of the multipath Rayleigh fading channel and unknown CFO are considered in this section. The notion of full adaptivity should be understood in this context so that, along with the CIR, the channel statistical parameters, specifically the channel AR parameters, are estimated recursively as well. Hence, the bandwidth of the CIR estimator, realized by the Kalman filter, can be effectively adjusted in accordance with the Doppler spread of the channel, and, thereby, an enhanced estimation accuracy may be obtained in different channel conditions. The rotation of the signal constellation due to the CFO can be compensated either implicitly by a feed-forward AR parameter estimator or explicitly by a feedback loop.

5.2.1 Feedforward adaptation

An alternative implementation of the AR parameter estimator is obtained via the steepest descent method. Specifically, at the ith iteration, the parameter estimator is given by

$$\hat{\boldsymbol{\Theta}}^{(i)} = \hat{\boldsymbol{\Theta}}^{(i-1)} + \Lambda \left. \frac{\partial \ln p(\mathbf{r}|\mathbf{s}, \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \right|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(i-1)}}, \tag{174}$$

where Λ is a diagonal matrix with step size parameters as diagonal elements. The score of the likelihood function can be computed with the aid of the Fisher's identity [161] as follows:

$$\frac{\partial}{\partial \boldsymbol{\Theta}} \ln p(\mathbf{r}|\mathbf{s}, \boldsymbol{\Theta}) = \left[\frac{\partial}{\partial \boldsymbol{\Theta}_0} Q\left(\boldsymbol{\Theta}_0 | \boldsymbol{\Theta}\right) \right]_{\boldsymbol{\Theta}_0=\boldsymbol{\Theta}}. \tag{175}$$

A recursive estimator for the parameter vector $\boldsymbol{\Theta}$ is now obtained by applying the *stochastic gradient approximation* technique to (174). Specifically,

$$\begin{aligned} \hat{\mathbf{A}}(k) &= \hat{\mathbf{A}}(k-1) \\ &\quad + \lambda_A \left. \frac{\partial E\left[\ln p\big(\mathbf{f}(k)|\mathbf{f}(k-1), \boldsymbol{\theta}_2\big) \Big| \{r(i)\}_{i=1}^{k}, \{\hat{\boldsymbol{\theta}}_2(i)\}_{i=1}^{k-1} \right]}{\partial \mathbf{A}} \right|_{\boldsymbol{\theta}_2=\hat{\boldsymbol{\theta}}_2(k-1)} \\ &= \hat{\mathbf{A}}(k-1) + \Lambda_{A,k}\Big(\boldsymbol{\Psi}_{k|k} - \hat{\mathbf{A}}(k-1)\boldsymbol{\Phi}_{k-1|k} \Big), \end{aligned} \tag{176}$$

where the diagonal matrix containing time-varying step-size parameters is given as $\Lambda_{A,k} = \mathrm{diag}\big(\lambda_A/\hat{g}_1(k-1), \ldots, \lambda_A/\hat{g}_{L+1}(k-1)\big)$, and, similarly to the formulation in (176), the recursive estimate for g_l is obtained as

$$\begin{aligned} \hat{g}_l(k) = (1-\lambda_g)\hat{g}_l(k-1) + \lambda_g\Big(\big[\boldsymbol{\Phi}_{k|k}\big]_{l,l} - 2\Re\big\{\hat{\mathbf{a}}_l^T(k)\boldsymbol{\psi}_{l,k|k}\big\} \\ + \hat{\mathbf{a}}_l^T(k)\boldsymbol{\Phi}_{k-1|k}\hat{\mathbf{a}}_l^*(k) \Big), \end{aligned} \tag{177}$$

where $\boldsymbol{\psi}_{l,k|k}$ denotes the lth column of $\boldsymbol{\Psi}_{k|k}$.

The error covariance matrices $\boldsymbol{\Phi}_{k|k}$, $\boldsymbol{\Phi}_{k-1|k}$ and $\boldsymbol{\Psi}_{k|k}$ are updated recursively, e.g.,

$$\boldsymbol{\Psi}_{k|k} = E\Big[\mathbf{f}(k-1)\mathbf{f}^H(k) \big| \{r(i)\}_{i=1}^{k}, \{\mathbf{s}(i)\}_{i=1}^{k}, \{\hat{\boldsymbol{\theta}}_2(i)\}_{i=1}^{k-1} \Big] \in \mathbb{C}^{(L+1)\times(L+1)}, \tag{178}$$

and they are computed by the one-step Kalman smoother. Thus, the recursive parameter estimators (176)-(177) are effectively cross-coupled with the recursive one-step

Kalman smoother. Importantly, as opposed to the AR parameter estimator proposed in [60], the error covariances of the CIR estimates, obtained from the Kalman smoother, are also exploited in the AR parameter estimation, in addition to the CIR estimates themselves. Based on the use of this extra information about the channel, the AR parameter estimators proposed here are expected to enjoy some performance advantage over the estimators presented in [60].

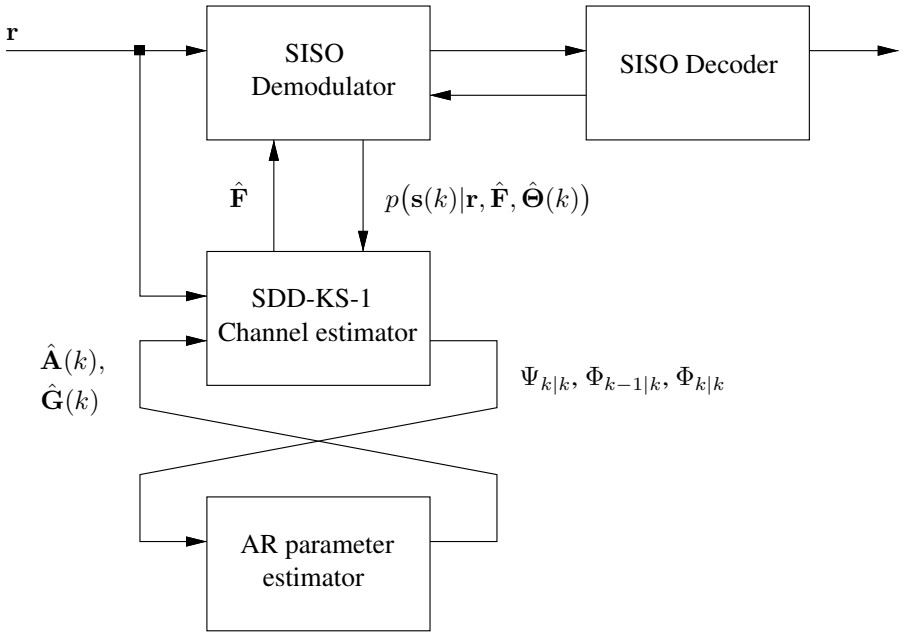

Fig 23. Block diagram of a fully adaptive turbo receiver in the presence of frequency offset.

A block diagram of the fully adaptive turbo receiver, including the cross-coupled SDD channel estimator and AR parameter estimator, is presented in Fig. 23. It is noted that the parameter updating itself does not need information on the data symbols, whereas the Kalman smoother can be operated either in the hard decision-directed (HDD) or the soft decision-directed (SDD) modes, in addition to the pilot mode. The HDD Kalman smoothers can exploit either final hard decisions or tentative hard decisions (e.g., via per-survivor processing) [215]. The SDD Kalman smoothers were derived in Chapter 4.

The complexity of the recursive adaptation in the frequency-selective channel is dominated by the vector Kalman smoothing since the parameter updating itself is computationally very simple. On the other hand, the complexity of the vector Kalman smoothing can be reduced by applying the MF approximation technique as stated in the previous section. Furthermore, the convergence rate of the gradient-based recursive estimator is solely determined by the step-size parameters. This is in contrast to the block-based EM algorithm whose convergence rate depends on the amount of information loss due to the missing information.

5.2.2 Feedback frequency recovery

Often the statistical characteristics of the fading channel change much slower than the CFO with respect to time. In that case, it is reasonable to perform the carrier frequency recovery explicitly while keeping the parameters of the state-space model itself fixed. So, in this section, the problem of CFO estimation and frequency recovery will be tackled in the case where the state-space model parameters are fixed and known. In particular, a novel frequency error detector, to be used as a building block of the feedback frequency recovery scheme, will be derived.

The Q-function for the frequency offset is obtained by inserting (44) into (154) and solving for $Q\big(\nu|\hat{\nu}^{(i-1)}\big)$, using (155) as a basis. Specifically in the case of the first order AR model, the Q-function can be expressed in the form

$$Q\big(\nu|\hat{\nu}^{(i-1)}\big) = 2\Re\Bigg\{e^{j2\pi\nu}\sum_{k=1}^{K}\mathrm{tr}\Big(\big(\mathbf{G}^H\mathbf{G}\big)^{-1}\bar{\mathbf{A}}\boldsymbol{\Psi}_k^{(i)}\Big)\Bigg\}, \tag{179}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of the argument matrix. Maximizing (179) with respect to

ν yields

$$\hat{\nu}^{(i)} = \frac{1}{2\pi} \arg\Bigg(\sum_{k=1}^{K} \mathrm{tr}\Big(\big(\mathbf{G}^H \mathbf{G}\big)^{-1} \bar{\mathbf{A}} \boldsymbol{\Psi}_k^{(i)} \Big) \Bigg). \tag{180}$$

The effects of the CFO can now be compensated explicitly with the aid of the estimate $\hat{\nu}$, for example, at the receiver FE. It is, however, more practical to employ a closed-loop frequency recovery system where the frequency error signal is generated for every received signal sample. This error signal is then filtered by a low-pass filter (LPF), and the filtered error signal is used to control the numerically-controlled oscillator (NCO) or the voltage-controlled oscillator (VCO) if the frequency offset is compensated in the analog domain.

The maximum-likelihood-based frequency error detector (FED) is defined in such a way that the error signal it produces can be used to recursively maximize the likelihood function $\mathcal{L}(\nu) = \ln p(\mathbf{r}|\mathbf{s}, \nu)$. Using Fisher's identity (175), the derivative of the likelihood function $\mathcal{L}(\nu)$ is given by

$$\begin{aligned} \frac{\partial}{\partial \nu} \ln p(\mathbf{r}|\mathbf{s}, \nu) &= -2 \sum_{k=1}^{K} \Im \Big\{ e^{j2\pi\nu} \mathrm{tr}\Big(\big(\mathbf{G}^H \mathbf{G}\big)^{-1} \bar{\mathbf{A}} \boldsymbol{\Psi}_k^{(i)} \Big) \Big\} \\ &= -2 \sum_{k=1}^{K} \Im \Big\{ \mathrm{tr}\Big(\big(\mathbf{G}^H \mathbf{G}\big)^{-1} \bar{\mathbf{A}} \bar{\boldsymbol{\Psi}}_k^{(i)} \Big) \Big\}, \end{aligned} \tag{181}$$

where $\Im\{\cdot\}$ denotes the imaginary part of the argument and the correlation matrix $\bar{\boldsymbol{\Psi}}_k^{(i)}$ is defined as $\bar{\boldsymbol{\Psi}}_k^{(i)} = E\Big[\mathbf{h}(k-1)\mathbf{h}^H(k) \Big| \mathbf{r}, \mathbf{s}, \hat{\nu}^{(i-1)} \Big]$. The second row of (181) is obtained by using (41). The zero-gradient point of the likelihood-function can be found recursively by applying the stochastic gradient approximation technique to (181). Specifically, the error signal at the output of the FED at time k is given by

$$e(k) = \Im \Big\{ \mathrm{tr}\Big(\big(\mathbf{G}^H \mathbf{G}\big)^{-1} \bar{\mathbf{A}} \bar{\boldsymbol{\Psi}}_{k|k} \Big) \Big\}, \tag{182}$$

where the correlation matrix $\bar{\boldsymbol{\Psi}}_{k|k}$ is defined as

$$\bar{\boldsymbol{\Psi}}_{k|k} = E\Big[\mathbf{h}(k-1)\mathbf{h}^H(k) \Big| \{r(i)\}_{i=1}^{k}, \{\mathbf{s}(i)\}_{i=1}^{k}, \{\hat{\nu}(i)\}_{i=1}^{k-1} \Big], \in \mathbb{C}^{(L+1)\times(L+1)} \tag{183}$$

and it can be obtained from the one-step Kalman smoother (KS-1) operating on the state-space model

$$\mathbf{h}(k) = \bar{\mathbf{A}}\mathbf{h}(k-1) + \mathbf{G}\bar{\mathbf{w}}(k-1) \tag{184}$$

$$r(k)\, e^{-j2\pi\hat{\nu}(k)\, k} = \mathbf{s}^T(k)\mathbf{h}(k) + v(k). \tag{185}$$

The frequency offset estimate may be updated by employing the first order loop filter as follows:

$$\hat{\nu}(k+1) = \hat{\nu}(k) + \mu\, e(k), \tag{186}$$

where μ is a step-size parameter. Like previously, the Kalman smoother can be operated either in the HDD mode or in the SDD mode, and the reduced complexity version of KS can be derived by applying the mean field approximation method. As a result, a computationally attractive closed-loop frequency recovery scheme, being able to fully exploit the available statistical information about the fading channel, is obtained.

In Fig. 24, a block diagram of the closed-loop frequency recovery circuit is illustrated in the case where the channel is assumed to be flat-fading and the channel estimation is performed in the HDD mode. As is well known, the acquisition properties of the feedback frequency recovery loop is determined by the so-called *S*-curve of the FED [164]. An example of the *S*-curve of the proposed FED obtained by simulations is given in Fig. 25.

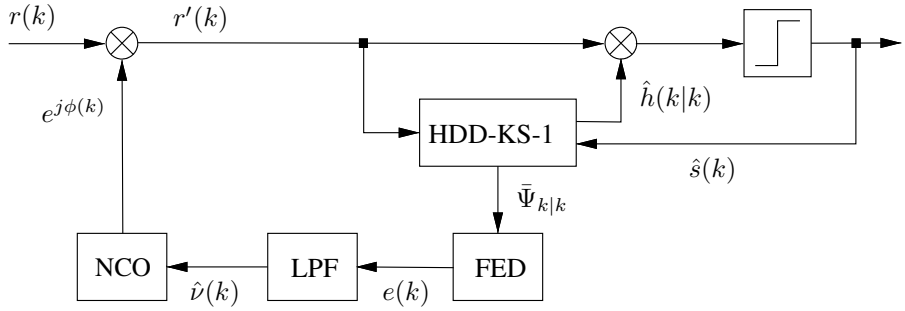

Fig 24. Block diagram of a fully adaptive receiver incorporating a frequency recovery loop. The channel is assumed to be flat-fading, and the Kalman smoother is operated in the hard decision-directed mode.

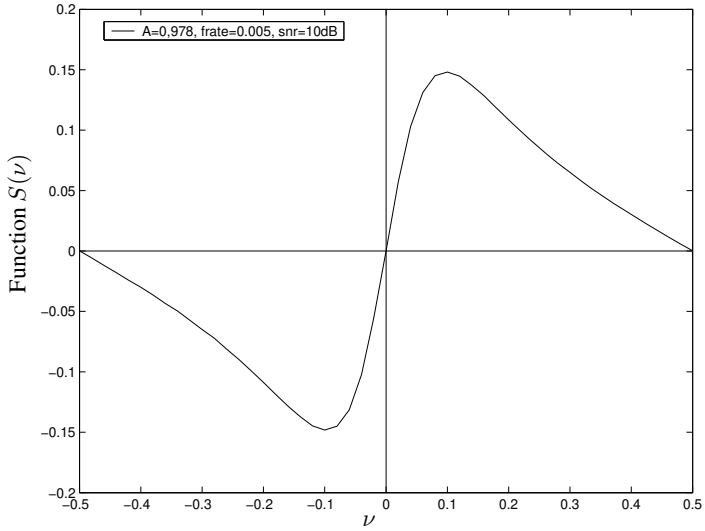

Fig 25. S-curve of the frequency error detector in a flat-fading channel with normalized fading rate $B_dT_s = 0.05$, $E_b/N_0 = 10$ dB, and first-order AR parameter $a = 0.978$.

5.3 CRLB analysis

The Cramer-Rao lower bound (CRLB) is a fundamental lower bound on the variance of any unbiased estimate [244], and it is known to be asymptotically achievable, under mild regularity conditions, for a large enough number of observations [191]. In this section, the problem of determining the CRLB for pilot-symbol-aided estimation of the CFO and the receiver noise variance in the presence of an unknown multipath Rayleigh fading channel is addressed, assuming that the second order statistics of the fading channel are known. A novel formulation of the CRLB adopted here does not only facilitate the involved computations but also allows to gain further understanding of the effects of the fading channel on the accuracy of the estimates. In particular, it will be shown that the impact of the fading phenomenon itself and the impact of incomplete knowledge of the fading process on the CRLB can be quantified separately.

For the sake of easier notation, the system model given by (26) is adopted here. In

the single antenna case considered here it reduces to

$$\mathbf{r} = \mathbf{\Omega X h} + \mathbf{v}, \tag{187}$$

where the matrix $\mathbf{\Omega}$ is a diagonal matrix containing the complex phasors as diagonal elements (see (23)), $\mathbf{X}$ denotes the data matrix (see (32) and (33)) consisting of known pilot symbols, and $\mathbf{h}$ denotes the channel vector given as $\mathbf{h} = \left[\mathbf{h}_1^T, \ldots, \mathbf{h}_{L+1}^T\right]^T$ where the component vector $\mathbf{h}_l$ contains the time-varying values of the lth channel tap gain. The covariance matrix of the fading channel is denoted as $\mathbf{C}_h$ and, under the WSSUS channel assumption, it is a block-diagonal matrix given by

$$\mathbf{C}_h = E\left[\mathbf{h}\mathbf{h}^H\right] = \text{diag}\left(\sigma_{h_1}^2 \mathbf{C}_{h_1}, \ldots, \sigma_{h_{L+1}}^2 \mathbf{C}_{h_{L+1}}\right), \tag{188}$$

where $\sigma_{h_l}^2$ and $\mathbf{C}_{h_l}$ represent the variance and normalized covariance of the lth channel tap gain, respectively. The unknown parameter vector in this case includes only real parameters and is given by $\boldsymbol{\theta} = \left[\nu, \sigma_v^2\right]^T$. The extension of the CRLB analysis theory to include the estimation of complex parameters and constrained complex parameters are considered in [243] and [120], respectively.

The variance of the estimated parameter $\hat{\theta}_i$ is lower bounded by the CRLB as follows:

$$E\left[|\hat{\theta}_i - \theta_i|^2\right] \geq \text{CRLB}(\theta_i) = \left[\boldsymbol{\mathcal{I}}^{-1}\right]_{i,i}, \tag{189}$$

where the (Fisher) information matrix $\boldsymbol{\mathcal{I}}$ is given by

$$\boldsymbol{\mathcal{I}} = -E\left[\frac{\partial^2 \ln p(\mathbf{r}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right]. \tag{190}$$

As shown in [193], the information matrix can be decomposed as follows:

$$\boldsymbol{\mathcal{I}} = -E\left[\frac{\partial^2 Q(\acute{\boldsymbol{\theta}}|\boldsymbol{\theta})}{\partial \acute{\boldsymbol{\theta}} \partial \acute{\boldsymbol{\theta}}^T}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}} - E\left[\frac{\partial^2 Q(\acute{\boldsymbol{\theta}}|\boldsymbol{\theta})}{\partial \acute{\boldsymbol{\theta}} \partial \boldsymbol{\theta}^T}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}} = \boldsymbol{\mathcal{I}}_c + \boldsymbol{\mathcal{I}}_m, \tag{191}$$

where the matrix $\boldsymbol{\mathcal{I}}_c$ is here referred to as the complete-data information matrix, and the matrix $\boldsymbol{\mathcal{I}}_m$ is referred to as the missing information matrix. The Q-function in (191) is the same as the Q-function of the EM algorithm when applied to the system model (187), i.e.,

$$Q(\acute{\boldsymbol{\theta}}|\boldsymbol{\theta}) = E\Big[\ln p(\mathbf{r}|\mathbf{h}, \acute{\boldsymbol{\theta}}) + \ln p(\mathbf{h})\Big|\mathbf{r}, \boldsymbol{\theta}\Big]. \tag{192}$$

Since the distribution $p(\mathbf{r}|\mathbf{h}, \boldsymbol{\theta})$ belongs to the family of regular exponential distributions [161], the order of differentiation and expectation in (191) can be reversed. Hence,

the matrix $\boldsymbol{\mathcal{I}}_c$ can be written as

$$\begin{aligned}\boldsymbol{\mathcal{I}}_c &= -E\left[\frac{\partial^2 Q(\acute{\boldsymbol{\theta}}|\boldsymbol{\theta})}{\partial\acute{\boldsymbol{\theta}}\partial\acute{\boldsymbol{\theta}}^T}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}} \\ &= -E\left[E\left[\frac{\partial^2 \ln p(\mathbf{r}|\mathbf{h},\acute{\boldsymbol{\theta}})}{\partial\acute{\boldsymbol{\theta}}\partial\acute{\boldsymbol{\theta}}^T}\Big|\mathbf{r},\boldsymbol{\theta}\right]\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}} \\ &= -E\left[\frac{\partial^2 \mathcal{L}_c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right],\end{aligned} \tag{193}$$

where $\mathcal{L}_c(\boldsymbol{\theta}) = \ln p(\mathbf{r}|\mathbf{h},\boldsymbol{\theta})$ is often called a complete-data log-likelihood function in the EM literature. Interestingly, the modified Cramer-Rao lower bound (MCRLB) for the parameter θ_i as defined in [56] is given as $\text{MCRLB}(\theta_i) = \left[\boldsymbol{\mathcal{I}}_c^{-1}\right]_{i,i}$. Thus, the matrix $\boldsymbol{\mathcal{I}}_c$ (and, hence, the MCRLB also) accounts for the amount of information available for the estimation of the parameter vector $\boldsymbol{\theta}$ under the assumption that the fading channel is perfectly known. On the contrary, the missing information matrix $\boldsymbol{\mathcal{I}}_m$ quantifies the loss in the amount of information available for the parameter estimation, given that the CIR is not known but is instead estimated by an optimal MMSE estimator. This fact follows directly from the statements provided by the following lemma, where some interesting properties of the missing information matrix $\boldsymbol{\mathcal{I}}_m$ are proved.

Lemma 3 *In the presence of the Rayleigh fading channel, the missing information pertaining to the estimation of the frequency offset ν, given by $\left[\boldsymbol{\mathcal{I}}_m\right]_{1,1}$, depends on the sensitivity of the MMSE estimate of $\mathbf{h}$, given by $\hat{\mathbf{h}} = E[\mathbf{h}|\mathbf{r},\boldsymbol{\theta}]$, to the changes in the value of $\boldsymbol{\theta}$, i.e.,*

$$\left[\boldsymbol{\mathcal{I}}_m\right]_{1,1} = f\left(\frac{\partial}{\partial\boldsymbol{\theta}}E\left[\mathbf{h}|\mathbf{r},\boldsymbol{\theta}\right]\right), \tag{194}$$

where $f(\cdot)$ should be read as "is a function of." Likewise, the missing information pertaining to the estimation of the noise variance σ_v^2, given by $\left[\boldsymbol{\mathcal{I}}_m\right]_{2,2}$, depends on the sensitivity of the MMSE estimate of $\mathbf{h}$ and on the sensitivity of the conditional covariance of $\mathbf{h}$, given by $E\left[\mathbf{h}\mathbf{h}^H|\mathbf{r},\boldsymbol{\theta}\right]$, to the changes in the value of $\boldsymbol{\theta}$, i.e.,

$$\left[\boldsymbol{\mathcal{I}}_m\right]_{2,2} = f\left(\frac{\partial}{\partial\boldsymbol{\theta}}E\left[\mathbf{h}|\mathbf{r},\boldsymbol{\theta}\right],\ \frac{\partial}{\partial\boldsymbol{\theta}}E\left[\mathbf{h}\mathbf{h}^H|\mathbf{r},\boldsymbol{\theta}\right]\right). \tag{195}$$

Proof. The missing informatin matrix $\boldsymbol{\mathcal{I}}_m$ can be expanded as follows:

$$\begin{aligned}
\boldsymbol{\mathcal{I}}_m &= -E\left[\frac{\partial^2 Q(\boldsymbol{\theta}|\acute{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\acute{\boldsymbol{\theta}}^T}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}} \\
&= -E\left[\frac{\partial^2 E\big[\ln p(\mathbf{r}|\mathbf{h},\boldsymbol{\theta}) + \ln p(\mathbf{h})\big|\mathbf{r},\acute{\boldsymbol{\theta}}\big]}{\partial\boldsymbol{\theta}\partial\acute{\boldsymbol{\theta}}^T}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}} \\
&= -E\left[\frac{\partial^2 E\big[-\frac{1}{\sigma_v^2}(\mathbf{r}-\boldsymbol{\Omega}\mathbf{X}\mathbf{h})^H(\mathbf{r}-\boldsymbol{\Omega}\mathbf{X}\mathbf{h})\,|\,\mathbf{r},\acute{\boldsymbol{\theta}}\big]}{\partial\boldsymbol{\theta}\partial\acute{\boldsymbol{\theta}}^T}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}} \\
&= -\frac{1}{\sigma_v^2}E\left[\frac{\partial^2\Big(2\Re\Big\{\mathbf{r}^H\boldsymbol{\Omega}\mathbf{X}\,E\big[\mathbf{h}|\mathbf{r},\acute{\boldsymbol{\theta}}\big]\Big\} - \mathrm{tr}\Big(\mathbf{X}\,E\big[\mathbf{h}\mathbf{h}^H|\mathbf{r},\acute{\boldsymbol{\theta}}\big]\,\mathbf{X}^H\Big)\Big)}{\partial\boldsymbol{\theta}\partial\acute{\boldsymbol{\theta}}^T}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}}.
\end{aligned} \tag{196}$$

From (196), a straightforward computation yields

$$\big[\boldsymbol{\mathcal{I}}_m\big]_{1,1} = -\frac{j2\pi}{\sigma_v^2}E\left[\Re\Big\{\mathbf{r}^H\boldsymbol{\Upsilon}\boldsymbol{\Omega}\mathbf{X}\frac{\partial E\big[\mathbf{h}|\mathbf{r},\acute{\boldsymbol{\theta}}\big]}{\partial\acute{\boldsymbol{\theta}}}\Big\}\right]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}}, \tag{197}$$

where the diagonal matrix $\boldsymbol{\Upsilon}$ is defined as

$$\boldsymbol{\Upsilon} \triangleq \mathrm{diag}\big(1,2,\ldots,k,\ldots,K\big) \in \mathbb{C}^{K\times K}. \tag{198}$$

Likewise, the missing information for the receiver noise can be expressed as

$$\begin{aligned}
\big[\boldsymbol{\mathcal{I}}_m\big]_{2,2} = -\frac{1}{(\sigma_v^2)^2}E\Bigg[&\Re\Big\{\mathbf{r}^H\boldsymbol{\Omega}\mathbf{X}\frac{\partial E\big[\mathbf{h}|\mathbf{r},\acute{\boldsymbol{\theta}}\big]}{\partial\acute{\boldsymbol{\theta}}}\Big\} \\
&- \mathrm{tr}\left(\mathbf{X}\,\frac{\partial E\big[\mathbf{h}\mathbf{h}^H|\mathbf{r},\acute{\boldsymbol{\theta}}\big]}{\partial\acute{\boldsymbol{\theta}}}\,\mathbf{X}^H\right)\Bigg]_{\acute{\boldsymbol{\theta}}=\boldsymbol{\theta}}.
\end{aligned} \tag{199}$$

□

When the channel tap gains are uncorrelated and the data symbols have constant energy, straightforward computation yields

$$\boldsymbol{\mathcal{I}}_c = \begin{bmatrix} 8\pi^2\rho\frac{K(K+1)(2K+1)}{6} & 0 \\ 0 & \frac{K}{(\sigma_v^2)^2} \end{bmatrix}, \tag{200}$$

where $\rho = \sum_i \sigma_{h_i}^2/\sigma_v^2$ denotes the signal-to-noise ratio (SNR) at the receiver. The missing information matrix $\boldsymbol{\mathcal{I}}_m$ also turns out to be a diagonal matrix whose diagonal

elements are obtained from (197) and (199) after some elaboration as follows:[28]

$$\begin{aligned}\left[\boldsymbol{\mathcal{I}}_m\right]_{1,1} = -\frac{8\pi^2}{\sigma_v^2}\mathrm{tr}\Big(&\mathbf{X}\mathbf{C}_h\mathbf{X}^H\left(\mathbf{X}\mathbf{C}_h\mathbf{X}^H+\sigma_v^2\mathbf{I}_K\right)^{-1}\\ &\boldsymbol{\Upsilon}\left(\mathbf{X}\mathbf{C}_h\mathbf{X}^H+\sigma_v^2\mathbf{I}_K\right)\boldsymbol{\Upsilon}\Big)\end{aligned} \tag{201}$$

$$\begin{aligned}\left[\boldsymbol{\mathcal{I}}_m\right]_{2,2} = -\frac{1}{(\sigma_v^2)^2}\mathrm{tr}\Big(&2\mathbf{X}\mathbf{C}_h\mathbf{X}^H\left(\mathbf{X}\mathbf{C}_h\mathbf{X}^H+\sigma_v^2\mathbf{I}_K\right)^{-1}\\ &-\mathbf{X}\mathbf{C}_h\mathbf{X}^H\left(\mathbf{X}\mathbf{C}_h\mathbf{X}^H+\sigma_v^2\mathbf{I}_K\right)^{-2}\mathbf{X}\mathbf{C}_h\mathbf{X}^H\Big).\end{aligned} \tag{202}$$

In the case of a flat fading channel and an equal energy signal constellation, the complete information matrix is given by (200), and the missing information is directly obtained from (201) and (202) by noting that the data matrix $\mathbf{X}$ becomes a diagonal unitary matrix and, therefore, can be cancelled out from the expression of $\boldsymbol{\mathcal{I}}_m$. In fact, the information matrix $\left[\boldsymbol{\mathcal{I}}\right]_{2,2}$, pertaining to the noise variance, can be expressed under the above assumptions as

$$\left[\boldsymbol{\mathcal{I}}\right]_{2,2} = \frac{1}{(\sigma_v^2)^2}\mathrm{tr}\Big(\left(\mathbf{I}_K-\mathbf{M}^{-1}\right)^2\Big), \tag{203}$$

where

$$\mathbf{M} \triangleq \mathbf{I}_K + \sigma_v^2\mathbf{C}_h^{-1}. \tag{204}$$

As expected, this is exactly the same formula for $\left[\boldsymbol{\mathcal{I}}\right]_{2,2}$ as given previously in [93] under the same assumptions. Recently, a related problem of determining the CRLB for CFO in a digital burst-mode transmission system affected by phase noise, where the time-varying phase noise is modeled with a random-walk (Wiener) process, has been treated in [20].

5.4 Numerical results

The performance of the proposed estimators was studied by using Monte Carlo simulations. In the simulations, QPSK-modulated training symbols were transmitted over the fading channel which was modeled by using a tapped-delay-line channel model with Rayleigh fading channel tap gains. In addition, an isotropic scattering environment was assumed, implying that the fading channel tap gains had Clarke's power spectrum [46]. The time-varying channel taps were implemented by using Jakes' method for

[28] Analytical expressions for the conditional mean and covariance of $\mathbf{h}$, required for the evaluation of $\boldsymbol{\mathcal{I}}_m$, are easily obtained for linear Gaussian channels, see, e.g., [132] for details.

generating a realization of the fading process [121]. Both one-tap flat-fading and three-tap frequency-selective channels, where the frequency-selective channels assumed the power-delay profile $\sigma^2_{\mathbf{h}} = [1/6\ 1/3\ 1/6]^T$, were considered in the simulations. The true normalized frequency offset was taken randomly, assuming a uniform distribution, from the interval $[-0.05\ 0.05]$. At the receiver, symbol-rate sampling was used.

In general, estimation of the statistical properties of the communication channel can be quite challenging, unless the channel is really changing quickly with respect to the signaling rate. This is because the Doppler bandwidth of the channel is significantly smaller than the overall signal bandwidth, and so the symbol rate sampling results in serious oversampling of the channel. The oversampling renders the estimation of the channel AR parameters very difficult since typically only a limited amount of training can be used for estimation. Particularly the higher order AR channel modeling seems to have limited applicability in communication systems due to problems in estimating the model parameters. In spite of these limitations, the frequency offset estimation via the first order AR modeling works suprisingly well, as demonstrated by the following simulation exercises where the performance of the proposed CFO estimators is studied in various channel conditions, using fairly short training sequences.

The mean square error (MSE) (MSE $\triangleq E\big[(\hat{\nu}-\nu)^2\big]$ for the frequency offset) was used as a performance criterion in all simulations. In addition, the MCRLB and CRLB curves, acting as effective lower bounds on the error variances of the examined estimators, are given as references in both the flat-fading and frequency-selective fading channel cases. While commonly used, the MSE criterion may not always represent a proper quality measure in communication systems where the primary goal is to transmit the information bits reliably over a communication link. In particular, a large MSE difference may in some cases be invisible in terms of the BER, whereas a small MSE difference may in some other cases considerably affect the BER. Therefore, the characterization of the proposed estimators in terms of their impact on the BER of the associated communication link would deserve a proper treatment, although it is considered to be out of scope of this thesis.

The MSE of the AR(1)-based frequency offset estimator in the presence of flat fading is presented as a function of SNR in Fig. 26. For comparison, the performance of two earlier published block-based frequency offset estimators is also presented. Specifically, the acronym uKF refers to the unweighted Kuo and Fitz frequency offset estimator [138], and the acronym uMMV refers to the unweighted Morelli, Mengali and Vitetta frequency offset estimator [172]. In [138] and [172], the frequency offset estimators are

based on weighted linear regressions for the phases or phase differences of the sample correlation sequences of the received signal. However, according to [30], the weighting gives only a slight performance advantage over the unweighted estimators, and, therefore, only the performance of the unweighted estimators (uKF and uMMV) is given as a benchmark here. The estimators make use of a pilot sequence with $K_{\text{pilot}} = 30$ symbols, and the normalized fading rate was assumed to be $B_d T_s = 0.005$. It can be noticed from the figure that the performance of the AR(1)-based frequency offset estimator is comparable with the performance of the benchmark estimators at high SNR while it gives some performance advantage over the others in the low SNR cases.

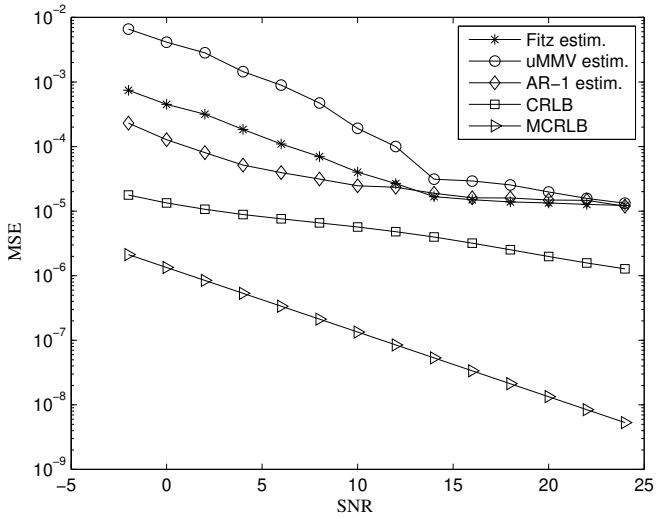

Fig 26. MSE performance of various frequency offset estimators as a function of SNR in a flat-fading channel with the normalized fading rate $B_d T_s = 0.005$.

The MSE performance curves for the weighted AR(1)-based frequency offset estimators in the presence of a three-tap Rayleigh fading channel are presented as a function of the normalized fading rate and as a function of SNR in Fig. 27 and Fig. 28, respectively. The proposed estimators are benchmarked against the *ad hoc* frequency offset estimator proposed by Morelli and Mengali (referred to as the MM-AHE estimator in the figure) in [171]. The frequency offset estimation was based on the use of a pilot

sequence with $K_{\text{pilot}} = 24$. The SNR in the former simulation setup was SNR $= 10$ dB, whereas the normalized fading rate in the latter case was assumed to be $B_d T_s = 0.005$. It may be noticed from the figures that the performance of both the weighted AR(1) and RC-AR(1) (cf. Section 5.1.1) estimators is comparable with the performance of the benchmark estimator, and that the RC-AR-based estimator essentially suffers only a minor performance penalty at high SNR values, as compared to the AR-based estimator. Moreover, in contrast to the MM-AHE estimator that requires a periodic training sequence, the proposed AR-based frequency offset estimators set no restrictions on the training signal structure.

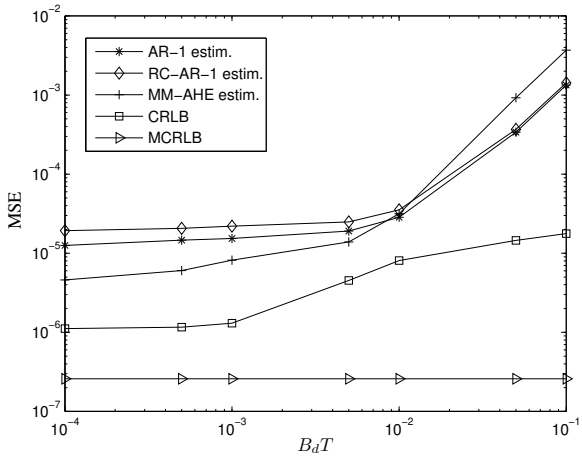

Fig 27. MSE performance of various frequency offset estimators as a function of the normalized fading rate B_dT_s in a three-tap frequency-selective fading channel for SNR $= 10$ dB.

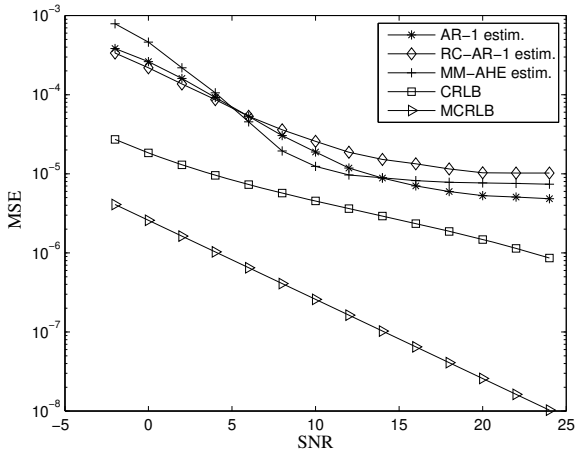

Fig 28. MSE performance of various frequency offset estimators as a function of SNR in a three-tap frequency-selective fading channel. The normalized fading rate was $B_dT_s = 0.005$.

In Fig. 29, the MSE performance of the receiver noise variance estimator in a flat-fading channel is presented as a function of SNR, along with the corresponding MCRLB and CRLB curves. The normalized fading rate was assumed to be $B_d T_s = 0.005$, and the estimator made use of the pilot sequence with $K_{\text{pilot}} = 30$ symbols. As seen from the figure, the performance of the noise variance estimator is close to CRLB at low/medium SNR-values, whereas at high SNR, the MSE of the noise variance estimator seems to saturate.

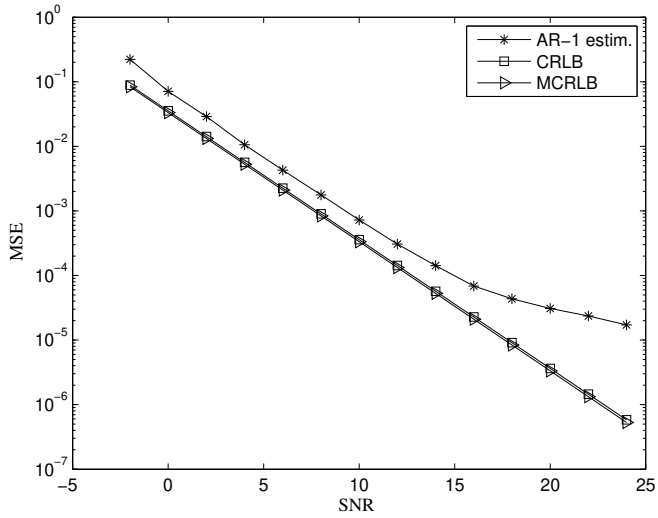

Fig 29. MSE performance of the AR1-based noise variance estimator as a function of SNR in a flat fading channel for a normalized fading rate $B_d T_s = 0.005$.

Furthermore, in Fig. 30, the averaged estimated fading rate value, where the estimator is based on the AR(1) modeling, is presented as a function of the true fading rate value for a flat-fading channel and SNR $= 10$ dB. For comparison, the performance of the fading rate estimator presented in [172] (referred to as the uMMV estimator) is also given. Despite a very long pilot sequence ($K_{\text{pilot}} = 5000$), both the estimators give highly biased estimates in low fading rate values. In these conditions, the problems due to serious oversampling of the Doppler bandwidth appear, indicating that the applicability of the AR modeling and estimation for the purpose of Doppler spread measurement may be quite limited. On the other hand, the overestimation of the fading rate should be acceptable in most systems, whereas the underestimation may be more disruptive. In this respect, the uMMV estimator can be regarded as a better choice of these two alternatives.

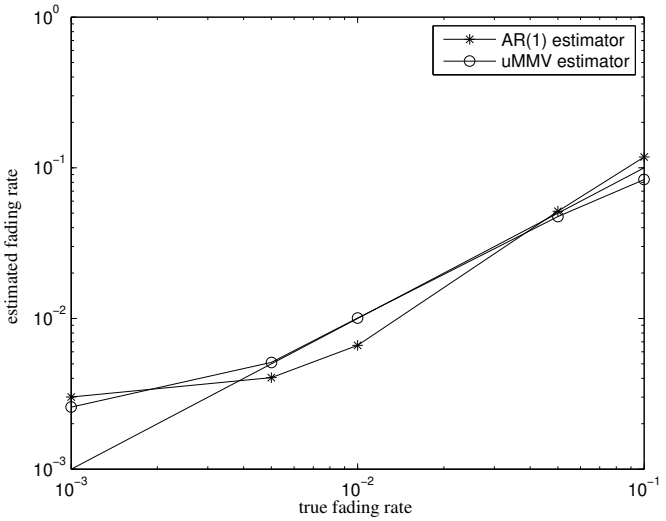

Fig 30. Estimated fading rate as a function of the true fading rate in a flat-fading channel for SNR $= 10$ dB.

The performance of the proposed recursive feed-forward frequency offset estimator is illustrated in Fig. 31 for a flat-fading channel with $B_d T_s = 0.005$. For a reference, the performance of the recursive AR parameter estimator described in [60] is also presented. This reference estimator is based on the cross-coupling of the Kalman-filter-based es-

timator for the CIR and the RLS estimator for the AR parameters (hence referred to as KF-AR estimator in the figure), whereas the recursive estimator proposed in Section 5.2 cross-couples the one-step Kalman-smoother and the AR parameter estimator (hence referred to as KS-AR estimator in the figure) in such a way that the error covariances along with the smoothed channel estimates are effectively exploited in the parameter estimation. As seen from the figure, the performance advantage of the latter approach is remarkable, particularly at the low SNR values.

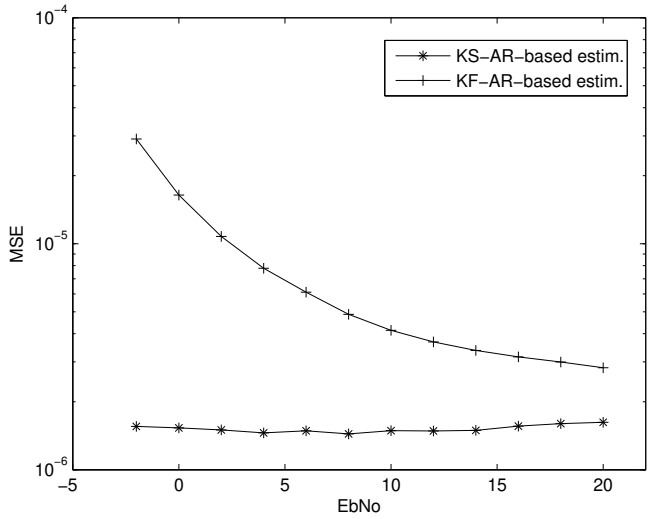

Fig 31. MSE performance of a feed-forward recursive frequency offset estimator in a flat-fading channel with the normalized fading rate $B_dT_s = 0.005$.

Finally, in Fig. 32, the MSE performance of the frequency offset estimators (referred to as KS-AR-FED estimators) that make use of the feedback frequency recovery loop in the presence of a flat-fading (FF) or a frequency-selective (FS) fading channel is presented as a function of SNR, along with the corresponding MCRLB and CRLB curves. The fading rate value was $B_dT_s = 0.005$. While it is noted that the accuracy of the feedback frequency recovery can be partly controlled by the loop filtering, the accuracy enhancement by tightening the loop filter bandwidth will always trade off the tracking capability of the loop. In these simulations, the loop bandwidth is adjusted so

that it effectively corresponds to the block length $K_{\text{equiv.}} = 100$ of an equivalent block estimator (the definition of the equivalent block-length of feedback recovery circuits is given in [164]). As can be noticed from the figure, the MSE of the estimator, as well as the corresponding CRLB, is lower in the FS channel case than in the FF channel case. This is obviously due to the diversity gain provided by multiple independently fading tap gains in the FS channel.

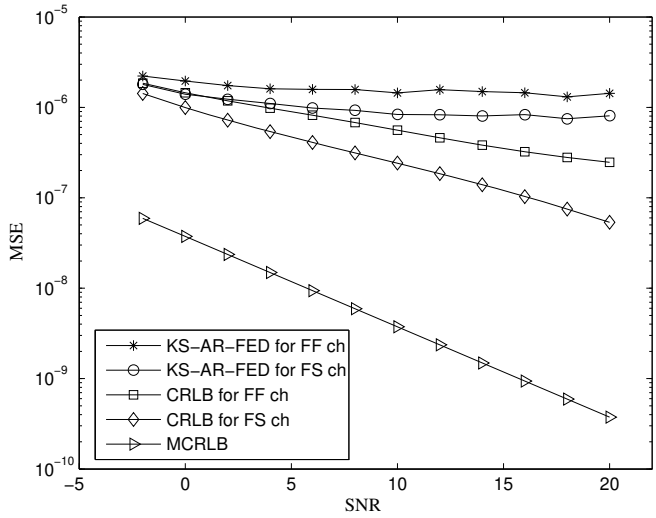

Fig 32. MSE performance of the frequency offset estimators as a function of SNR when a feedback frequency recovery is used. The channel is either flat-fading (FF) or frequency-selective (FS) fading while the normalized fading rate in both cases is $B_d T_s = 0.005$ and the equivalent block length of the estimator is $K_{\text{equiv.}} = 100$.

5.5 Summary and conclusions

The focus in this chapter was on the difficult but important problem of estimating the statistical properties of the multipath communication channel. Based on the complex autoregressive modeling of the received signal in the presence of a multipath Rayleigh fading channel, a block-based joint estimator for the CFO and all relevant statistical parameters of the channel was derived via the EM algorithm. A reduced-complexity

variant of the joint estimator was obtained via the mean field approximation. Moreover, computationally attractive fully adaptive recursive joint synchronization and channel estimation structures, based either on feed-forward adaption or on feedback frequency recovery, were derived as well.

The computer simulations showed that the proposed CFO estimators have good performance in various channel conditions. At the same time, they possess a number of highly desirable properties, such as a high estimation range and high flexibility in terms of channel configuration and the training signal structure. Although a single transmitter antenna system was adopted in this chapter, the derived estimators are also applicable to the multiple transmitter antenna systems, even in cases where the different multiple antenna signals may not be synchronized with each other.

6 Adaptive iterative detection for phase-uncertain channels via variational bounding

So far in this thesis, adaptive iterative detection has been considered mainly for fading multipath channels. In this chapter, the focus is on iterative detection/decoding of data symbols transmitted over the AWGN channel in the presence of phase uncertainty. As summarized in the introduction, a number of receiver algorithms with various amounts of suboptimality have been proposed in the literature to deal with the phase uncertainties. They are based on the modeling of the phase uncertainty either as an unknown deterministic variable/process or random variable/process with a known *a priori* probability density function.

Despite the wealth of literature on the topic, there has been no published treatment of the problem using the VFEM algorithm before. Therefore, based upon the VFEM framework presented in Chapter 2, a set of new adaptive iterative detection algorithms will be obtained in this chapter by applying the variational bounding techniques. In particular, efficient iterative joint estimation and detection/decoding schemes will be derived for a constant phase model as well as for a dynamic phase model by using the variational Bayesian (VB) variant of the VFEM framework.

Optimization via the variational Bayesian formulation has been widely used in the machine-learning community, for example for learning and scoring of graphical model structures [25, 26], but may be less familiar to people working with digital communications. Therefore, the basic principles of the variational bounding specifically tailored for joint iterative phase estimation and data inference are introduced first, followed by detailed derivation of the actual receiver algorithms. Importantly, the close connection of the VB-based receivers to the classical EM-based receiver structures as well as to the optimal noncoherent receivers will be pointed out at various points along the way.

Performance of the proposed VB-based receivers in the presence of strong dynamic phase noise is compared by simulations to the performance of some existing FG-processing-based receivers. Moreover, it is demonstrated by numerical simulations that the convergence of the iterative receivers may be greatly accelerated in many cases by applying the incremental scheduling of the VB or EM algorithm [176], instead of

using their exact implementations.

The rest of this chapter is structured as follows. In Section 6.1, for the sake of clarity, a system model in the presence of either constant phase uncertainty or dynamic phase noise is presented, adhering closely to the generic system model presented in Chapter 2. In addition, modeling issues related to the oscillator phase noise processes will also be discussed. A generic VB framework for iterative estimation and detection is presented in Section 6.2. Then, practical receiver structures based on the generic VB framework are derived. These are obtained for coded data transmission over the constant-phase channels in Section 6.3 and over dynamic-phase AWGN channels in Section 6.4. Section 6.5 is devoted to the discussion on complexities of the proposed detectors versus the ones proposed in the literature. In Section 6.6, numerical results for the performance of the proposed receiver structures are provided. Finally, in Section 6.7, the concluding remarks are presented.

6.1 On system modeling for phase-uncertain channels

Modeling issues relating to the transmission of a sequence of linearly modulated complex data symbols $s(k)$, taking values in the discrete space $\mathcal{S}$, over the AWGN channel affected by *a priori* unknown carrier phase noise is discussed in this section. In general, the phase noise in such a transmission system is caused by oscillators at the transmitter and the receiver, together. However, considering, e.g., a downlink of a wireless system,[29] the oscillator used at the base station is, in practise, sufficiently stable to disregard its phase noise. This is also the case to be considered more closely in the chapter.

Departing momentarily from the generic system model presented in Chapter 2, let us consider a continuous-time transmission system depicted in Fig. 33. Under the above system assumptions, the received continuous-time signal prior to receiver filtering can be expressed as

$$r'(t) = \left(\sum_{k=1}^{K} s(k) p(kT - t) + v'(t) \right) \theta(t), \tag{205}$$

where $p(t)$ denotes the impulse response of a pulse-shaping filter, $v'(t)$ is a zero-mean complex circular white Gaussian noise, and $\theta(t)$ denotes the multiplicative phasor process due to the receiver phase noise. The phasor process can be defined as

$$\theta(t) \triangleq e^{j\phi(t)}, \tag{206}$$

[29] The downlink in this context means the transmission link from a base station to a portable terminal.

where $\phi(t)$ denotes the receiver phase noise, i.e., the instantaneous difference between the phase of the carrier and the phase of the local oscillator at the receiver. It is assumed to be a zero-mean random process with Gaussian statistics. Moreover, the system is assumed to be perfectly frequency-locked, and so there is no deterministic frequency offset at the received signal.

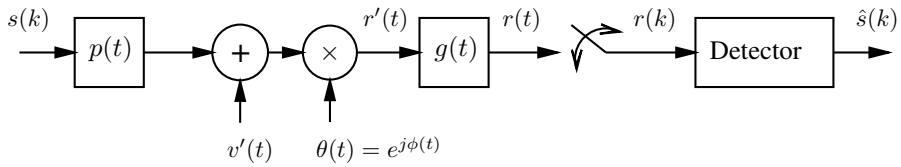

Fig 33. System model for transmission over phase-uncertain channels.

There are basically two types of phase noise processes: finite-power and infinite-power processes. The former type of phase noise is caused by an oscillator controlled by a phase-locked loop (PLL), and it can be approximated as a zero-mean colored Gaussian process that is wide sense stationary (WSS) [200]. Its power spectrum $P_\phi(f)$ depends on the oscillator and on the PLL, but typically it is a low-pass shaped spectrum whose bandwidth is denoted as B_ϕ. Since the PLL tries to keep $\phi(t)$ small, it may be assumed that $|\phi(t)| \ll 1$, and, hence, the following approximation holds: $\theta(t) = e^{j\phi(t)} \approx 1 + j\phi(t)$. In [147], this approximation was used as a basis for deriving variational inference-based receiver algorithms for the OFDM system in the presence of phase noise. A direct consequence of the above approximation is that $P_\theta(f) \approx P_\phi(f)$, where $P_\theta(f)$ denotes the power spectrum of the phasor process $\theta(t)$.

The infinite-power phase noise is a concomitant of a free-running oscillator [200, 275]. In this case, the phase noise $\phi(t)$ is usually modeled as a Wiener-Lévy process (also called continuous-path Brownian motion) that is nonstationary, with its power growing with time. Specifically,

$$\phi(t) = \int_0^t \vartheta(\tau) d\tau, \tag{207}$$

where $\vartheta(t)$ is a zero-mean white Gaussian noise process. Its power is $E\big[\phi^2(t)\big] = 2\pi B_\theta |t|$, where B_θ denotes the two-sided 3-dB bandwidth of the power spectrum of the

phasor process $\theta(t)$ [275]. The power spectrum of $\phi(t)$ is given by

$$P_\phi(f) = \frac{B_\theta}{2\pi f^2}, \tag{208}$$

whereas the spectrum of the phasor process, $P_\theta(f)$, is Lorentzian [200]. In fact, the shape of the spectrum $P_\theta(f)$ matches that of RC-filtered white noise [78], where the acronym RC refers to resistor-capacitor.

At the receiver FE, the signal $r'(t)$ is filtered as

$$r(t) = r'(t) * g(t), \tag{209}$$

where $*$ denotes the convolution operation and $g(t)$ denotes the impulse response of the combined pulse-shaping filter and noise whitening filter, referred to later as the receiver FE filter (sometimes also called the whitened matched filter). Assuming symbol-rate sampling with proper timing and phase variations slow enough for no intersymbol interference to be caused, the sampled output of the receiver FE filter at time k can be expressed as

$$r(k) = s(k)e^{j\phi(k)} + v(k). \tag{210}$$

The sequence $\mathbf{s}$ of K data symbols may be a codeword of some channel code $\mathcal{C}$, constructed by using an arbitrary coding rule.

In this chapter, two specific discrete-time random phase models are considered. In the first one, the phase is assumed to be constant during the transmission of a data block, i.e., $\phi(k) = \phi\ \forall k$, but changes randomly from block to block. The phase samples at different data blocks are modeled as i.i.d. random variables uniformly distributed in the interval $[0,\ 2\pi)$. In the second one, a discrete-time equivalent of Wiener-Lévy phase noise, called the discrete-time random-walk process, is assumed. Specifically, the phase noise $\phi(k)$ is expressed as

$$\phi(k) = \phi(k-1) + \Delta(k), \tag{211}$$

where $\Delta(k)$ is a sample of the real-valued zero mean white Gaussian process with variance σ_Δ^2, implying that $\Delta(k)$ for $k = 1, 2, \ldots, K$ forms a sequence of independently and identically distributed random noise samples. While the values for the noise variances σ_v^2 and σ_Δ^2 are here assumed to be known *a priori*, they can also be estimated efficiently using the methods discussed in the previous chapter. The phase noise model (211) may be considered as a worst case approximation of a real phase noise at the

receiver, since the phasor process $\theta(t)$ was assumed to pass unchanged through the receiver FE filter. On the other hand, if $B_\theta \ll B_s$, where B_s denotes the bandwidth of the transmitted signal, the error due to omitting the effects of receiver FE filtering on the phase noise modeling is probably very small.

6.2 Framework for joint inference and estimation via variational Bayesian algorithm

An optimal MAP symbol detection receiver computes

$$\hat{s}(k)^{\text{MAP}} = \arg\max_{s(k)} p(s(k)|\mathbf{r}) = \arg\max_{s(k)} \sum_{\substack{\mathbf{s}\setminus s(k) \\ \mathbf{s}\in\mathcal{C}}} \int_{\boldsymbol{\phi}} p(\mathbf{s}|\mathbf{r},\boldsymbol{\phi}) p(\boldsymbol{\phi}) d\boldsymbol{\phi}, \tag{212}$$

where $\boldsymbol{\phi} = \{\phi(1), \ldots, \phi(K)\}$ denotes the sequence of phase samples. This receiver is hereafter referred to as the optimal noncoherent receiver due to the fact that the phase uncertainties are averaged out from the decision statistics. In general, the noncoherent receiver is not immune to the exponential time complexity that is known to plague most of the exact inference algorithms in the presence of either channel or parametric uncertainty induced memory. However, for a class of block-constant phase models, symbol-by-symbol soft-decisions under the min-sum criterion can be exactly evaluated with only polynomial complexity with respect to the sequence length [174, 114]. Nevertheless, optimal receivers, even with polynomial complexity in K, are rather prohibitive for practical receiver design. In fact, they are computationally unattractive even in the coherent system if powerful turbo-like codes are involved. Therefore, practical joint data detection and phase estimation algorithms will next be pursued by appealing to the variational bounding technique.

The VFEM algorithm which is specifically tailored for joint inference and estimation in the presence of random parameters is generally referred to as the variational Bayesian (VB) algorithm. In the following, the generic VB algorithm is first introduced by emphasizing its character as being a lower bound for the log-likelihood of $\mathbf{r}$. It is then later used as a basis for detailed derivations of adaptive iterative receiver structures for the two different phase models specified in the previous section.

An essential feature of (212) is a marginalization of the objective function $p(\mathbf{s}, \boldsymbol{\phi}|\mathbf{r})$. Let us now look at a related computational problem—a problem of computing the log-likelihood of $\mathbf{r}$, defined as $\ln p(\mathbf{r})$. Appealing to Jensen's inequality [52], the log-

likelihood of $\mathbf{r}$ can be lower bounded as follows:

$$\begin{aligned}\ln p(\mathbf{r}) &= \ln \sum_{\mathbf{s}} \int_{\boldsymbol{\phi}} p(\mathbf{r}, \mathbf{s}, \boldsymbol{\phi}) d\boldsymbol{\phi} \\ &= \ln \sum_{\mathbf{s}} \int_{\boldsymbol{\phi}} Q(\mathbf{s}, \boldsymbol{\phi}) \frac{p(\mathbf{r}, \mathbf{s}, \boldsymbol{\phi})}{Q(\mathbf{s}, \boldsymbol{\phi})} d\boldsymbol{\phi} \\ &\geq \sum_{\mathbf{s}} \int_{\boldsymbol{\phi}} Q(\mathbf{s}, \boldsymbol{\phi}) \ln \frac{p(\mathbf{r}, \mathbf{s}, \boldsymbol{\phi})}{Q(\mathbf{s}, \boldsymbol{\phi})} d\boldsymbol{\phi}.\end{aligned} \tag{213}$$

The lower bound is achieved when the trial distribution $Q(\mathbf{s}, \boldsymbol{\phi})$ is defined as $Q(\mathbf{s}, \boldsymbol{\phi}) = p(\mathbf{s}, \boldsymbol{\phi}|\mathbf{r})$. However, formulating the lower bound in terms of an unrestricted set of variational distributions provides little guidance in the design of computationally efficient receiver architectures. Therefore, the VB algorithm considered here searches for the maximum of the lower bound over the family of factorized (mean field) distributions: $Q(\mathbf{s}, \boldsymbol{\phi}) = q_{\mathbf{s}}(\mathbf{s}) q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$. With this restriction,

$$\ln p(\mathbf{r}) \geq \sum_{\mathbf{s}} \int_{\boldsymbol{\phi}} q_{\mathbf{s}}(\mathbf{s}) q_{\boldsymbol{\phi}}(\boldsymbol{\phi}) \ln \frac{p(\mathbf{r}, \mathbf{s}, \boldsymbol{\phi})}{q_{\mathbf{s}}(\mathbf{s}) q_{\boldsymbol{\phi}}(\boldsymbol{\phi})} d\boldsymbol{\phi} = \mathcal{G}\big(q_{\mathbf{s}}(\mathbf{s}), q_{\boldsymbol{\phi}}(\boldsymbol{\phi}), \mathbf{r}\big), \tag{214}$$

where the difference between the log-likelihood $\ln p(\mathbf{r})$ and its lower bound can be expressed in terms of the KL divergence as

$$\ln p(\mathbf{r}) - \mathcal{G}\big(q_{\mathbf{s}}(\mathbf{s}), q_{\boldsymbol{\phi}}(\boldsymbol{\phi}), \mathbf{r}\big) = D\big(q_{\mathbf{s}}(\mathbf{s}) q_{\boldsymbol{\phi}}(\boldsymbol{\phi}) \| p(\mathbf{s}, \boldsymbol{\phi}|\mathbf{r})\big). \tag{215}$$

Thus, finding the maximum of the inverse free energy functional $\mathcal{G}$ amounts to minimizing the KL divergence between the factorized trial distribution $q_{\mathbf{s}}(\mathbf{s}) q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ and the original objective function $p(\mathbf{s}, \boldsymbol{\phi}|\mathbf{r})$. Since the KL divergence is often used as a distance measure between distributions, this result may be interpreted so that the VB algorithm provides the best approximation to the optimal noncoherent receiver amongst the solutions that are based on the factorized posteriors. The rationale behind using the VB approach is that the marginalization of $q_{\mathbf{s}}(\mathbf{s})$, a prerequisite for computing approximate symbol APPs, is computationally much simpler than the marginalization in (212).

Starting from some initial values, the VB algorithm proceeds by iteratively maximizing the functional $\mathcal{G}$ alternatively over each of the free distributions $q_{\mathbf{s}}(\mathbf{s})$ and $q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ while holding the other fixed, which is essentially a coordinate ascent in the function space of the factorized distributions. Taking functional derivatives of $\mathcal{G}$ with respect to $q_{\mathbf{s}}(\mathbf{s})$ and $q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$, a straightforward computation yields the following update equations

for the free distributions at the ith iteration:

$$\texttt{VBE}-\texttt{step}:\ q_{\boldsymbol{\phi}}^{(i)}(\boldsymbol{\phi}) \propto p(\boldsymbol{\phi}) \cdot e^{\sum_{\mathbf{s}} q_{\mathbf{s}}^{(i-1)}(\mathbf{s}) \ln p(\mathbf{r}|\mathbf{s},\boldsymbol{\phi})} \tag{216}$$

$$\texttt{VBM}-\texttt{step}:\ q_{\mathbf{s}}^{(i)}(\mathbf{s}) \propto p(\mathbf{s}) \cdot e^{\int_{\boldsymbol{\phi}} q_{\boldsymbol{\phi}}^{(i)}(\boldsymbol{\phi}) \ln p(\mathbf{r}|\mathbf{s},\boldsymbol{\phi}) d\boldsymbol{\phi}}, \tag{217}$$

where $p(\boldsymbol{\phi})$ and $p(\mathbf{s})$ denote *a priori* distributions of the sequence of phase noise samples and the sequence of data symbols, respectively. As reflected by the naming of the update steps above, a close connection of the VB algorithm to the standard EM algorithm can be easily found. In fact, the exponent in (216) can be recognized as an objective function of the EM algorithm which is specified for estimating $\boldsymbol{\phi}$ in the presence of unknown data. On the other hand, the exponent in (217) can similarly be recognized as an objective function of the EM algorithm which is at this time specified for estimating (or rather detecting) data symbols in the presence of random phase disturbances. Due to this close connection, the VB algorithm is sometimes called the *generalized* EM algorithm or the variational Bayesian EM (VBEM) algorithm in the literature [25, 26].

Based upon the factorized Bayesian variational approximation, a general purpose variational message passing (VMP) algorithm, applicable to the Bayesian networks containing conjugate-exponential distributions, were proposed in [260]. Like BP, VMP proceeds by sending messages between nodes in the graph and updating posterior beliefs using local computations at each node. On the basis of (214), each such update increases the lower bound on the log-likelihood. Next, however, the generic VB framework described above is used in deriving practical receiver structures for two phase models specified in Section 6.1.

6.3 Adaptive turbo receivers for constant-phase channels

In this section, the emphasis is put on the detailed derivation of the VB-based iterative receiver structures for the constant-phase channels, where the value of the phase may change from block to block. A close connection of the obtained receiver to the classical EM-based receiver is explicitly pointed out. Also, some words are reserved for discussing the computation of the CRLB in the presence of unknown data symbols—a problem which may be formidable to solve analytically. This is particularly so if channel coding is involved.

6.3.1 VB-based turbo receivers

Consider the constant phase model, where ϕ is a random variable with uniform distribution in $[0,\ 2\pi)$. Inserting this *a priori* information about ϕ into (216), the update equation for the phase posterior distribution (VBE-step) at the ith iteration can be expressed as

$$\begin{aligned} q_{\phi}^{(i)}(\phi) &= C \cdot p(\phi) \cdot e^{-\frac{1}{\sigma_v^2}\sum_{\mathbf{s}} q_{\mathbf{s}}^{(i-1)}(\mathbf{s})\|\mathbf{r}-\mathbf{s}e^{j\phi}\|^2} \\ &= C \cdot p(\phi) \cdot e^{-\frac{1}{\sigma_v^2}\left(\|\mathbf{r}-\bar{\mathbf{s}}^{(i-1)}e^{j\phi}\|^2+\sum_{\mathbf{s}} q_{\mathbf{s}}^{(i-1)}(\mathbf{s})\|\mathbf{s}-\bar{\mathbf{s}}^{(i-1)}\|^2\right)} \\ &= C \cdot p(\phi) \cdot e^{-\frac{1}{\sigma_v^2}\|\mathbf{r}-\bar{\mathbf{s}}^{(i-1)}e^{j\phi}\|^2} \\ &= C \cdot p\big(\mathbf{r}, \phi|\bar{\mathbf{s}}^{(i-1)}\big), \end{aligned} \tag{218}$$

where the soft data decisions at the $(i-1)$th iteration are defined as

$$\bar{s}(k)^{(i-1)} = \sum_{s(k)\in\mathcal{S}} s(k)\, q_{s(k)}^{(i-1)}\big(s(k)\big), \tag{219}$$

and the marginal distribution $q_{s(k)}^{(i-1)}\big(s(k)\big)$ is computed as follows:

$$q_{s(k)}^{(i-1)}\big(s(k)\big) = \sum_{\substack{\mathbf{s}\setminus s(k)\\ \mathbf{s}\in\mathcal{C}}} q_{\mathbf{s}}^{(i-1)}(\mathbf{s}). \tag{220}$$

It should be noted that the constant C, incorporating factors independent of ϕ, may take a different value in each line of (218). In particular, defining the constant C so that the distribution $q_{\phi}^{(i)}(\phi)$ integrates to one yields

$$q_{\phi}^{(i)}(\phi) = p\big(\phi|\mathbf{r}, \bar{\mathbf{s}}^{(i-1)}\big). \tag{221}$$

Considering now the VBM-step, the following lemma provides an intermediate result needed for obtaining the data posterior distribution $q_{\mathbf{s}}(\mathbf{s})$. Specifically, the outcome of the integration in (217) for the constant phase model is given as follows.

Lemma 4

$$\int_{\phi} q_{\phi}^{(i)}(\phi)\ln p(\mathbf{r}|\mathbf{s},\phi)d\phi = -\frac{1}{\sigma_v^2}\Big(\mathbf{r}^H\mathbf{r} - 2\,\zeta^{(i)}\cdot\Re\big\{\mathbf{r}^H\mathbf{s}\,e^{-j\arg(\mathbf{r}^H\bar{\mathbf{s}}^{(i-1)})}\big\} + \mathbf{s}^H\mathbf{s}\Big), \tag{222}$$

where $\arg(\cdot)$ *denotes the phase of the (complex) argument, and*

$$\zeta^{(i)} = \frac{I_1\left(\frac{2}{\sigma_v^2}\left|\mathbf{r}^H\bar{\mathbf{s}}^{(i-1)}\right|^2\right)}{I_0\left(\frac{2}{\sigma_v^2}\left|\mathbf{r}^H\bar{\mathbf{s}}^{(i-1)}\right|^2\right)}, \tag{223}$$

where $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind, given by

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{\pm x \cos(\phi)} d\phi, \tag{224}$$

and $I_1(\cdot)$ is the first-order modified Bessel function of the first kind, expressed as

$$I_1(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos(\phi)} \cos(\phi)\, d\phi. \tag{225}$$

Proof. See Appendix 4. □

Using Lemma 4, the symbol sequence probability at ith iteration can now be expressed in the product form as

$$q_{\mathbf{s}}^{(i)}(\mathbf{s}) = \prod_{k=1}^{K} p\big(s(k)\big) \cdot c_k \cdot e^{-\frac{\zeta^{(i)}}{\sigma_v^2}\left(\left|r(k)-s(k)e^{-j\hat{\phi}^{(i-1)}}\right|^2+\left(\frac{1}{\zeta^{(i)}}-1\right)\left|s(k)\right|^2\right)}, \tag{226}$$

where $\hat{\phi}^{(i-1)} \triangleq \arg\big(\mathbf{r}^H \bar{\mathbf{s}}^{(i-1)}\big)$. It has additionally been assumed that the *a priori* symbol sequence probability fully factorizes.[30] Moreover, c_k denotes the normalizing constant at time k. Based on the above formulation, the VB algorithm at the ith iteration can simply be expressed by the equations (223) and (226). A block diagram of the turbo receiver consisting of a VB-based SISO detector and a SISO decoder, which are iteratively interconnected, is shown in Fig. 34. The fact that the APP detector and the SISO decoder exchange extrinsic information with each other is depicted explicitly in the figure.

[30]This is a reasonable assumption in turbo processing where pseudorandom interleavers are typically employed.

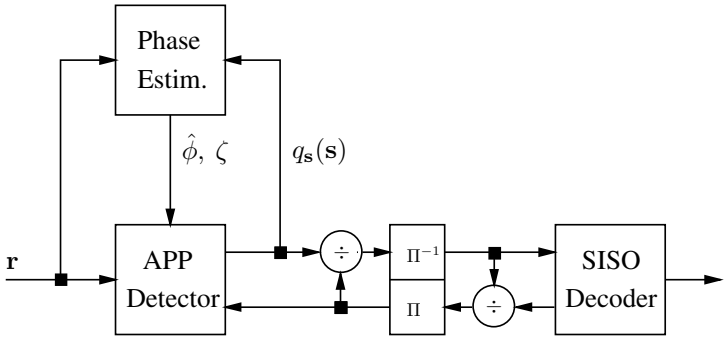

Fig 34. Block diagram of a VB-based iterative turbo receiver in the presence of block-constant phase offset.

Interestingly, by approximating $\zeta^{(i)} \approx 1 \quad \forall i$, it can be easily verified that the VB algorithm reduces to the non-Bayesian EM algorithm (see, e.g., [188] for iterative detection via the classical EM algorithm). It is also easy to see that the EM algorithm as such could also be obtained as a result of variational Bayesian optimization if the free distribution $q_\phi(\phi)$ is restricted to be a point density, i.e., $q_\phi(\phi) = \delta(\phi - \hat{\phi})$, where $\delta(\cdot)$ denotes a Dirac delta function and $\hat{\phi}$ is a free parameter to be optimized.

Since $0 < \zeta^{(i)} < 1$, the VB solution can be interpreted so that, as compared to the EM algorithm, it effectively increases the noise power by the factor $1/\zeta^{(i)}$ at the ith iteration in order to take into account the estimation error in the phase estimation. In the case of M-QAM signaling, it additionally biases the detection slightly towards those constellation points which have small energy. In other words, regarding the computation of the soft data decisions, the constellation points which have large energy are penalized in favor of those constellation points which have small energy. While this can be regarded as a nonobvious and interesting feature coming out of the VB framework, it should at the same time be noted that, in practical systems with moderate to large data block sizes ($K >> 1$), the phase posterior (221) is typically sharply peaked around the most probable (mean) value and, hence, the VB and EM algorithms in these cases perform equally also for M-QAM constellations. This is especially true in the context of iterative turbo receivers.

Since the free distribution $q_{\mathbf{s}}(\mathbf{s})$ according to (226) factorizes into symbol APPs, the

inverse free energy functional $\mathcal{G}\big(q_{\mathbf{s}}(\mathbf{s}), q_\phi(\phi), \mathbf{r}\big)$ can alternatively be written as

$$\mathcal{G}\big(q_{\mathbf{s}}(\mathbf{s}), q_\phi(\phi), \mathbf{r}\big) = \sum_{k=1}^{K} \mathcal{G}_k\big(q_{s(k)}, q_\phi, r(k)\big) + \int_\phi q_\phi(\phi) \ln \frac{p(\phi)}{q_\phi(\phi)}, \tag{227}$$

where

$$\mathcal{G}_k\big(q_{s(k)}, q_\phi, r(k)\big) = \sum_{s(k)\in\mathcal{S}} \int_\phi q_{s(k)}\big(s(k)\big) q_\phi(\phi) \cdot \ln \frac{p\big(r(k), s(k)|\phi\big)}{q_{s(k)}\big(s(k)\big)} d\phi. \tag{228}$$

In practical implementations, different scheduling algorithms can be used to update the free distributions in (227), each defining a specific coordinate ascent on the functional $\mathcal{G}$. For example, the ith iteration of the VB algorithm could be defined as

$$\begin{aligned} q_\phi^{(i)}(\phi) &= \arg\max_{q_\phi} \mathcal{G}\big(q_{\mathbf{s}}(\mathbf{s}), q_\theta(\phi), \mathbf{r}\big), \text{ s.t. } q_{\mathbf{s}}(\mathbf{s}) = q_{\mathbf{s}}^{(i-1)}(\mathbf{s}) \\ q_{s(l)}^{(i)}\big(s(l)\big) &= q_{s(l)}^{(i-1)}\big(s(l)\big) \text{ for } l \neq k \\ q_{s(k)}^{(i)}\big(s(k)\big) &= \arg\max_{q_{s(k)}} \mathcal{G}_k\big(q_{s(k)}, q_\phi, r(k)\big), \text{ s.t. } q_\phi(\phi) = q_\phi^{(i)}(\phi). \end{aligned} \tag{229}$$

In this scheduling, the phase posterior at the VBE-step is updated using the data posteriors from the previous iteration and, at the VBM-step, only one of the data posteriors is updated while retaining the previously computed values for the rest. Thus, the phase posterior $q_\phi(\phi)$ is updated for each (incremental) symbol APP update. From now on, all the scheduling algorithms, where only a part of the symbol APPs are updated at any particular iteration, are collectively referred to as incremental VB algorithms. Specifically, by defining $k = i$ in (229) and initializing $q_{\mathbf{s}}^{(0)}(\mathbf{s}) = \mathbf{0}$, the symbol APPs can be computed by the following *recursive* VB algorithm:

`For` $k = 1, \ldots, K$

$$\hat{\phi}(k) = -\arg\left(\sum_{l=1}^{k-1} r^*(l)\bar{s}(l)\right)$$

$$\zeta^{(k)} = \frac{I_1\left(\frac{2}{\sigma_v^2}\left|\sum_{l=1}^{k-1} r^*(l)\bar{s}(l)\right|^2\right)}{I_0\left(\frac{2}{\sigma_v^2}\left|\sum_{l=1}^{k-1} r^*(l)\bar{s}(l)\right|^2\right)}$$

$$q_{s(k)}\big(s(k)\big) = c_k \cdot p\big(s(k)\big) \cdot e^{-\frac{\zeta^{(k)}}{\sigma_v^2}\left(\left|r(k)-s(k)e^{j\hat{\phi}(k)}\right|^2 + \left(\frac{1}{\zeta^{(k)}}-1\right)\left|s(k)\right|^2\right)}$$

`end`.

Again, assuming that $\zeta^{(k)} \approx 1\ \forall k$, the incremental VB algorithm reduces to the incremental EM algorithm [176] for which interesting convergence results are available:

the algorithm converges, under mild assumptions on the model, to stationary points in likelihood, but the favorable property of monotonically increasing the likelihood at each iteration is lost [96]. However, since the soft data decisions are used immediately in phase estimation, instead of being held until all soft decisions are found, the incremental VB/EM algorithm can be expected to converge much faster than the standard EM algorithm. As will be later demonstrated by numerical examples, the convergence rate of the iterative turbo receiver may indeed be significantly improved by replacing the exact VB/EM algorithm with the incremental VB/EM algorithm at the first iteration of the turbo processing. In practice, this means that the number of decoding iterations can be reduced in order to reach the target performance, potentially entailing a considerable decrease in overall receiver complexity.

6.3.2 On CRLB in the presence of unknown data symbols

The CRLB, being a lower bound on the error variance of any unbiased estimate, serves as a useful benchmark for practical phase estimators. For an unknown carrier phase, it is computed as

$$\mathrm{CRLB}_{\phi} = E_{p(\mathbf{r}|\phi)}\left[\left(\frac{\partial}{\partial\phi}\ln p(\mathbf{r}|\phi)\right)^2\right]^{-1} = \mathcal{I}(\phi)^{-1}, \tag{230}$$

where $\mathcal{I}(\phi)$ is called Fisher information on ϕ. In particular, for pilot-only (PO) estimation, assuming that only K_{pilot} pilot symbols are used for estimation, the CRLB is given by

$$\mathrm{CRLB}_{\phi,\mathrm{PO}} = \frac{1}{2K_{\text{pilot}}}\frac{\sigma_v^2}{E_s}, \tag{231}$$

where E_s denotes the averaged energy per symbol [213]. Similarly, in the case of data-aided (DA) estimation, where K_{data} data symbols together with K_{pilot} pilot symbols ($K = K_{\text{pilot}} + K_{\text{data}}$) are exploited in estimation,[31] the CRLB is given by

$$\mathrm{CRLB}_{\phi,\mathrm{DA}} = \frac{1}{2K}\frac{\sigma_v^2}{E_s}. \tag{232}$$

The DA (or equivalently genie-aided) CRLB (232) is often employed as a lower bound for non-data-aided phase estimators, i.e., phase estimators which exploit the received

[31] It would probably be more appropriate to talk about the genie-aided estimation in this case, since the information symbols are unrealistically assumed to be known by the receiver.

samples affected by unknown data symbols. While important, DA-CRLB may be quite a loose bound at low SNR values.

Unfortunately, for the NDA estimation case, an analytical evaluation of an exact CRLB is in most cases not feasible. This is especially so for coded systems. However, a closed form expression for the NDA CRLB is provided in [53] in the case of BPSK and QPSK signaling and has later been extended to M-QAM modulated symbols in [212]. Further extensions of these CRLB formulas to support the pilot-symbol assisted transmission are provided in [191]. A drawback of these formulas is that their evaluation requires numerical integration, limiting their usefulness in general. Moreover, they do not apply to transmission systems which employ coded symbols.

On the other hand, using the identity [161]

$$\frac{\partial}{\partial\phi}\ln p(\mathbf{r}|\phi) = \left[\frac{\partial}{\partial\phi_0}E\big[\ln p(\mathbf{r}|\mathbf{s},\phi_0) \mid \mathbf{r},\phi\big]\right]_{\phi_0=\phi}, \tag{233}$$

the Fisher information on ϕ can alternatively be expressed as

$$\mathcal{I}(\phi) = \frac{4K}{(\sigma_v^2)^2}E_{p(\mathbf{r}|\phi)}\left[\left(\sum_{k=1}^{K}\Im\big\{r(k)\bar{s}^*(k)e^{-j\phi}\big\}\right)^2\right], \tag{234}$$

where the soft symbol decisions are determined in terms of the symbol APPs, i.e., $\bar{s}(k) = \sum_{s(k)\in\mathcal{S}} s(k)\,p\big(s(k)|\mathbf{r},\phi\big)$. These, on the other hand, are computed under the restrictions imposed by coding on the signal structure.[32] Consequently, the formula (234) may be considered as a basis for the numerical evaluation of the CRLB in general settings, including the case of encoded symbols. Specifically, approximating (234) by using empirical observed information yields

$$\text{CRLB}_{\phi,\text{NDA}} \approx \big(\mathcal{I}_e(\phi,\mathbf{r})\big)^{-1}, \tag{235}$$

where the empirical observed information $\mathcal{I}_e(\phi,\mathbf{r})$, so termed by Meilijson [162], is obtained simply by omitting the statistical expectation in (234) as

$$\mathcal{I}_e(\phi) = \frac{4K}{(\sigma_v^2)^2}\left(\sum_{k=1}^{K}\Im\big\{r(k)\bar{s}^*(k)e^{-j\phi}\big\}\right)^2. \tag{236}$$

Of course, a better approximation for the exact Fisher information may be obtained by approximating the statistical expectation in (234) by an arithmetical averaging. While this approach was first proposed by the author of this thesis in [188], a similar analysis was later carried out in an independent work reported in [190].

[32] An insertion of pilot symbols into the frame structure can also be considered a specific form of coding.

6.4 Adaptive turbo receivers in the presence of Wiener phase noise

In an AWGN channel distorted by a dynamic phase noise, an adaptive receiver must track the evolving phase trajectory, either implicitly or explicitly. Derivation of iterative turbo receivers for this type of channel is a topic of this section. First, the classical EM-based phase estimators are tailored for a changing phase environment via the "sliding window averaging," resulting in two types of soft decision-directed (SDD) digital phase locked loops (DPLL) for phase tracking. Secondly, the VB framework is directly applied to the *locally* linearized system model, resulting in an SDD extended Kalman smoother for phase tracking.

6.4.1 Recursive EM-based feedback phase recovery schemes

Soft decision-directed phase trackers for the time-varying phase can be built without explicitly taking into account the phase dynamics as such. For example, based on Fisher's identity (233), a stochastic gradient based maximum likelihood phase tracker, referred to as the SDD first-order digital phase locked loop of type 1 (DPLL-T1), can be formulated as

$$\hat{\phi}(k+1) = \hat{\phi}(k) + \mu' \, \Im\big\{ r(k)\bar{s}^*(k) e^{-j\hat{\phi}(k)} \big\}, \tag{237}$$

where μ' is a step-size parameter. This is a well-known phase tracker, proposed also independently in [191], which due to exploitation of the identity (233) could be called a *recursive* EM-based phase tracker.

In the case of BPSK modulation, the phase tracker in (237) can alternatively be written as

$$\hat{\phi}(k+1) = \hat{\phi}(k) + \mu' \, \tanh\bigg(\frac{2}{\sigma_v^2} \Re\Big\{ r(k) e^{-j\hat{\phi}(k)} \Big\} \bigg) \, \Im\Big\{ r(k) e^{-j\hat{\phi}(k)} \Big\}. \tag{238}$$

Furthermore, exploiting the fact that for the uncoded PSK signaling with $J \geq 4$ the symbol APPs can be expressed as

$$\bar{s}(k) = \frac{1}{\sum_{j=1}^{J/2} \cosh\Big(\frac{2}{\sigma_v^2} \Re\big\{ r(k)\alpha_j^* e^{-j\hat{\phi}(k)} \big\} \Big)} \sum_{j=1}^{J/2} \alpha_j \, \sinh\Big(\frac{2}{\sigma_v^2} \Re\big\{ r(k)\alpha_j^* e^{-j\hat{\phi}(k)} \big\} \Big), \tag{239}$$

where $\alpha_j \in \mathcal{S}$ and J denotes the cardinality of the symbol space, an SDD phase tracker in the case of uniform *a priori* symbol distribution may be written as

$$\hat{\phi}(k+1) = \hat{\phi}(k) + \frac{\mu'}{\sum_{j=1}^{J/2} \cosh\Big(\frac{2}{\sigma_v^2}\Re\big\{r(k)\alpha_j^* e^{-j\hat{\phi}(k)}\big\}\Big)} \times \sum_{j=1}^{J/2} \Im\big\{r(k)\alpha_j^* e^{-j\hat{\phi}(k)}\big\} \sinh\Big(\frac{2}{\sigma_v^2}\Re\big\{r(k)\alpha_j^* e^{-j\hat{\phi}(k)}\big\}\Big). \quad (240)$$

At low SNR values, using approximations $\sinh(x) \approx x + x^3/3!$ and $\cosh(x) \approx 1$, the phase tracker in (240) for $J \geq 4$ may be further simplified as

$$\hat{\phi}(k+1) = \hat{\phi}(k) + \mu \sum_{j=1}^{J/2} \Im\big\{r(k)\alpha_j^* e^{-j\hat{\phi}(k)}\big\} \Big[\Re\big\{r(k)\alpha_j^* e^{-j\hat{\phi}(k)}\big\}\Big]^3, \quad (241)$$

where the step size parameter μ is dependent on the receiver noise variance σ_v^2. In fact, (241) provides a mathematical description of the blind Costas loop, presented earlier in [164]. Hence, based upon the above formulation, the blind Costas loop can be interpreted as a low SNR approximation of the SDD-DPLL algorithm.

Another possibility is to introduce a forgetting factor into the recursive version of the incremental EM algorithm. Thus, we obtain an SDD phase tracker given as

$$\hat{\phi}(k) = \arg\Big\{\sum_{i=1}^{k} \lambda^{k-i}\, r^*(i)\bar{s}(i)\Big\}, \quad (242)$$

where λ is a forgetting factor. This phase tracker is henceforth referred to as the SDD digital phase locked loop of type 2 (DPLL-T2) in order to separate it from the gradient-based phase tracker. In fact, the argument of the right hand side of (242) can be recognized as an SDD version of the vector tracker proposed in [57].

6.4.2 VB-based turbo receivers

The sliding window averaging techniques were used above to make the classical EM-based phase estimators able to track changes in the value of the carrier phase. Next, a novel iterative receiver algorithm will be derived through an employment of the VB algorithm, with the main objective being a better adaptivity to the phase dynamics modeled by the Wiener process (211). Unfortunately, the nonlinearity of the system model makes a direct application of the VB algorithm very difficult, if not impossible. In order

to circumvent the difficulties due to nonlinearity, the model is first *locally* linearized by approximating $e^{j\phi(k)}$ with its first-order Taylor-series expansion around the previously estimated value of the phase $\hat{\phi}(k-1|k-1)$,[33] i.e.,

$$e^{j\phi(k)} \approx e^{j\hat{\phi}(k-1|k-1)} + je^{j\hat{\phi}(k-1|k-1)}\big(\phi(k) - \hat{\phi}(k-1|k-1)\big). \tag{243}$$

By inserting (243) into (210), the following linearized system model is obtained:

$$r(k) = j\, s(k)e^{j\hat{\phi}(k-1|k-1)}\phi(k) + s(k)e^{j\hat{\phi}(k-1|k-1)}\big(1 - j\hat{\phi}(k-1|k-1)\big) + v(k). \tag{244}$$

Before defining the VB algorithm for this linearized system model, a lemma, proving an equivalence which greatly facilitates the formulation of the VB solution for a system employing constant amplitude signaling, is provided.

Lemma 5 *Considering the constant amplitude signaling and linearized system model given by (244), the following equivalence holds true:*

$$\sum_{\mathbf{s}} q_{\mathbf{s}}(\mathbf{s}) \ln p(\mathbf{r}|\mathbf{s}, \boldsymbol{\phi}) = \ln p(\tilde{\mathbf{r}}|\boldsymbol{\phi}), \tag{245}$$

where the kth element of $\tilde{\mathbf{r}}$ is given as

$$\tilde{r}(k) = r(k)\bar{s}^*(k) - e^{j\hat{\phi}(k-1|k-1)}\big(1 - j\,\hat{\phi}(k-1|k-1)\big). \tag{246}$$

Proof. For the linearized system model (244),

$$\begin{aligned} E_{q_{\mathbf{s}}}\big[\ln p(\mathbf{r}|\mathbf{s}, \boldsymbol{\phi})\big] = -\frac{1}{\sigma_v^2}\sum_k E_{q_{\mathbf{s}}}\Big[&|r(k)|^2 \\ &- 2\Re\big\{r^*(k)s(k)e^{j\hat{\phi}(k-1|k-1)}\big(1 + j\phi(k) - j\hat{\phi}(k-1|k-1)\big)\big\} \\ &+ |s(k)|^2\Big(1 - 2\phi(k)\hat{\phi}(k-1|k-1) + |\hat{\phi}(k-1|k-1)|^2 + |\phi(k)|^2\Big)\Big] + C, \end{aligned} \tag{247}$$

where C is a constant independent of ϕ. If $|s(k)|^2 = 1$, (247) can be further written as (see [262] for similar formulation)

$$\begin{aligned} E_{q_{\mathbf{s}}}\big[\ln p(\mathbf{r}|\mathbf{s}, \boldsymbol{\phi})\big] &= -\frac{1}{\sigma_v^2}\sum_k \Big(|r(k)|^2 - 2\Re\big\{r^*(k)\bar{s}(k)e^{j\hat{\phi}(k-1|k-1)} \\ &\quad \times \big(1 + j\phi(k) - j\hat{\phi}(k-1|k-1)\big)\big\} \\ &\quad + \Big(1 - 2\phi(k)\hat{\phi}(k-1|k-1) + |\hat{\phi}(k-1|k-1)|^2 + |\phi(k)|^2\Big)\Big) \\ &\quad + C \\ &= \ln p(\tilde{\mathbf{r}}|\boldsymbol{\phi}). \end{aligned} \tag{248}$$

[33] The notation $\hat{\phi}(k-1|k-1)$ denotes the estimated value of the phase at time $k-1$ using the received signal samples up to time $k-1$.

□

On the other hand, for M-QAM signaling, the following approximation can be justified

$$\sum_{\mathbf{s}} q_{\mathbf{s}}(\mathbf{s}) \ln p(\mathbf{r}|\mathbf{s}, \boldsymbol{\phi}) \approx \ln p(\mathbf{r}|\bar{\mathbf{s}}, \boldsymbol{\phi}). \tag{249}$$

The approximation is tight if $\mathrm{Var}_{q_{s(k)}}\big(s(k)\big) \triangleq E_{q_{s(k)}}\big[|s(k) - \bar{s}(k)|^2\big]$ is small. Based on Lemma 5 and (249), the free distribution $q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ at the ith iteration is given by $q_{\boldsymbol{\phi}}^{(i)}(\boldsymbol{\phi}) = p\big(\boldsymbol{\phi}|\tilde{\mathbf{r}}^{(i-1)}\big)$ in the case of M-PSK signaling and $q_{\boldsymbol{\phi}}^{(i)}(\boldsymbol{\phi}) \approx p\big(\boldsymbol{\phi}|\mathbf{r}, \bar{\mathbf{s}}^{(i-1)}\big)$ in the case of M-QAM signaling. Due to joint Gaussianity of $\tilde{\mathbf{r}}$ (or $\mathbf{r}$) and $\boldsymbol{\phi}$, $q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ is in both cases also Gaussian. It should be noted that the phase posterior $q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ could be computed exactly also for the M-QAM modulated signals by applying the double-looping technique used in Chapter 4 for the SISO demodulation in the presence of a frequency-selective fading channel. This, however, would increase the computational complexity significantly.

Given that the estimated phase posterior $q_{\boldsymbol{\phi}}^{(i)}(\boldsymbol{\phi})$ is Gaussian, a straightforward formulation yields the following update equation for the product of symbol APPs:

$$q_{\mathbf{s}}^{(i)}(\mathbf{s}) = \prod_{k=1}^{K} p\big(s(k)\big) \cdot c_k \cdot e^{-\frac{1}{\sigma_v^2}\left(\left|r(k) - s(k)e^{j\hat{\phi}^{(i)}(k|K)}\right|^2 + |s(k)|^2 P^{(i)}(k|K)\right)}, \tag{250}$$

where

$$\hat{\phi}^{(i)}(k|K) = E_{q_{\boldsymbol{\phi}}^{(i)}}\big[\phi(k)\big] \tag{251}$$

$$P^{(i)}(k|K) = E_{q_{\boldsymbol{\phi}}^{(i)}}\Big[\big|\phi(k) - \hat{\phi}^{(i)}(k|K)\big|^2\Big] \tag{252}$$

denotes the posterior mean and variance of the phase sample $\phi(k)$ under $q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$, respectively. They are computed by the Kalman smoother operating on the locally linearized state space model given for the M-PSK modulated signals as (the iteration index i is omitted for notational convenience)

$$\phi(k) = \phi(k-1) + \Delta(k) \tag{253}$$

$$\tilde{r}(k) = je^{j\hat{\phi}(k-1|k-1)}\phi(k) + v(k). \tag{254}$$

The forward recursion of the Kalman smoother operating on the above state-space model has the following update equations:

$$P(k|k-1) = P(k-1|k-1) + \sigma_{\Delta}^2 \tag{255}$$

$$G_k = \frac{P(k|k-1)}{P(k|k-1) + \sigma_v^2} \tag{256}$$

$$\hat{\phi}(k|k) = \hat{\phi}(k-1|k-1) + G_k \, \Im\big\{ r(k)\bar{s}^*(k) e^{-j\hat{\phi}(k-1|k-1)} \big\} \tag{257}$$

$$P(k|k) = G_k \, \sigma_v^2. \tag{258}$$

The update equations for the backward processing are given as

$$S(k-1) = \frac{P(k-1|k-1)}{P(k-1|k-1) + \sigma_{\Delta}^2} \tag{259}$$

$$\hat{\phi}(k-1|K) = \hat{\phi}(k-1|k-1) + S(k-1)\big(\hat{\phi}(k|K) - \hat{\phi}(k-1|k-1)\big) \tag{260}$$

$$P(k-1|K) = P(k-1|k-1) + S(k-1)\big(P(k|K) - P(k|k-1)\big)S(k-1). \tag{261}$$

Using the approximation $\mathrm{Var}_{q_{s(k)}}\big(s(k)\big) \approx 0$ and proceeding along similar lines as above, the same KS update equations can be shown to be valid for M-QAM modulated signals as well, with the exception of the Kalman gain factor G_k that is computed as $G_k = P(k|k-1)/(P(k|k-1)|\bar{s}(k)|^2 + \sigma_v^2)$ for nonconstant amplitude symbols.

Just as the VB algorithm for the constant phase model was shown be closely related to the EM algorithm, the VB inference and estimation algorithm when applied to the linearized system model (244) is closely related to the Bayesian EM algorithm. In fact, as shown in Chapter 4, the Bayesian EM algorithm results from the VFEM framework by forcing the free distribution $q_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ to be composed of a train of spikes, i.e., $q_{\boldsymbol{\phi}}(\boldsymbol{\phi}) = \prod_{k=1}^{K} \delta\big(\phi(k) - \hat{\phi}(k)\big)$. The two algorithms are exactly the same if M-PSK signaling is assumed. This is an interesting result since it essentially shows that, whenever the original objective function is factorized into two independent factors, one being a function of the phase process while the other one is a function of data symbols, it is enough that the full description of our knowledge of $\boldsymbol{\phi}$, $p(\boldsymbol{\phi}|\mathbf{r})$, is being approximated by a delta function. The shape of the estimated phase posterior, no matter how sophisticated an approximation it may be, is not exploited anyhow by the VB-based receiver.

In contrast, the symbol decisions produced by the VB algorithm for the M-QAM constellations are biased towards small constellation points by the amount determined

by the error variance $P(k|K)$. Although the constellation points with large energy are penalized in a way analogous to the constant phase model, it should be noted that in many practical cases the phase posterior at any particular time instant is sharply peaked around the mean value (i.e., $P(k|K)$ is very small), and, hence, the VB algorithm in these circumstances performs equally to the Bayesian EM algorithm.

6.5 Complexity considerations

A short survey of complexity of SISO detectors in the presence of phase noise is presented in this section, partly building on the considerations made earlier in the context of adaptive SISO detectors in the unknown fading channels. As already pointed out earlier, the complexity of the exact APP detector is, in general, exponential in the sequence length. However, under certain circumstances, the worst case complexity of the APP detector in block-constant phase channels is only polynomial in the sequence length [174, 114]. On the other hand, in the case where the dynamic phase noise process can be assumed, for all practical purposes, to have a limited-length coherence time, the exact APP detector has complexity which is exponential in the channel coherence time [8, 266].

A suboptimal SISO detectors for all kinds of phase uncertain channels can be obtained via the PSP technique, with complexity which is scalable and can be easily adjusted by design so that a desired tradeoff between complexity and performance is reached [9]. The complexity issues related to the PSP receivers were earlier discussed in Section 4.5 to some extent. In cases where the transmitted signal is encoded by a code, which can be described with an FSM or a network of FSMs, such as a turbo-code, the PSP processing can be applied to the trellis that is associated with the code structure [9]. Thus, the complexity related to detection is closely connected to the complexity of the decoder, if such is needed.

The iterative EM- and BEM-based SISO detectors, encompassing thereby the incremental version of the EM algorithm, have complexity that is linear, per iteration, in the sequence length. Only one phase tracker is needed, and it is operated in an SDD manner. When iterative decoding is assumed, the SDD phase trackers can be conveniently embedded into the decoding algorithm, and, thus, the increased complexity, compared to the coherent receiver, is mainly due to channel tracking. On the other hand, the iterative decoder may require some extra decoding iterations because of the phase uncertainty. Similar remarks concerning complexity are valid also for the factor-graph

processing-based SISO detectors/decoders, such as the detector based on the Tikhonov parametrization of the phase distribution [48]. Differences between EM/BEM detectors and FG-processing-based detectors appear mainly in details of the phase estimator/tracker which in both approaches is operated in the SDD manner.

6.6 Numerical results

Studying the performance of the proposed adaptive turbo receivers, a constant phase model is considered first. The performance of the VB-based and incremental VB-based receivers is evaluated via numerical simulations for an uncoded and a coded QPSK signaling over the noncoherent AWGN channel. In fact, the VB algorithm performed equally with the EM algorithm in the selected simulation setups, and so the less complex EM algorithm was actually selected for final simulations. The phase is assumed to be constant during the block but changes randomly from block to block, where each block consists of 8 pilot symbols followed by 200 data symbols.

The MSE results for various phase estimators as a function of SNR are presented for an uncoded system in Fig. 35. The DA estimation in this context implies that both the pilot and data symbols are 'magically' known to the receiver, whereas the NDA phase estimator knows only the pilot symbols but additionally exploits the statistical information about the phase carried by the data symbols. The EM algorithm assumes five iterations, whereas the incremental EM algorithm updates phase estimates recursively, resulting in a noniterative receiver. In addition, the Cramer-Rao lower bound (CRLB) is plotted for the pilot-only, data-aided, and non-data-aided phase estimators using the empirical averaging technique outlined earlier in this chapter.

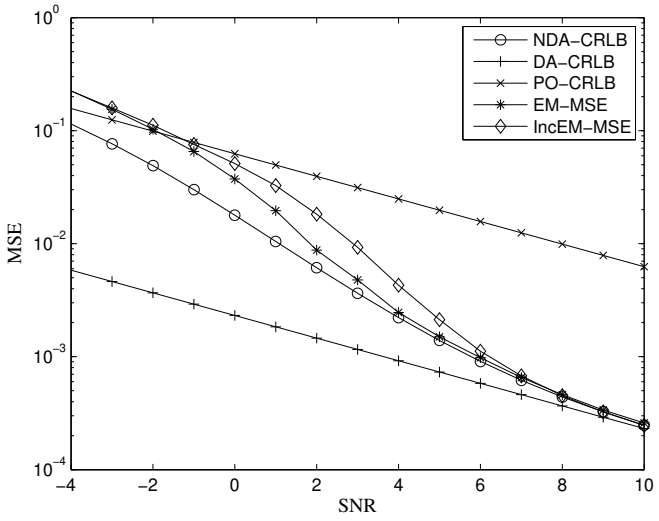

Fig 35. MSE performance of the phase estimators as a function of SNR for an uncoded QPSK-modulated system.

Examining Fig. 35, it may be noticed that, despite much smaller complexity, the performance of the incremental phase estimator in the MSE sense is very close to that of the EM-based estimator with five iterations. Actually, the performance of both estimators coincide with the $\mathrm{CRLB}_{\phi,\mathrm{NDA}}$ at high SNR values while some deviation takes place at low SNR values. Similar results, when channel coding is employed, are presented in Fig. 36. In this case, a convolutional code (CC) with 64 states and code rate 1/2 (code generators in octal form are (133,171)) is used for channel coding and a SISO decoder is assumed at the receiver. It is noticed that the exploitation of the coding-induced information in estimation via soft symbol decisions improves the performance of the iterative EM-based phase estimator at moderate SNR values. A similar trend can be noticed in the $\mathrm{CRLB}_{\phi,\mathrm{NDA}}$ curve for the code-aided (CA) estimation. The performance of the incremental EM-based estimator, which remains unchanged compared to the uncoded case (since feedback from the decoder is not used), is given as a reference.

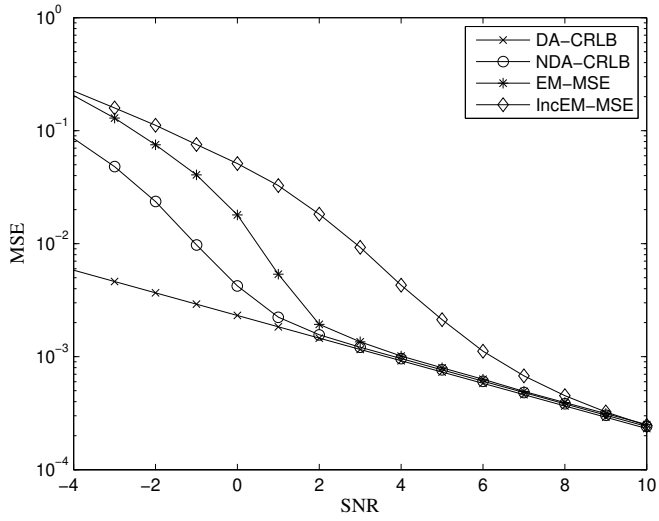

Fig 36. MSE performance of the phase estimators as a function of SNR for a coded QPSK-modulated system.

Despite considerable difference in MSE performance between the iterative EM-based and recursive (incremental) phase estimators, e.g., at SNR = 2 dB the difference is almost an order of magnitude, the performance of the corresponding detectors in terms of bit error rate (BER) differs just slightly at the same SNR point as shown in Fig. 37. On the basis of this single numerical example, it seems that the relationship between the BER of an adaptive detector and the MSE of the associated parameter estimator is highly nonlinear in nature and definitely an interesting issue to look at further.

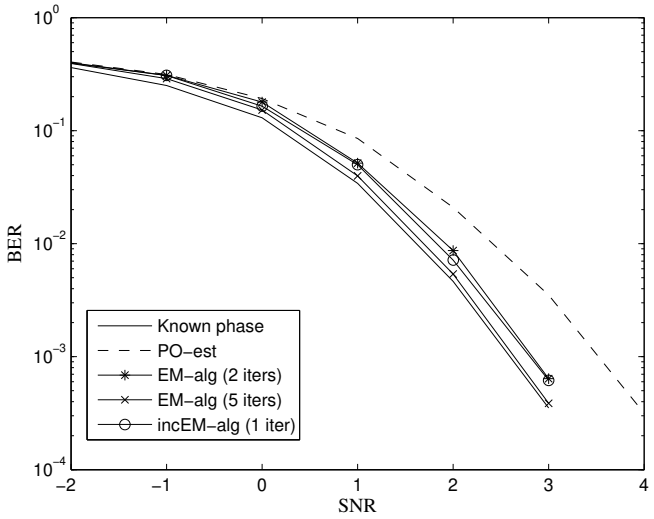

Fig 37. BER versus SNR for a coded QPSK modulated system in the presence of constant phase noise. The acronym PO-est stands for pilot only estimation.

In the second numerical example, the performance of the proposed iterative receiver schemes was assessed in a dynamic phase environment by computer simulations. In particular, there was a system under examination where turbo-coded bits were first interleaved with a random interleaver, then mapped into channel symbols using various modulation schemes, and finally transmitted over the AWGN channel distorted by the Wiener phase noise with $\sigma_\Delta = 3^\circ$. The turbo code was constructed according to the 3GPP specifications, and the code rate was 1/3. The coding block consisted of 12 sub-blocks, where each sub-block was comprised of 4 pilot symbols followed by 50 data symbols. At the beginning of each coding block, the number of pilot symbols was 20. The turbo decoder assumed 29 iterations, and, during the decoding process, the phase estimates were updated 5 times by utilizing the preliminary soft data estimates from the decoder. The numbers of decoding iterations for different decoding occasions were set at $[2, 3, 4, 5, 7, 8]$, summing up into 29 iterations in total.

For the purpose of comparison with the existing algorithms, the performance of the iterative receiver scheme based on the Tikhonov canonical distribution and SPA [48] as well as the performance of the receiver based on the SDD Kalman smoother

(KS) described in [73] are also simulated along with the proposed receiver schemes. Particularly the receiver based on the Tikhonov parameterization (hereafter referred to as the Tikhonov receiver) can be regarded as a state-of-the art solution for iterative detection in the presence of strong phase noise.

The BER results as a function of E_b/N_0 for BPSK, 16-QAM, and 64-QAM modulation schemes are presented in Fig. 38, Fig. 39, and Fig. 40, respectively. The label 'EKS' refers to the SDD extended Kalman smoother (obtained via the VB algorithm), whereas the label 'DPPL-T1' refers to the stochastic gradient-based SDD DPLL phase tracker (see (237)), and the label 'DPPL-T2' refers to the incremental EM-based SDD DPLL (see (242)). The step-size parameter for the DPPL-T1 estimator and the forgetting factor for the DPLL-T2 estimator were optimized separately at every simulated E_b/N_0 value. The basic difference between the EKS and KS estimators is that the KS estimator estimates the time-varying phasor as such, whereas the EKS estimator is based on the locally linearized system model enabling the tracking of the dynamic phase itself. Both of these are operated in the SDD manner.

As expected, the BPSK modulated system is less sensitive to phase estimation errors. In fact, the performance of the Tikhonov receiver is the same as the performance of the Kalman smoother (both EKS and KS)-based receivers, while the DPLL-based receivers experience a small performance loss compared to those. In the cases of higher order modulation, the Tikhonov- and EKS-based receivers still seem to have similar performance, while the KS- and DPLL-based receivers are not performing as well. Of the two DPLL estimators, the one based on the incremental EM algorithm seems to perform slightly better, particularly at high SNR values. Importantly, it can be noticed that both the Tikhonov- and EKS-based receivers have excellent performance even in the 64-QAM signaling system when compared to the performance of the coherent receiver.[34] It may also be noticed that the performance loss of the KS receiver as compared to the EKS receiver gets bigger as the size of the constellation alphabet increases, being roughly 1.5 dB for 64-QAM signaling.

It should, however, be noted that various system parameters, like the number of pilot symbols, pilot spacing, etc., may affect the performance of different receivers in a different way, and, thus, the comparisons made here may not be fully conclusive. Nevertheless, the presented simulation results strongly suggest that the Tikhonov receiver and the EKS-based receiver represent two efficient though distinct detection schemes

[34]It should be noted here that the coherent receiver assumes the same pilot overhead as the other receivers, and, thus, the performance gaps between different curves are solely due to the phase estimation.

with roughly the same performance and complexity. This may even be regarded as a main result of the numerical simulations reported in this section.

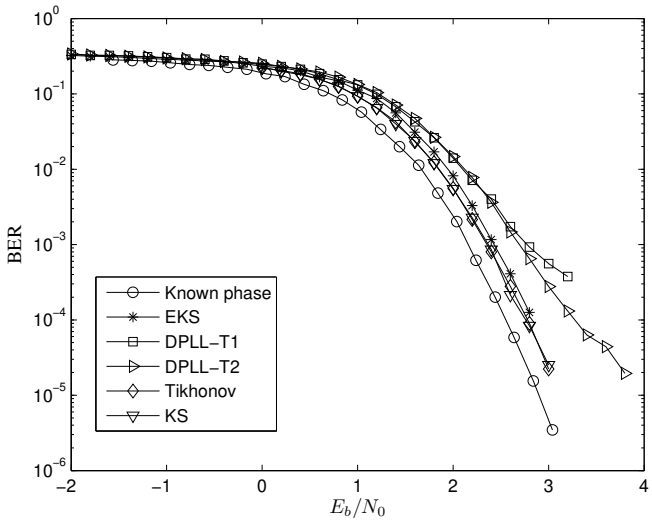

Fig 38. BER versus E_b/N_0 for a turbo-coded BPSK modulated system in the presence of Wiener phase noise.

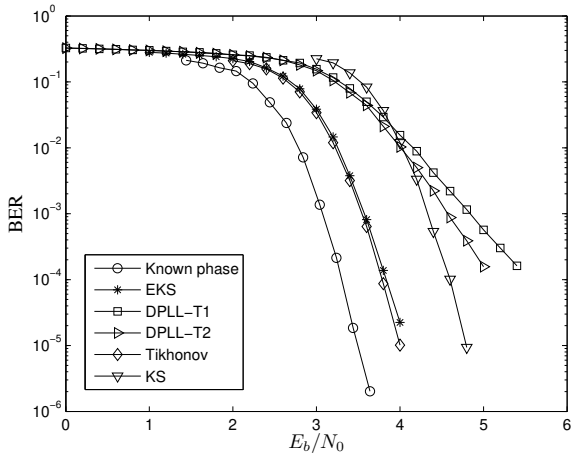

Fig 39. BER versus E_b/N_0 for a turbo-coded 16-QAM modulated system in the presence of Wiener phase noise.

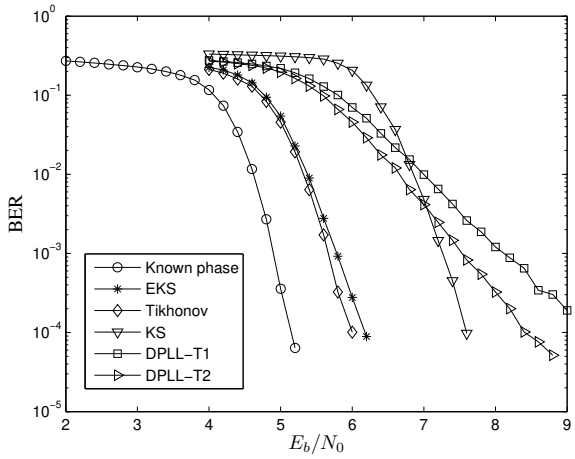

Fig 40. BER versus E_b/N_0 for a turbo-coded 64-QAM modulated system in the presence of Wiener phase noise.

6.7 Summary and conclusions

In this chapter, the problem of iterative detection in the presence of phase uncertainty was looked at from the variational Bayesian perspective. The variational Bayesian approach was shown to bridge the gap between the optimal noncoherent receivers and the classical EM-based receivers. In particular, the generic VB algorithm was applied for deriving novel receiver structures for the constant-phase channels as well as for the dynamic phase models. Moreover, it was demonstrated by numerical simulations that the incremental versions of the VB or EM algorithms may allow the reduction of the computational cost of the iterative receivers for the constant-phase channels. Furthermore, in the presence of high phase dynamics, the VB approach was shown to produce low-complexity yet powerful iterative receiver algorithms that can be regarded as a real alternative to the FG-processing-based receivers.

In parallel work reported in [147], the variational approximation technique has been used in deriving joint detection and phase estimation algorithms for OFDM systems. Although some conceptual similarities can be found in the work reported in [147] and in the one reported in this chapter, the assumed phase models and details in algorithmic development, in addition to assumed transmission formats of the signals, are essentially different in the respective contributions of these works, thus resulting in totally distinct sets of algorithms.

7 Summary and future work

So far, existing receiver algorithms have been reviewed and variational bounding methodology has been introduced and exploited in characterizing the existing turbo processing algorithms as well as in developing new ones. In addition, new algorithms for jointly estimating the radio channel parameters have been derived. In the following, the main results and findings of the thesis are shortly summarized and some prospects for future work are discussed.

7.1 Summary

In recent years, graphical models have gained popularity in the design and analysis of iterative turbo-processing receivers. They greatly enhance the representational power of the underlying probability models through qualitative characterization of their properties. Typically, this also leads to greater efficiency in terms of the computational algorithms that empower such representations. However, in many cases, for example in coded communications over the wideband channel, the exact probabilistic inference and estimation may become infeasible. With this in mind, variational approximations were introduced in Chapter 2 via four simple yet illuminating examples. The main aim was to point out the principle of convex duality as a solid foundation for obtaining lower bounds for the probabilities of interest. The emphasis was not on full generality but on demonstrating the utility of dual representations in transforming the log-likelihood function into a computationally feasible form.

Subsequently, in Chapter 3, the variational methodology was used in transforming the exact graphical models, representing the SISO demodulators in single antenna or MIMO systems, into approximate graphical models which render the probabilistic calculations in the context of turbo receivers more tractable. Specifically, the new formulation of existing linear SISO detectors through the mean field bounding method was shown to provide not only unified and rigorous justification for them but also suggestions on how to enhance their performance. Moreover, by exploiting the properties of the information divergence, more light was shed onto the domain characteristics in which one turbo receiver is preferable to others.

In Chapter 4, the problem of SISO demodulation in unknown frequency-selective

channels was addressed by relying again on the variational methodology in seeking computationally efficient yet close to optimal turbo receivers. First, the EM- and Bayesian EM-based SISO demodulators were obtained as appropriately defined mean field approximations to the exact APP demodulators in the constant deterministic and time-varying Rayleigh fading channels, respectively. Secondly, several recursive versions of the classical Baum-Welch algorithm and Bayesian EM algorithm were derived in an unified manner. Unlike the earlier attempts to cast these algorithms into recursive forms, the new formulations provided in Chapter 4 were shown to lead to computationally attractive soft decision-directed channel estimators which avoid matrix inversions by applying sequential processing over the *time* and *branch* indices of the associated trellis diagram. Moreover, they were shown to be readily integrable with the standard forward-backward processing SISO algorithms designed for known channels. Numerical results via computer simulations showed that the resulting approximate APP demodulators, when iteratively interconnected with the SISO decoder via turbo-processing, can achieve better performance than existing reference receivers, although their complexity is either comparable or significantly smaller.

The Bayesian channel estimator, whether taking a block or a recursive form, requires that the channel statistical properties are either known or appropriately estimated. In addition, frequency offset is typically present in the system, and, hence, the spectrum of the received signal is not exactly at the baseband but is instead shifted in frequency by the amount dictated by the frequency offset. In the generic system model presented in Chapter 2, the frequency offset was incorporated into the Rayleigh fading ISI channel via the complex bandpass AR modeling. Such a modeling concept was shown to be empowered by the assumption of circularity of the process noise. Later in Chapter 5, the complex bandpass modeling was shown to greatly facilitate the derivation of a joint estimator for the frequency offset and the channel AR parameters which, together, provide the parametric description of the fading statistics. The main advantage of the proposed joint estimator as compared to "competing" schemes is that it can perform equally well in a wide variety of scattering environments. This was shown to be valid also for extremely high-mobility environments. Furthermore, it was proved that the complexity of the EM-based estimators can be significantly reduced by applying the mean field approximation technique, while causing only minor penalty in performance.

Moreover, a fully adaptive recursive joint synchronization and channel estimation algorithm as well as a novel one-step Kalman smoother-based frequency error detector were proposed. Both the forward-processing adaptation scheme and the feedback-

processing frequency recovery scheme, incorporating the proposed error detector, were shown to be able to operate in either the hard decision-directed or soft decision-directed manner, which is a virtue in the presence of a limited amount of training. The simulation results showed that both the block-processing as well as the recursive frequency offset and AR estimators not only have excellent performance but also possess a number of highly favorable properties, such as a large estimation range and great flexibility with regard to the system and channel configurations and training signal structure. While the algorithms and numerical results were obtained for a single antenna system, they could be readily extended to systems employing multiple antennas at the transmitter. This is valid even in situations where the signals arriving from different transmitter antennas are not synchronized in frequency.

Iterative detection or decoding in the presence of phase uncertainty was addressed in Chapter 6. The problem was considered from the Bayesian modeling point of view. The optimal receivers obtained via the exact Bayesian formalism are generally known to be costly and often infeasible. This is due to the fact that the phase uncertainty has to be integrated out from the decision metric. In Chapter 6, a variational Bayesian approximation was used in deriving efficient iterative joint estimation and inference algorithms for constant-phase models as well as for dynamic phase models. In fact, the VB approach provides a lower bound on the likelihood function. Such a bound yields much more information than mere approximations and provides an inherent error metric for assessing the quality of the approximation. These desirable properties, concomitant of the variational methods, are unlikely to arise as a result of any other deterministic or stochastic approximation to the optimal noncoherent receiver. In addition, a close relationship between the VB-based receivers and the classical EM-based receivers was established, showing that the latter ones can be interpreted as special cases of the former ones.

In the class of non-Bayesian parameter estimators, the Cramer-Rao lower bound is known to provide a lower bound for the MSE of any unbiased parameter estimator. The computation of the CRLB can, however, become difficult, if not totally infeasible, in the presence of nuisance parameters that are interfering wih the estimation of the parameter of interest. In Chapter 5, the CRLB of the carrier frequency offset and receiver noise variance in the presence of the frequency-selective Rayleigh fading channel was formulated in an elegant way. This new formulation did not only facilitate the computation of the bound itself but also allowed to gain new insight into the factors affecting the accuracy of the estimate. Later in Chapter 6, the computation of the CRLB of the

carrier phase, when unknown data symbols were playing the role of nuisance parameters, was considered. In particular, exact Fisher information of the carrier phase was approximated with the corresponding empirical Fisher information. This was shown to make the numerical evaluation of the CRLB feasible also in cases where the information symbols are encoded.

7.2 Future work

There are still many open questions, pertaining to advanced receiver technology in general and to iterative detection and estimation in particular, that are not thoroughly examined in this thesis nor properly addressed in the open literature, although admittedly warrant detailed investigation. In fact, research carried out for the thesis has in some areas raised many more questions than it has definitely answered. Some of those open issues are shortly discussed in this section.

First, there are several issues relating to the characterization of the variational methods which are still, at least partially, unresolved. For example, a pertinent question is which dependencies will be lost and which are imposed by the variational approximation. The question as such is somewhat ambiguous since the answer, while being naturally contingent on the particular approximation method in use, depends also on the optimization schedule adopted for the variational free parameters. Then, there are still not many systematic answers to the question why a certain variational approximation may work fine in some circumstances while badly failing in others. This problem has so far been tackled mainly by examining example cases. For example, in some statistical mechanics systems, characterized by densely connected graphical models with uniformly weak couplings between the nodes, the mean field approximation does not only provide a good solution, but in fact an exact solution [198]. On the other hand, the same methodology in the context of turbo receivers is known to give often a very bad approximation as compared to the optimal receiver. It typically converges to some locally optimum solution, which is manifested by an error floor in the performance of the turbo receiver. Such a behavior could also be anticipated simply by viewing the graphical model associated with the SISO detector. It is typically densely connected, and dependencies between the nodes are relatively strong. That indicates that the objective function to be approximated may be multimodal. However, it would be nice if we would have a more principled way to evaluate the accuracy of certain variational approximation in the model of interest.

A big step forward in characterization of the variational approximations in general exponential-family graphical models has recently been taken in [251], however. Specifically, it was shown that a variety of approximate inference algorithms — including loopy belief propagation, general cluster variational methods, and mean field approximations — can be interpreted as methods for solving particular relaxations of the general variational principle that in turn was built on the solid foundation of convex optimization. Although not answering directly the above questions, the fresh viewpoint obtained via marginal polytopes does, however, provide a unified theoretical framework for comparing different approximations in the class of exponential graphical models. While important, the full theoretical characterization of variational methods in the context of an arbitrary statistical model seems, however, to be overly difficult. Instead, extensive experimental evaluation and comparison of different variational approximations, e.g., when applied to inference in communication systems, could be beneficial in the sense of increasing our understanding of the practical utility of these methods. Moreover, it would also be interesting to compare, by using numerical simulations, the variational methods to the various pruning-based methods as well as to the stochastic sampling-based methods [85].

In order to offer a healthy balance between performance and complexity, it could be advantageous to combine different variational methods in approximating a certain graphical model. This is an extension of the idea presented in [218], where the original graph was proposed to be divided into subgraphs, and the interfaces between different subgraphs were simplified by using variational methodology. At the same time, the exact inference was supposed to take place within each subgraph. In fact, the turbo receiver (equalizer) in its original form [65] is a realization of this principle. Along with these lines, extending the work presented in Chapter 3 could involve answering questions such as:

1. Can we find some systematic methodology to detect those edges in the graph whose removal variationally affects least the accuracy of the resulting approximation?
2. Could the mixing of the methods be done, for example, so that some edges in the graph are eliminated by applying the mean field approximation, while operating the belief propagation algorithm along the remaining edges?
3. How could we cope with the time-varying dependency structure of the model?
4. Can the variational optimization setting be reconfigured from block to block in some convenient way?

Such explorations should most obviously be carried out by bearing some practical application, such as turbo receivers, in mind. Yet another extension of the work in this thesis is to combine the variational methods with the stochastic sampling-based methods [85, 256]. However, it is less apparent how this can be done efficiently.

The problem of iterative detection in the presence of random continuous-valued parameters was dealt with in Chapters 4 and 6. Specifically, the variational methodology was applied for SISO demodulation in an unknown ISI channel and for iterative decoding in the presence of phase uncertainty. Another intuitively appealing solution for this type of problem is to model the continuous valued parameters as nodes in the graph and then operate the standard BP algorithm on it [261]. The computational inconvenience resulting from the requirement of computing high dimensional integrations has been tackled by parameterizing the messages originating from the parameter nodes. Thus, only a limited number of parameters, instead of full descriptions of the parameter distributions, need to be updated each time the node is visited. On the other hand, the EP algorithm has been shown to be better, at least in the tested statistical models, than the parameterizing approach in terms of performance [168]. One further option for extending the work of this thesis is to compare these three mainstream solutions at both the conceptual and the experimental levels.

Finally, it would be interesting and obviously useful to develop practical iterative receiver structures which would possess an improved capability to combat against the intra and intercell interference in a multiuser cellular environment [39]. This is particularly important in challenging cellular network topologies which are targeting for a reuse factor one. An interesting approach would be to test the applicability of the variational methodology in this area as well.

References

1. Abe T & Matsumoto T (2003) Space-time turbo equalization in frequency-selective MIMO channels. IEEE Trans. Veh. Technol. 52(3): 469–475.
2. Abend K & Fritchman BD (1970) Statistical detection for communication channels with intersymbol interference. Proc. IEEE 779–785.
3. Agrawal D & Vardy A (2001) The turbo decoding algorithm and its phase trajectories. IEEE Trans. Inform. Theory 47(2): 699–722.
4. Aji SM & McEliece RJ (2000) The generalized distributive law. IEEE Trans. Inform. Theory 46(2): 325–343.
5. Alexander PD, Reed MC, Asenstorfer JA & Schlegel CB (1999) Iterative multiuser interference reduction: Turbo CDMA. IEEE Trans. Commun. 47(7): 1008–1014.
6. Amari S (1985) Differential-geometrical methods in statistics. Springer-Verlag, Berlin.
7. Amari S (2001) Information geometry on a hierarchy of probability distributions. IEEE Trans. Inform. Theory 47(4): 1701–1711.
8. Anastasopoulos A & Chugg KM (2000) Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty. IEEE Trans. Commun. 48(10): 1638–1649.
9. Anastasopoulos A & Chugg KM (2001) Adaptive iterative detection for phase tracking in Turbo-coded systems. IEEE Trans. Commun. 49(12): 2135–2144.
10. Anastasopoulos A, Chugg KM, Colavolpe G, Ferrari G & Raheli R (2007) Iterative detection for channel with memory. Proceedings of the IEEE 95(6): 1272–1294.
11. Anderson BDO & Moore JB (1979) Optimal filtering. Prentice Hall, Englewood Cliffs, NJ.
12. Anderson JB & Mohan S (1984) Sequential coding algorithms: A survey cost analysis. IEEE Trans. Commun. 32(2): 169–176.
13. Antón-Haro C, Fonollosa JAR, Zvonar Z & Fonollosa JR (1998) Probabilistic algorithms for blind adaptive multiuser detection. IEEE Trans. Signal Proc. 46(11): 2953–2966.
14. Baccarelli E & Cusani R (1998) Combined channel estimation and data detection using soft statistics for frequency-selective fast-fading digital links. IEEE Trans. Commun. 46(4): 424–427.
15. Baccarelli E & Cusani R (1998) A novel adaptive receiver with enhanced channel tracking capability for TDMA-based mobile radio communications. IEEE J. Select. Areas Commun. 16(12): 1630–1639.
16. Bahl LR, Cocke J, Jelinek F & Raviv J (1974) Optimal decoding of linear codes for minimizing symbol error rate. IEEE Trans. Inform. Theory 20(2): 284–287.
17. Balasubramanian R & Fitz MP (2000) Soft-output detection of CPM signals in frequency flat, Rayleigh fading channels. IEEE J. Sel. Areas Commun. 18(7): 1145–1152.
18. Balasubramanian R, Fitz MP & Krogmeier JV (1998) Optimal and suboptimal symbol-by-symbol demodulation of continuous phase modulated signals. IEEE Trans. Commun. 46(12): 1662–1668.
19. Bar-David I & Elia A (1999) Augmented APP (A^2P^2) module for *a posteriori* probability calculation and channel parameter tracking. IEEE Communications Letters 3(1): 18–20.
20. Barbieri A & Colavolpe G (2007) On the Cramer-Rao bound for carrier frequency estimation in the presence of phase noise. IEEE Trans. Wireless Commun. 6(2): 575–582.
21. Barbieri A & Colavolpe G (2007) Soft-output decoding of rotationally invariant codes over

channels with phase noise. IEEE Trans. Commun. 55(11): 2125–2133.
22. Baskaran G, Fu Y & Anderson PW (1986) On the statistical mechanics of the traveling salesman problem. Journal of Statistical Physics 45: 1–25.
23. Bauch G & Al-Dhahir N (2002) Reduced-complexity space-time turbo-equalization for frequency-selective MIMO channels. IEEE Trans. Wireless Commun. 1(10): 819–828.
24. Baum LE, Petrie T, Soules G & Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Statist. 41(1): 164–171.
25. Beal MJ & Ghahramani Z (2003) The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. In: Bayesian Statistics, volume 7, 453–464. Oxford University Press.
26. Beal MJ & Ghahramani Z (2004) Variational Bayesian learning of directed graphical models with hidden variables. Bayesian Analysis 1(4): 651–980.
27. Benedetto S, Divsalar D, Montorsi G & Pollara F (1998) Soft-input soft-output modules for the construction and distributed decoding of code networks. Europ. Trans. Telecommun. (ETT) 9: 155–172.
28. Berrou C, Glavieux A & Thitimajshima P (1993) Near Shannon limit error-correcting coding and decoding: Turbo-codes. Proc. IEEE Int. Conf. Commun., 1064–1070.
29. Berthet AO, Ünal BS & Visoz R (2001) Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels. IEEE J. Select. Areas Commun. 19(9): 1729–1743.
30. Besson O & Stoica P (2001) On frequency offset estimation for flat-fading channels. IEEE Commun. Lett. 5(10): 402–404.
31. Bethe HA (1935) Statistical theory of superlattices. Proc. Roy. Soc. London A 150(871): 552–575.
32. Bhashyam S, Sayeed AM & Aazhang B (2000) Time-selective signaling and reception for communication over multipath fading channels. IEEE Trans. Commun. 48(1): 83–94.
33. Bilbro GL, Snyder WE, Garnier SJ & Gault JW (1992) Mean field annealing: a formalism for constructing GNC-like algorithms. IEEE Trans. Neural Networks 3(1): 131–138.
34. Bottomley GE & Chennakeshu S (1998) Unification of MLSE receivers and extension to time-varying channels. IEEE Trans. Commun. 46(4): 464–472.
35. Boutros J & Caire G (2002) Iterative multiuser joint decoding: unified framework and asymptotic analysis. IEEE Trans. Inform. Theory 48(7): 1772–1793.
36. Chandler D (1987) Introduction to modern statistical mechanics. Oxford University Press, Oxford, UK.
37. Chang RW & Hancock JC (1966) On receiver structures for channels having memory. IEEE Trans. Inform. Theory 12(4): 463–468.
38. Chiavaccini E & Vitetta GM (2001) MAP symbol estimation on frequency-flat Rayleigh fading channels via a Bayesian EM algorithm. IEEE Trans. Commun. 49(11): 1869–1872.
39. Choi J (2007) An EM-based iterative receiver for MIMO-OFDM under interference-limited environments. IEEE Trans. Wireless Commun. 6(11): 3994–4003.
40. Christensen LPB & Larsen J (2006) On data and parameter estimation using variational Bayesian EM-algorithm for block-fading frequency-selective MIMO channels. Proc. IEEE ICASSP'06, 4: 465–468.
41. Chugg KM (1998) Blind acquisition characteristics of PSP-based sequence detectors. IEEE J. Select. Areas Commun. 16(10): 1518–1529.

42. Chugg KM (1998) The condition for the applicability of the Viterbi algorithm with implications for fading channel MLSD. IEEE Trans. Commun. 46(9): 1112–1116.
43. Chugg KM & Polydoros A (1996) MLSE for an unknown channel—Part I: optimality considerations. IEEE Trans. Commun. 44(7): 836–846.
44. Cimini, Jr LJ (1985) Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. IEEE Trans. Commun. 33(7): 665–675.
45. Cirpan HA & Tsatsanis MK (2001) Maximum-likelihood estimation of FIR channels excited by convolutionally encoded inputs. IEEE Trans. Commun. 49(7): 1125–1128.
46. Clarke RH (1968) A statistical theory of mobile radio reception. Bell Syst. Tech. J. 47: 957–1000.
47. Colavolpe G (2006) On LDPC codes over channels with memory. IEEE Trans. Wireless Commun. 5(7): 1757–1766.
48. Colavolpe G, Barbieri A & Caire G (2005) Algorithms for iterative decoding in the presence of strong phase noise. IEEE J. Select. Areas Commun. 23(9): 1748–1757.
49. Colavolpe G, Ferrari G & Raheli R (2000) Noncoherent iterative (Turbo) decoding. IEEE Trans. Commun. 48(9): 1488–1498.
50. Colavolpe G, Ferrari G & Raheli R (2001) Extrinsic information in iterative decoding: a unified view. IEEE Trans. Commun. 49(12): 2088–2094.
51. Colavolpe G & Germi G (2005) On the application of factor graphs and the sum-product algorithm to ISI channels. IEEE Trans. Commun. 53(5): 818–825.
52. Cover T & Thomas J (1991) Elements of information theory. John Wiley & Sons Inc., New York.
53. Cowley WG (1996) Phase and frequency estimation for PSK packets: bounds and algorithms. IEEE Trans. Commun. 44(1): 26–28.
54. Csiszar I & Tusnady G (1984) Information geometry and alternating minimization procedures. In: Dudewics *et al* EJ (ed) Recent results in estimation theory and related topics. Statistics and Decision, Supplement Issue No. 1.
55. Dai Q & Shwedyk E (1994) Detection of bandlimited signals over frequency selective Rayleigh fading channels. IEEE Trans. Commun. 42(2/3/4): 941–950.
56. D'Andrea AN, Mengali U & Reggiannini R (1994) The modified Cramer-Rao bound and its application to synchronization problems. IEEE Trans. Commun. 42(2/3/4): 1391–1399.
57. D'Andrea AN, Mengali U & Vitetta GM (1994) Approximate ML decoding of coded PSK with no explicit carrier phase reference. IEEE Trans. Commun. 42(2/3/4): 1033–1039.
58. Dauwels J & Loeliger HA (2003) Joint decoding and phase estimation: an exercise in factor graphs. Proc. IEEE ISIT'03, 231.
59. Dauwels J & Loeliger HA (2004) Phase estimation by message passing. Proc. IEEE ICC'04, 523–527.
60. Davis LM, Collings IB & Evans RJ (1998) Coupled estimators for equalization of fast-fading mobile channels. IEEE Trans. Commun. 46(10): 1262–1265.
61. Davis LM, Collings IB & Hoeher P (2001) Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels. IEEE Trans. Commun. 49(12): 2106–2114.
62. Dempster AP, Laird NM & Rubin DB (1977) Maximum-likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. 39: 1–17.
63. Divsalar D, Dolinar S & Pollara F (2001) Iterative turbo decoder analysis based on density evolution. IEEE J. Select. Areas Commun. 19(5): 891–907.

64. Dong B, Wang X & Doucet A (2003) A new class of soft MIMO demodulation algorithms. IEEE Trans. Signal Proc. 51(11): 2752–2763.
65. Douillard C, Jézéquel M, Berrou C, Picard A, Didier P & Glavieux A (1995) Iterative correction of intersymbol interference: Turbo-equalization. Europ. Trans. Telecommun. (ETT) 6: 507–511.
66. Drost RJ & Singer AC (2007) Factor-graph algorithms for equalization. IEEE Trans. Signal Proc. 55(5): 2052–2065.
67. El Gamal H & Geraniotis E (2000) Iterative multiuser detection for coded CDMA signals in AWGN and fading channels. IEEE J. Select. Areas Commun. 18(1): 30–41.
68. Erfanian J, Pasupathy S & Gulak G (1994) Reduced complexity symbol detectors with parallel structures for ISI channels. IEEE Trans. Commun. 42(2/3/4): 1661–1671.
69. Falconer D, Ariyavisitakul SL, Benyamin-Seeyar A & Eidson B (2002) Frequency domain equalization for single-carrier broadband wireless communication. IEEE Commun. Magazine 40(4): 58–66.
70. Falconer DD & Salz J (1977) Optimal reception of digital data over the Gaussian channel with unknown delay and phase jitter. IEEE Trans. Inform. Theory 23(1): 117–126.
71. Feder M & Catipovic JA (1991) Algorithms for joint channel estimation and data recovery—application to equalization in underwater communications. IEEE J. Oceanic Eng. 16(1): 42–55.
72. Feder M & Weinstein E (1988) Parameter estimation of superimposed signals using the EM algorithm. IEEE Trans. Signal Proc. 36(4): 477–489.
73. Ferrari G, Colavolpe G & Raheli R (2004) Detection algorithms for wireless communications with applications to wired and storage systems. John Wiley & Sons.
74. Ferrari G, Colavolpe G & Raheli R (2005) A unified framework for finite-memory detection. IEEE J. Select. Areas Commun. 23(9): 1697–1706.
75. Fisher RFH (2002) Precoding and signal shaping for digital transmission. John Wiley & Sons, Inc.
76. Forney GD (1972) Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference. IEEE Trans. Inform. Theory 18(3): 363–378.
77. Foschini GJ & Gans MJ (1998) On limits of wireless communications in a fading environment when using multiple antennas. Wireless Pers. Commun. 6: 311–335.
78. Foschini GJ & Vannucci G (1988) Characterizing filtered light waves corrupted by phase noise. IEEE Trans. Inform. Theory 34(6): 1437–1448.
79. Fragouli C, Al-Dhahir N, Diggavi SN & Turin W (2002) Prefiltered space-time M-BCJR equalizer for frequency-selective channels. IEEE Trans. Commun. 50(5): 742–753.
80. Frey BJ (1998) Graphical models for machine learning and digital communication. MIT Press, Cambridge, MA.
81. Frey BJ, Koetter R & Vardy A (2001) Signal-space characterization of iterative decoding. IEEE Trans. Inform. Theory 47(2): 766–781.
82. Frey BJ, Patrascu R, Jaakkola TS & Moran J (2000) Sequentially fitting "inclusive" trees for inference in noisy-OR networks. Proc. Advances in Neural Information Processing Systems, 13.
83. Gallager RG (1962) Low density parity check codes. IRE Trans. Inform. Theory 8(1): 21–28.
84. Gazor S, Rabiei AM & Pasupathy S (2002) Synchronized per survivor MLSD receiver using a differential Kalman filter. IEEE Trans. Commun. 50(3): 364–368.

85. Geman S & Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Analysis and Machine Intelligence (11): 721–741.
86. Georghiades CN (1990) Optimal delay and sequence estimation from incomplete data. IEEE Trans. Inform. Theory 36(1): 202–208.
87. Georghiades CN & Han JC (1997) Sequence estimation in the presence of random parameters via the EM algorithm. IEEE Trans. Commun. 45(3): 300–308.
88. Georghiades CN & Snyder DL (1991) The expectation-maximization algorithm for symbol unsynchronized sequence detection. IEEE Trans. Commun. 39(1): 54–61.
89. Gertsman MJ & Lodge JH (1997) Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels. IEEE Trans. Commun. 45(7): 788–799.
90. Ghahramani Z & Jordan MI (1997) Factorial hidden Markov models. Machine Learning 29: 245–275.
91. Giallorenzi TR & Wilson SG (1996) Multiuser ML sequence estimator for convolutionally coded asynchronous DS-CDMA systems. IEEE Trans. Commun. 44(8): 997–1008.
92. Giallorenzi TR & Wilson SG (1996) Suboptimal multiuser receivers for convolutionally coded asynchronous DS-CDMA systems. IEEE Trans. Commun. 44(9): 1183–1196.
93. Gini F, Luise M & Reggiannini R (1998) Cramer-Rao bounds in the parametric estimation of fading radiotransmission channels. IEEE Trans. Commun. 46(10): 1390–1398.
94. Giridhar K, Shynk JJ & Iltis RA (1996) Adaptive MAPSD algorithms for symbol and timing recovery of mobile radio TDMA signals. IEEE Trans. Commun. 44(4): 976–987.
95. Gulati V & Heung-no Lee (2002) Low-complexity iterative per-antenna MAP equalizer for MIMO frequency selective fading channels. Proc. IEEE GLOBECOM'02, 2.
96. Gunawardana A & Byrne W (2005) Convergence theorems for generalized alternating minimization procedures. Journal of Machine Learning Research 6: 2049–2073.
97. Guo D & Verdu S (2005) Randomly spread CDMA: asymptotics via statistical physics. IEEE Trans. Inform. Theory 51(6): 1983–2010.
98. Guo D & Wang X (2003) Blind detection in MIMO systems via sequential Monte Carlo. IEEE J. Select. Areas Commun. 21(4): 464–473.
99. Haeb R & Meyr H (1989) A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels. IEEE Trans. Commun. 37(7): 748–754.
100. Hansson U & Aulin TM (1999) Aspects on single symbol signaling on the frequency flat Rayleigh fading channel. IEEE Trans. Commun. 47(6): 874–883.
101. Hart B & Taylor DP (1997) Extended MLSE diversity receiver for the time- and frequency-selective channel. IEEE Trans. Commun. 45(3): 322–333.
102. Hart BD, Borah DK & Pasupathy S (2000) Autocovariance preserving estimator (APE) interpretation of the MLSD metric for Rayleigh fading channels. IEEE Trans. Commun. 48(10): 1614–1617.
103. Hart BD & Pasupathy S (2000) Innovations-based MAP detection for time-varying frequency-selective channels. IEEE Trans. Commun. 48(9): 1507–1519.
104. Hart BD & Taylor DP (1998) Maximum-likelihood synchronization, equalization, and sequence estimation for unknown time-varying frequency-selective Rician channels. IEEE Trans. Commun. 46(2): 211–221.
105. Hart BD & Taylor DP (2000) On the irreducible error floor in fast fading channels. IEEE Trans. Veh. Technol. 49(3): 1044–1047.
106. Haykin S (1996) Adaptive Filter Theory. Prentice Hall, Upper Saddle River, NJ, third

edition.
107. Hebley MG & Taylor DP (1998) The effect of diversity on a burst-mode carrier-frequency estimator in the frequency-selective multipath channel. IEEE Trans. Commun. 46(4): 553–560.
108. Heo J, Chugg KM & Anastasopoulos A (2000) A comparison of forward-only and bi-directional fixed-lag adaptive SISOs. Proc. IEEE Int. Conf. Commun., 1660–1664.
109. Herzet C, Ramon V & Vandendorpe L (2007) A theoretical framework for iterative synchronization based on the sum-product and the expectation-maximization algorithms. IEEE Trans. Signal Proc. 55(5): 1644–1658.
110. Heskes T (2003) Stable fixed points of loopy belief propagation are minima of the Bethe free energy. Proc. Advances in Neural Information Processing Systems, 15: 343–350.
111. Heskes T, Albers K & Kappen B (2003) Approximate inference and constraint optimization. Proc. the 19th Conference on Uncertainty in Artificial Intelligence.
112. Hochwald BM & ten Brink S (2003) Achieving near-capacity on a multiple-antenna channel. IEEE Trans. Commun. 51(3): 389–399.
113. Hoeher P & Lodge J (1999) "Turbo DPSK": iterative differential PSK demodulation and channel decoding. IEEE Trans. Commun. 47(6): 837–843.
114. Hsu CH & Anastasopoulos A (2005) Design and analysis of joint data detection and frequency/phase estimation algorithms. IEEE J. Select. Areas Commun. 23(9): 1707–1717.
115. Iacono DL, Messina E, Volpe C & Spalvieri A (2005) Serial block processing for multi-code WCDMA frequency domain equalization. Proc. IEEE Wireless Commun. and Networking Conference (WCNC), 164–170.
116. Ikeda S, Tanaka T & Amari SI (2004) Information geometry of turbo and low-density parity-check codes. IEEE Trans. Inform. Theory 50(6): 1097–1114.
117. Iltis RA (1992) A Bayesian maximum-likelihood sequence estimation algorithm for *a priori* unknown channels and symbol timing. IEEE J. Sel. Areas Commun. 10(4): 579–588.
118. Jaakkola T (2000) Tutorial on variational approximation methods. In: Advanced mean field methods: theory and practice. MIT Press.
119. Jaakkola T, Saul LK & Jordan MI (1996) Fast learning by bounding likelihoods in Sigmoid type of belief networks. Proc. Advances in Neural Information Processing Systems, 8: 528–534.
120. Jagannatham AK & Rao BD (2004) Cramer-Rao lower bound for constrained complex parameters. IEEE Signal Proc. Lett. 11(11): 875–878.
121. Jakes WC (1974) Microwave mobile communications. John Wiley & Sons, New York, NY.
122. Jeong ER, Choi G & Lee YH (2001) Data-aided frequency estimation for PSK signaling in frequency-selective fading. IEEE J. Select. Areas Commun. 19(7): 1408–1419.
123. Jeong ER, Jo SK & Lee YH (2001) Least squares frequency estimation in frequency-selective channels and its application to transmissions with antenna diversity. IEEE J. Select. Areas Commun. 19(12): 2369–2380.
124. Jordan MI (2004) Graphical models. Statistical Science (Special Issue on Bayesian Statistics) 19: 140–155.
125. Jordan MI, Ghahramani Z, Jaakkola TS & Saul LK (1999) An introduction to variational methods for graphical models. Machine Learning 37: 183–233.
126. Jordan MI & Weiss Y (2002) Probabilistic inference in graphical models. In: Arbib M (ed) Handbook of Neural Networks and Brain Theory. MIT Press.
127. Kailath T (1960) Correlation detection of signals perturbed by a random channel. IRE

Trans. Inform. Theory 6(2): 361–366.
128. Kailath T (1969) A general likelihood-ratio formula for random signals in Gaussian noise. IEEE Trans. Inform. Theory 15(3): 350–361.
129. Kaleh GK & Vallet R (1994) Joint parameter estimation and symbol detection for linear and nonlinear unknown channels. IEEE Trans. Commun. 42(7): 2406–2413.
130. Kam PY & Teh CH (1983) Reception of PSK signals over fading channels via quadrature amplitude estimation. IEEE Trans. Commun. 31(8): 1024–1027.
131. Kappen H & Rodriguez P (1998) Efficient learning in Boltzmann machines using linear response theory. Neural Computation 10: 1137–1156.
132. Kay SM (1993) Fundamentals of statistical signal processing: estimation theory. Prentice-Hall, Englewood Cliffs, NJ.
133. Kittel C & Kroemer H (1980) Thermal physics. W.H. Freeman & Co., New York.
134. Kocarev L, Lehmann F, Maggio GM, Scanavino B, Tasev Z & Vardy A (2006) Nonlinear dynamics of iterative decoding systems: analysis and applications. IEEE Trans. Inform. Theory 52(4): 1366–1384.
135. Krasny L, Arslan H, Koilpillai D & Chennakeshu S (2001) Doppler spread estimation in mobile radio systems. IEEE Commun. Lett. 5(5): 197–199.
136. Kschischang FR & Frey BJ (1998) Iterative decoding of compound codes by probability propagation in graphical models. IEEE J. Select. Areas Commun. 16(2): 219–230.
137. Kschischang FR, Frey BJ & Loeliger HA (2001) Factor graphs and the sum-product algorithm. IEEE Trans. Inform. Theory 47(2): 498–519.
138. Kuo WY & Fitz MP (1997) Frequency offset compensation of pilot symbol assisted modulation in frequency flat fading. IEEE Trans. Commun. 45(11): 1412–1416.
139. Kurkoski BM, Siegel PH & Wolf JK (2002) Joint message-passing decoding of LDPC codes and partial-response channels. IEEE Trans. Inform. Theory 48(6): 1410–1422.
140. Laot C, Glavieux A & Labat J (2001) Turbo equalization: adaptive equalization and channel decoding jointly optimized. IEEE J. Select. Areas Commun. 19(9): 1744–1752.
141. Lauritzen SL (1996) Graphical models. Oxford Univ. Press, Oxford, U.K.
142. Lauritzen SL & Spiegelhalter DJ (1988) Local computations with probabilities on graphical structures and their application to expert systems. J. R. Statist. Soc. B 50(2): 157–224.
143. Li X & Wong TF (2007) Turbo equalization with nonlinear Kalman filtering for time-varying frequency-selective fading channels. IEEE Trans. Wireless Commun. 6(2): 691–700.
144. Li Y, Vucetic B & Sato Y (1995) Optimal soft-output detection for channels with intersymbol interference. IEEE Trans. Inform. Theory 41(3): 704–713.
145. Lin DD & Lim TJ (2005) A variational free energy minimization interpretation of multiuser detection in CDMA. Proc. IEEE GLOBECOM'05, 3: 1570–1575.
146. Lin DD & Lim TJ (2007) Multiuser detection of M-QAM symbols via bit-level equalization and soft detection. Proc. IEEE ISIT'07.
147. Lin DD & Lim TJ (2007) The variational inference approach to joint data detection and phase noise estimation in OFDM. IEEE Trans. Signal Proc. 55(5): 1862–1874.
148. Lin DD & Lim TJ (2007) A variational inference framework for soft-in-soft-out detection in interference channels. IEEE Trans. Inform. Theory (submitted).
149. Lindbom L, Sternad M & Ahlén A (2001) Tracking of time-varying mobile radio channels—Part I: the Wiener LMS algorithm. IEEE Trans. Commun. 49(12): 2207–2217.
150. Lodge JH & Moher ML (1990) Maximum likelihood sequence estimation of CPM signals

transmitted over Rayleigh flat-fading channels. IEEE Trans. Commun. 38(6): 787–794.
151. Loeliger HA (2001) On hybrid factor graphs and adaptive equalization. Proc. IEEE Int. Symp. Information Theory, 268.
152. Logothetis A & Krishnamurthy V (1999) Expectation maximization algorithms for MAP estimation of jump Markov linear systems. IEEE Trans. Signal Proc. 47(8): 2139–2156.
153. Lopes RR & Barry JR (2006) The soft-feedback equalizer for turbo equalization of highly dispersive channels. IEEE Trans. Commun. 54(5): 783–788.
154. Lottici V & Luise M (2004) Embedding carrier phase recovery into iterative decoding of Turbo-coded linear modulations. IEEE Trans. Commun. 52(4): 661–669.
155. Lucky RW, Salz J & Weldon (Jr) EJ (1968) Principles of data communication. McGraw-Hill, New York.
156. MacKay DJC (1997) Introduction to Monte Carlo methods. In: Jordan MI (ed) Learning in graphical models. Kluwer Academic Publishers.
157. Mämmelä A, Polydoros A & Järvensivu P (2002) Data and channel estimators: a systematic classification (invited plenary paper). Proc. X National Symposium of Radio Science (URSI).
158. Marple SL (1987) Digital spectral analysis with applications. Prentice Hall, Englewood Cliffs, NJ.
159. McEliece RJ, MacKay DJC & Cheng JF (1998) Turbo decoding as an instance of Pearl's 'belief propagation' algorithm. IEEE J. Select. Areas Commun. 16(2): 140–152.
160. McEliece RJ & Yildirim M (2002) Belief propagation on partially ordered sets. In: Mathematical Systems Theory in Biology, Communications, Computation, and Finance, 275–300.
161. McLachlan GJ & Krishnan T (1997) The EM algorithm and extensions. John Wiley & Sons Inc.
162. Meilijson I (1989) A fast improvement to the EM algorithm on its own terms. J. Roy. Statist. Soc. 51: 127–138.
163. Mendel JM (1971) Computational requirements for a discrete Kalman filter. IEEE Trans. Automatic Control 16(6): 748–758.
164. Mengali U & D'Andrea AN (1997) Synchronization techniques for digital receivers. Plenum Press, New York.
165. Meyr H, Moeneclaey M & Fechtel SA (1998) Digital communication receivers: synchronization, channel estimation, and signal processing. John Wiley & Sons, Inc.
166. Meyr H, Oerder M & Polydoros A (1994) On sampling rate, analog frefiltering, and sufficient statistics for digital receivers. IEEE Trans. Commun. 42(12): 3208–3214.
167. Miller MI & Snyder DL (1987) The role of likelihood and entropy in incomplete-data problems: applications to estimating point-process intensities and Toeplitz covariances. Proc. IEEE 75: 892–907.
168. Minka TP (2001) Expectation propagation for approximate Bayesian inference. Proc. the 17th Conference on Uncertainty in Artificial Intelligence, 362–369.
169. Moher M (1998) An iterative multiuser decoder for near-capacity communications. IEEE Trans. Commun. 46(7): 870–880.
170. Moher M & Gulliver TA (1998) Cross-entropy and iterative decoding. IEEE Trans. Inform. Theory 44(7): 3097–3104.
171. Morelli M & Mengali U (2000) Carrier-frequency estimation for transmissions over selective channels. IEEE Trans. Commun. 48(9): 1580–1589.

172. Morelli M, Mengali U & Vitetta GM (1998) Further results in carrier frequency estimation for transmissions over flat fading channels. IEEE Commun. Lett. 2(12): 327–330.
173. Morley (Jr) RE & Snyder DL (1979) Maximum likelihood sequence estimation for randomly dispersive channels. IEEE Trans. Commun. 27(6): 833–839.
174. Motedayen-Aval I & Anastasopoulos A (2003) Polynomial-complexity noncoherent symbol-by-symbol detection with application to adaptive iterative decoding of turbo-like codes. IEEE Trans. Commun. 51(2): 197–207.
175. Motedayen-Aval I, Krishnamoorthy A & Anastasopoulos A (2007) Optimal joint detection/estimation in fading channels with polynomial complexity. IEEE Trans. Inform. Theory 53(1): 209–223.
176. Neal RM & Hinton GE (1998) A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Jordan MI (ed) Learning in graphical models. Kluwer Academic Publishers.
177. Nissilä M & Pasupathy S (2003) Adaptive Bayesian and EM-based detectors for frequency-selective fading channels. IEEE Trans. Commun. 51(8): 1325–1336.
178. Nissilä M & Pasupathy S (2006) Joint estimation of carrier frequency offset and statistical parameters of the multipath fading channel. IEEE Trans. Commun. 54(6): 1038–1048.
179. Nissilä M & Pasupathy S (2007) Adaptive iterative detectors for phase-uncertain channels via variational bounding. IEEE Trans. Commun. (accepted for publication).
180. Nissilä M & Pasupathy S (2007) Soft-input soft-output equalizers for turbo receivers: a statistical physics perspective. IEEE Trans. Commun. 55(7): 1300–1307.
181. Nissilä M (2007) Variational Bayesian perspectives on iterative detection in the presence of phase uncertainty. Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC).
182. Nissilä M & Pasupathy S (2002) Adaptive Baum-Welch algorithms for frequency-selective fading channels. Proc. IEEE Int. Conf. Commun., 79–83.
183. Nissilä M & Pasupathy S (2002) Bayesian EM-based demodulators for frequency-selective fading channels. Proc. IEEE GLOBECOM, 1147–1151.
184. Nissilä M & Pasupathy S (2003) EM-based estimators for ARMA parameters of multipath fading channels. Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 399–402.
185. Nissilä M & Pasupathy S (2004) Joint estimation of frequency offset and fading rate in multipath fading channel using autoregressive channel modeling. Proc. IEEE 59th Vehicular Technology Conference (VTC), 1063–1067.
186. Nissilä M & Pasupathy S (2004) Low-complexity turbo receivers for multiple-antenna space-time coded systems via variational inference. Proc. IEEE 59th Vehicular Technology Conference (VTC), 530–534.
187. Nissilä M & Pasupathy S (2006) Turbo equalizers for MIMO systems: optimality consideration. Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC).
188. Nissilä MJ, Pasupathy S & Mämmelä A (2001) An EM approach to carrier phase recovery in AWGN channel. Proc. IEEE Int. Conf. Commun., 2199–2203.
189. Noels N, Herzet C, Dejonghe A, Lottici V, Steendam H, Moeneclaey M, Luise M & Vandendorpe L (2003) Turbo synchronization: an EM algorithm interpretation. Proc. IEEE Int. Conf. Commun., 2933–2937.
190. Noels N, Steendam H & Moeneclaey M (2003) The Cramer-Rao bound for phase estimation

from coded linearly modulated signals. IEEE Commun. Lett. 7(5): 207–209.
191. Noels N, Steendam H, Moeneclaey M & Bruneel H (2005) Carrier phase and frequency estimation for pilot-symbol assisted transmission: bounds and algorithms. IEEE Trans. Signal Proc. 53(12): 4578–4587.
192. Nuriyev R & Anastasopoulos A (2003) Pilot-symbol-assisted coded transmission over the block-noncoherent AWGN channel. IEEE Trans. Commun. 51(9): 953–963.
193. Oakes D (1999) Direct calculation of the information matrix via the EM algorithm. J. R. Statist. Soc. B 61: 479–482.
194. Omori H, Asai T & Matsumoto T (2001) A matched filter approximation for SC/MMSE iterative equalizers. IEEE Commun. Letters 5(7): 310–312.
195. Otnes R & Tüchler M (2004) Iterative channel estimation for turbo equalization of time-varying frequency-selective channels. IEEE Trans. Wireless Commun. 3(11): 1918–1913.
196. Pakzad P & Anantharam V (2005) Estimation and marginalization using Kikuchi approximation methods. Neural Computation 17: 1836–1873.
197. Pakzad P & Anantharam V (2006) Kikuch approximation method for joint decoding of LDPC codes and partial-response channels. IEEE Trans. Commun. 54(7): 1149–1153.
198. Parisi G (1988) Statistical field theory. Addison-Wesley, Redwood City, CA.
199. Pearl J (1988) Probabilistic reasoning in intelligent systems. Morgan Kaufman, San Mateo.
200. Piazzo L & Mandarini P (2002) Analysis of phase noise effects in OFDM modems. IEEE Trans. Commun. 50(10): 1696–1705.
201. Picinbono B (1994) On circularity. IEEE Trans. Signal Proc. 42(12): 3473–3482.
202. Picinbono B & Bondon P (1997) Second-order statistics of complex signals. IEEE Trans. Signal Proc. 45(2): 411–420.
203. Plischke M & Bergersen B (1989) Equilibrium statistical physics. Prentice-Hall, Englewood Cliffs, NJ.
204. Proakis JG (1989) Digital communications. McGraw-Hill, New York, second edition.
205. Qi Y & Minka TP (2007) Window-based expectation propagation for adaptive signal detection in flat-fading channels. IEEE Trans. Wireless Commun. 6(1): 348–355.
206. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77: 257–286.
207. Rabiner LR & Juang BH (1993) Fundamentals of speech recognition. Prentice Hall, Englewood Cliffs, NJ.
208. Raheli R, Marino G & Castoldi P (1996) Per-survivor processing and tentative decisions: what is in between? IEEE Trans. Commun. 44(2): 127–129.
209. Raheli R, Polydoros A & Tzou CK (1995) Per-survivor processing: a general approach to MLSE in uncertain environments. IEEE Trans. Commun. 43(2/3/4): 354–364.
210. Reed MC, Schlegel CB, Alexander PD & Asenstorfer JA (1998) Iterative multiuser detection for CDMA with FEC: near-single-user performance. IEEE Trans. Commun. 46(12): 1693–1699.
211. Reynolds D & Wang X (2001) Low-complexity turbo-equalization for diversity channels. Signal Processing 81(5): 989–995.
212. Rice F, Cowley B, Moran B & Rice M (2001) Cramer-Rao lower bounds for QAM phase and frequency estimation. IEEE Trans. Commun. 49(9): 1582–1591.
213. Rife DC & Boorstyn RR (1974) Single-tone parameter estimation from discrete-time observations. IEEE Trans. Inform. Theory 20(5): 591–598.
214. Rockafellar R (1972) Convex analysis. Princeton University Press.

215. Rollins ME & Simmons SJ (1997) Simplified per-survivor Kalman processing in fast frequency-selective channels. IEEE Trans. Commun. 45(5): 544–553.
216. Rustagi J (1976) Variational methods in statistics. Academic Press, New York.
217. Saul LK, Jaakkola T & Jordan MI (1996) Mean field theory for sigmoid belief networks. Journal of Artificial Intelligence Research 4: 61–76.
218. Saul LK & Jordan MI (1996) Exploiting tractable substructures in intractable networks. Proc. Advances in Neural Information Processing Systems, 8.
219. Sayed AH & Kailath T (1994) A state-space approach to adaptive RLS filtering. IEEE Signal Process. Mag. 11(7): 18–60.
220. Sayeed AM & Aazhang B (1999) Joint multipath-Doppler diversity in mobile wireless communications. IEEE Trans. Commun. 47(1): 123–132.
221. Sellathurai M & Haykin S (2002) TURBO-BLAST for wireless communications: theory and experiments. IEEE Trans. Signal Proc. 50(10): 2538–2546.
222. Seymour JP & Fitz MP (1995) Near-optimal symbol-by-symbol detection schemes for flat Rayleigh fading. IEEE Trans. Commun. 43(2/3/4): 1525–1533.
223. Shumway RH & Stoffer DS (1982) An approach to time series smoothing and forecasting using the EM algorithm. J. Time Series Anal. 3(4): 253–264.
224. Snyder DL, O'Sullivan JA & Miller MI (1989) The use of maximum-likelihood estimation for forming images of diffuse radar targets for delay-Doppler data. IEEE Trans. Inform. Theory 35(3): 536–548.
225. Song S, Singer AC & Sung KM (2004) Soft input channel estimation for turbo equalization. IEEE Trans. Signal Proc. 52(10): 2885–2894.
226. Steingrimsson B, Luo Z & Wong KM (2003) Soft quasi-maximum-likelihood detection for multiple-antenna wireless channels. IEEE Trans. Signal Proc. 51(11): 2710–2719.
227. Sutskover I & Shamai S (2006) Iterative decoding of low-density parity-check codes over compound channels. IEEE Trans. Commun. 54(2): 308–318.
228. Tanaka T (2002) A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors. IEEE Trans. Inform. Theory 48(11): 2888–2910.
229. Tanaka T & Okada M (2005) Approximate belief propagation, density evolution, and statistical neurodynamics for CDMA multiuser detection. IEEE Trans. Inform. Theory 51(2): 700–706.
230. Tarable A, Montorsi G & Benedetto S (2005) A linear front end for iterative soft interference cancellation and decoding in coded CDMA. IEEE Trans. Wireless Commun. 4(3): 507–518.
231. Tarokh V, Seshadri N & Calderbank AR (1998) Space-time codes for high data rate wireless communication: performance criterion and code construction. IEEE Trans. Inform. Theory 44(2): 744–765.
232. Taylor D, Vitetta G, Hart B & Mämmelä A (1998) Wireless channel equalisation. European Trans. Telecommun. 9: 117–143.
233. Taylor DP (1973) The estimate feedback equalizer: a suboptimal nonlinear receiver. IEEE Trans. Commun. 21(9): 979–990.
234. Teh YW & Welling M (2002) The unified propagation and scaling algorihm. Proc. Advances in Neural Information Processing Systems, 953–960.
235. ten Brink S (2001) Convergence behavior of iteratively decoded parallel concatenated codes. IEEE Trans. Commun. 49(10): 1727–1737.
236. Tong L & Perreau S (1998) Multichannel blind identification: from subspace to maximum

likelihood methods. Proc. IEEE 1951–1968.
237. Tsatsanis MK, Giannakis GB & Zhou G (1996) Estimation and equalization of fading channels with random coefficients. Signal Processing 53: 211–229.
238. Tüchler M, Koetter R & Singer AC (2002) Turbo equalization: principles and new results. IEEE Trans. Commun. 50(5): 754–767.
239. Tüchler M, Singer AC & Koetter R (2002) Minimum mean squared error equalization using *a priori* information. IEEE Trans. Signal Proc. 50(3): 673–683.
240. Tugnait JK, Tong L & Ding Z (2000) Single-user channel estimation and equalization. IEEE Signal Proc. Mag. (5): 17–28.
241. Ungerboeck G (1974) Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems. IEEE Trans. Commun. 22(5): 624–636.
242. Valenti MC & Woerner BD (2001) Iterative channel estimation and decoding of pilot symbol assisted Turbo codes over flat-fading channels. IEEE J. Select. Areas Commun. 19(9): 1697–1705.
243. van den Bos A (1994) A Cramér-Rao lower bound for complex parameters. IEEE Trans. Signal Proc. 42(10): 2859.
244. Van Trees HL (1968) Detection, estimation, and modulation theory, Part I. John Wiley and Sons, New York.
245. Viswanathan H & Krishnamoorthy R (2001) A frequency offset estimation technique for frequency-selective fading channels. IEEE Commun. Lett. 5(4): 166–168.
246. Vitetta GM & Taylor DP (1995) Maximum likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat-fading channels. IEEE Trans. Commun. 43(11): 2750–2758.
247. Vitetta GM & Taylor DP (1996) Multisampling receivers for uncoded and coded PSK signal sequences transmitted over Rayleigh frequency-flat fading channels. IEEE Trans. Commun. 44(2): 130–133.
248. Vogelbruch F & Haar S (2003) Improved soft ISI cancellation for turbo equalization using full soft output channel decoder's information. Proc. IEEE GLOBECOM'03, 3: 1736–1740.
249. Vogelbruch F & Haar S (2005) Low complexity turbo equalization based on soft feedback interference cancelation. IEEE Commun. Letters 9(7): 586–588.
250. Wainwright MJ, Jaakkola TS & Willsky AS (2003) Tree-based reparameterization framework for analysis of sum-product and related algorithms. IEEE Trans. Inform. Theory 49(5): 1120–1146.
251. Wainwright MJ & Jordan MI (2003) Variational inference in graphical models: The view from the marginal polytope. Proc. Allerton Conf. Communication, Control, and Computing.
252. Wainwright MJ & Jordan MI (2006) A variational principle for graphical models. In: New directions in statistical signal processing: from systems to brains. MIT Press.
253. Wang D, Hua J, Gao X, You X, Weckele M & Costa E (2004) Turbo detection and decoding for single-carrier block transmission systems. Proc. IEEE PIMRC'04, 1163–1167.
254. Wang D, Zhao J, Gao X & You X (2004) Space-time turbo detection and decoding for MIMO block transmission systems. Proc. IEEE Int. Conf. Commun., 2914–2918.
255. Wang X & Poor HV (1999) Iterative (Turbo) soft interference cancellation and decoding for coded CDMA. IEEE Trans. Commun. 47(7): 1046–1061.
256. Wang X & Poor HV (2004) Wireless communication systems: advanced techniques for

signal reception. Prentice Hall PTR, cop., Upper Saddle River, NJ.
257. Welling M & Teh YW (2001) Belief optimization for binary networks: a stable alternative to loopy belief propagation. Proc. the 17th Conference on Uncertainty in Artificial Intelligence.
258. Welling M & Teh YW (2003) Approximate inference in Boltzmann machines. Artificial Intelligence 143(1): 19–50.
259. Wiberg N, Loeliger HA & Kötter R (1995) Codes and iterative decoding on general graphs. Eur. Trans. Telecommun. 6: 513–525.
260. Winn J & Bishop CM (2005) Variational message passing. Journal of Machine Learning Research 6: 661–695.
261. Worthen AP & Stark WE (2001) Unified design of iterative receivers using factor graphs. IEEE Trans. Inform. Theory 47(2): 843–849.
262. Yan M & Rao BD (2003) Soft decision-directed MAP estimate of fast Rayleigh flat fading channels. IEEE Trans. Commun. 51(12): 1965–1969.
263. Yedidia JS (2001) An idiosyncratic journey beyond mean field theory. In: Opper M & Saad D (eds) Advanced Mean Field Methods, Theory and Practice. MIT Press.
264. Yedidia JS, Freeman WT & Weiss Y (2003) Understanding belief propagation and its generalizations. In: Lakemeyer G & Nebel B (eds) Exploring Artificial Intelligence in the New Millenium. Morgan Kaufmann.
265. Yedidia JS, Freeman WT & Weiss Y (2005) Constructing free-energy approximations and generalized belief propagation algorithms. IEEE Trans. Inform. Theory 51(7): 2282–2312.
266. Yu X & Pasupathy S (1995) Innovations-based MLSE for Rayleigh fading channels. IEEE Trans. Commun. 43(2/3/4): 1534–1544.
267. Yue G, Wang X & Madihian M (2007) Design of rate-compatible irregular repeat accumulate codes. IEEE Trans. Commun. 55(6): 1153–1163.
268. Yuille AL (2002) CCCP algorithms to minimize the Bethe amd Kikuchi free energies: convergent alternatives to belief propagation. Neural Computation 14(7): 1691–1722.
269. Zamiri-Jafarian H & Pasupathy S (1999) Adaptive MLSDE using the EM algorithm. IEEE Trans. Commun. 47(8): 1181–1193.
270. Zhang J (1992) The mean field theory in EM procedures for Markov random fields. IEEE Trans. Signal Proc. 40(10): 2570–2583.
271. Zhang J (1993) The mean field theory in EM procedures for blind Markov random field image restoration. IEEE Trans. Image Proc. 2(1): 27–39.
272. Zhang J & Hanauer GG (1995) The application of mean field theory to image motion estimation. IEEE Trans. Image Proc. 4(1): 19–33.
273. Zhang J, Modestino JW & Langan DA (1994) Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. IEEE Trans. Image Proc. 3(4): 404–420.
274. Zhang Y, Fitz MP & Gelfand SB (1997) Soft output demodulation of frequency-selective Rayleigh fading channels using AR channel models. Proc. IEEE GLOBECOM '97, Phoenix, AZ, USA, 327–330.
275. Zou Q, Tarighat A & Sayed AH (2007) Compensation of phase noise in OFDM wireless systems. IEEE Trans. Signal Proc. 55(11): 5407–5424.

Appendix 1 Proofs of lemmas and propositions of Chapter 3

Purely for notational simplicity and without any loss of generality, the proofs of lemmas and propositions of Chapter 3 are given here only in the special case of a single transmitter antenna system. The extension to the more general MIMO system case is straightforward and, therefore, a routine exercise.

Proof of Lemma 1:

Proof. A solution to the minimization in (64) simply follows from the following inequality

$$\begin{aligned}
&D\Big(p_r(\mathbf{s})\Big\|\prod_k q_k\Big) - D\Big(p_r(\mathbf{s})\Big\|\prod_k p_r\big(s(k)\big)\Big) \\
&= \sum_{\mathbf{s}} p_r(\mathbf{s}) \ln \frac{p_r(\mathbf{s})}{\prod_k q_k} - \sum_{\mathbf{s}} p_r(\mathbf{s}) \ln \frac{p_r(\mathbf{s})}{\prod_k p_r\big(s(k)\big)} \\
&= \sum_{\mathbf{s}} p_r(\mathbf{s}) \sum_k \ln \frac{p_r\big(s(k)\big)}{q_k} = \sum_k D\Big(p_r\big(s(k)\big)\big\|q_k\Big) \geq 0.
\end{aligned}$$

□

Proof of Proposition 1:

Proof. Using (63) and (67), the variational free energy $\mathcal{F}(Q)$ under the constraint of the fully factorized trial function, i.e., $Q(\mathbf{s}) = \prod_k q_k(s(k))$, can be expressed as

$$\begin{aligned}
\mathcal{F}(Q) &= \frac{1}{T}\sum_{\mathbf{s}} Q(\mathbf{s})\mathcal{E}(\mathbf{s}) + \sum_{\mathbf{s}} Q(\mathbf{s}) \ln Q(\mathbf{s}) \\
&= \frac{1}{T}\sum_{\mathbf{s}}\prod_l q_l \|\mathbf{r} - \mathbf{H}\mathbf{s}\|^2 - \sum_{\mathbf{s}}\sum_l q_l \ln \lambda_{2,l} \\
&\quad + \sum_{\mathbf{s}}\sum_l q_l \ln q_l + C,
\end{aligned} \tag{262}$$

where C denotes a constant. The fixed point equations for solving the minimum of $\mathcal{F}(Q)$ are obtained by finding the zero-gradient point of $\mathcal{F}(Q)$ with respect to each factor of $Q(\mathbf{s})$. The gradient with respect to q_k yields

$$\frac{\partial}{\partial q_k}\mathcal{F}(Q) = \frac{1}{T}\sum_{\mathbf{s}}\prod_{l\neq k} q_l \|\mathbf{r} - \mathbf{H}\mathbf{s}\|^2 - \ln \lambda_{2,k} + \ln q_k + 1. \tag{263}$$

By setting (263) to zero, the distribution $q_k(s_k)$ is obtained as

$$q_k(s(k)) = \lambda_{2,k}\frac{1}{\gamma_k}e^{-\frac{1}{T}\sum_{\mathbf{s}}\prod_{l\neq k}q_l(s(l))\|\mathbf{r}-\mathbf{H}\mathbf{s}\|^2} = \lambda_{2,k}\frac{1}{\gamma_k}e^{-\frac{1}{T}\mathcal{E}_k(s(k))}. \tag{264}$$

By exploiting the decoupling property of the fully factorized variational trial distribution $Q(\mathbf{s})$, the mean field symbol energy term $\mathcal{E}_k(s(k))$ can be expressed as

$$\begin{aligned}\mathcal{E}_k(s_k) &= \sum_{\mathbf{s}}\prod_{l\neq k}q_l(s(l))\big\|\mathbf{r}-\mathbf{H}\mathbf{s}\big\|^2 \\ &= \mathbf{r}^H\mathbf{r} - E_{Q_{\setminus k}}\big[\mathbf{s}^H\big]\mathbf{H}^H\mathbf{r} - \mathbf{r}^H\mathbf{H}E_{Q_{\setminus k}}\big[\mathbf{s}\big] + \mathrm{tr}\Big(E_{Q_{\setminus k}}\big[\mathbf{s}\mathbf{s}^H\big]\mathbf{H}^H\mathbf{H}\Big) \\ &= \Big\|\mathbf{r}-\mathbf{H}\overline{\mathbf{s}}+\mathbf{H}\mathbf{e}_k\big(\overline{s}(k)-s(k)\big)\Big\|^2 + C,\end{aligned} \tag{265}$$

where $E_{Q_{\setminus k}}\big[\cdot\big]$ denotes the expectation under the distribution $\prod_{l\neq k}q_l(s(l))$, tr denotes the trace of the matrix, C is a constant which is independent of $s(k)$, and $\overline{s}(k)$ denotes the soft symbol value defined in (74). By expanding (265) and incorporating the terms which are independent of $s(k)$ into the normalizing constant γ_k, equations (68)–(74) are found after straightforward elaboration. □

Proof of Proposition 2:

Proof. The variational free energy in this case can be formulated as

$$\begin{aligned}\mathcal{F}(Q) &= \mathcal{F}(\hat{\mathbf{s}},\boldsymbol{\Sigma}_o) \\ &= -\ln\det\boldsymbol{\Sigma}_o + \frac{1}{\sigma_v^2}\int_{\mathbf{s}}Q(\mathbf{s})\Big(\mathbf{r}^H\mathbf{r}-\mathbf{r}^H\mathbf{H}\mathbf{s}-\mathbf{s}^H\mathbf{H}^H\mathbf{r}+\mathbf{s}^H\mathbf{H}^H\mathbf{H}\mathbf{s}\Big)d\mathbf{s} \\ &\quad + \int_{\mathbf{s}}Q(\mathbf{s})\Big((\mathbf{s}-\overline{\mathbf{s}})^H\boldsymbol{\Sigma}_i^{-1}(\mathbf{s}-\overline{\mathbf{s}})\Big)d\mathbf{s} + C \\ &= -\ln\det\boldsymbol{\Sigma}_o - \frac{1}{\sigma_v^2}\Big(\big(\mathbf{H}^H\mathbf{r}+\sigma_v^2\boldsymbol{\Sigma}_i^{-1}\overline{\mathbf{s}}\big)^H\hat{\mathbf{s}} + \hat{\mathbf{s}}^H\big(\mathbf{H}^H\mathbf{r}+\sigma_v^2\boldsymbol{\Sigma}_i^{-1}\overline{\mathbf{s}}\big) \\ &\quad -\hat{\mathbf{s}}^H\big(\mathbf{H}^H\mathbf{H}+\sigma_v^2\boldsymbol{\Sigma}_i^{-1}\big)\hat{\mathbf{s}} - \mathrm{tr}\big(\boldsymbol{\Sigma}_o(\mathbf{H}^H\mathbf{H}+\sigma_v^2\boldsymbol{\Sigma}_i^{-1})\big)\Big) + C.\end{aligned} \tag{266}$$

Looking for the zero gradient point of $\mathcal{F}(\hat{\mathbf{s}},\boldsymbol{\Sigma}_o)$ with respect to $\hat{\mathbf{s}}$, standard matrix manipulation yields

$$\begin{aligned}\hat{\mathbf{s}} &= \big(\mathbf{H}^H\mathbf{H}+\sigma_v^2\boldsymbol{\Sigma}_i^{-1}\big)^{-1}\big(\mathbf{H}^H\mathbf{r}+\sigma_v^2\boldsymbol{\Sigma}_i^{-1}\overline{\mathbf{s}}\big) \\ &= \overline{\mathbf{s}} + \boldsymbol{\Sigma}_i\mathbf{H}^H\big(\mathbf{H}\boldsymbol{\Sigma}_i\mathbf{H}^H+\sigma_v^2\mathbf{I}\big)^{-1}\big(\mathbf{r}-\mathbf{H}\overline{\mathbf{s}}\big).\end{aligned} \tag{267}$$

Similarly, by finding the zero gradient point of $\mathcal{F}(\hat{\mathbf{s}},\boldsymbol{\Sigma}_o)$ with respect to $\sigma_{o,k}^2$, Eq. (77) is readily obtained. □

Appendix 2 Proofs of lemmas and theorems of Chapter 4

In this appendix, proofs of two theorems and one lemma of Chapter 4 of this thesis are provided. In particular, three different forms of the BEM algorithm for estimating the Gaussian frequency selective fading channel in the presence of unknown data symbols are given by these theorems and lemma. In general, instead of trying to calculate the MAP sequence estimate of random CIR vectors $\mathbf{\Theta} = \{\mathbf{h}(1), \ldots, \mathbf{h}(K)\}$ directly, the BEM algorithm pursues the maximum of the posterior density function $p(\mathbf{\Theta}|\mathbf{r})$ by alternately estimating the posterior pdf of the *complete* data set $\mathcal{X}$, given the *incomplete* data set $\mathcal{Y}$ and the current set of CIR estimates, and maximizing the estimated complete data pdf with respect to $\mathbf{\Theta}$. Specifically, the E-step and the M-step of the BEM algorithm at ith iteration can be expressed as

$$\mathtt{E-step}: \quad Q\big(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(i-1)}\big) = E\Big[\ln p(\mathcal{X}, \mathbf{\Theta})\Big|\mathcal{Y}, \hat{\mathbf{\Theta}}^{(i-1)}\Big] \tag{268}$$

$$\mathtt{M-step}: \quad \hat{\mathbf{\Theta}}^{(i)} = \arg\max_{\mathbf{\Theta}} Q\big(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(i-1)}\big). \tag{269}$$

For notational convenience and without losing any generality, the following proofs are given for the first order AR channel model ($p = 1$) and, in addition, the iteration indices are suppressed.

Proof of Theorem 1:

Proof. By defining the complete data set as $\mathcal{X} = \{\mathbf{r}, \mathbf{s}\}$ and the incomplete data set as $\mathcal{Y} = \{\mathbf{r}\}$, the objective function $Q(\mathbf{\Theta}|\hat{\mathbf{\Theta}})$ can be expanded as follows:

$$\begin{aligned}
Q(\mathbf{\Theta}|\hat{\mathbf{\Theta}}) &= E\Big[\ln p(\mathcal{X}|\mathbf{\Theta})\Big|\mathbf{r}, \hat{\mathbf{\Theta}}\Big] + \ln p(\mathbf{\Theta}) \\
&= C - \frac{1}{\sigma_v^2}\sum_k\sum_j\Big(r(k)r^*(k) - 2\Re\big\{r(k)\mathbf{h}^H(k)\boldsymbol{\xi}_j^*\varphi(k,j)\big\} \\
&\quad + \mathbf{h}^H(k)\boldsymbol{\xi}_j^*\boldsymbol{\xi}_j^T\mathbf{h}(k)\varphi(k,j)\Big) + \ln p(\mathbf{\Theta}) \\
&= C - \frac{1}{\sigma_v^2}\sum_k\sum_j\Big(\tilde{r}_j(k)\tilde{r}_j^*(k) - 2\Re\big\{\tilde{r}_j(k)\mathbf{h}^H(k)\tilde{\mathbf{d}}_j^*(k)\big\} \\
&\quad + \mathbf{h}^H(k)\tilde{\mathbf{d}}_j^*(k)\tilde{\mathbf{d}}_j^T(k)\mathbf{h}(k)\Big) + \ln p(\mathbf{\Theta}) + [\text{terms independent of } \mathbf{\Theta}] \\
&= \ln p\big(\{\tilde{\mathbf{r}}(k)\}_{k=1}^K, \{\tilde{\mathbf{D}}(k)\}_{k=1}^K, \mathbf{\Theta}\big) + [\text{terms independent of } \mathbf{\Theta}],
\end{aligned} \tag{270}$$

where $\tilde{r}_j(k) = r(k)\sqrt{\varphi(k,j)}$ and $\tilde{\mathbf{d}}_j(k) = \boldsymbol{\xi}_j\sqrt{\varphi(k,j)}$. Furthermore, the vector

$\tilde{\mathbf{r}}(k)$ and the matrix $\tilde{\mathbf{D}}(k)$ are defined in (126) and (127), respectively. Due to the joint Gaussianity of sequences $\boldsymbol{\Theta}$ and $\{\tilde{\mathbf{r}}\}_1^K$, the M-step of the BEM algorithm can be written as (see, e.g., [132])

$$\hat{\mathbf{h}}(k) = E\Big[\mathbf{h}(k)\Big|\{\tilde{\mathbf{r}}\}_1^K, \{\tilde{\mathbf{D}}\}_1^K\Big], \quad k = 1, \ldots, K, \tag{271}$$

and, hence, the sequence estimate $\hat{\boldsymbol{\Theta}}$ is computed by the fixed-interval Kalman smoother operating on the state-space model given in (124) and (125). This vector Kalman smoother is iteratively cross-coupled with the BCJR algorithm, since the branch APPs $\varphi(k, j)$ are obtained from the BCJR algorithm. □

Proof of Lemma 2:

Proof. Given that complete data set is defined as $\mathcal{X} = \big\{\{x_l(k)\}_{l=1,\ldots,L+1}^{k=1,\ldots,K}, \mathbf{s}\big\}$ and the incomplete data set as $\mathcal{Y} = \{\mathbf{r}\}$, the objective function of the BEM algorithm can be written as

$$\begin{aligned}
Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}) &= E\bigg[\sum_l \sum_k \ln p(x_l(k)|\boldsymbol{\Theta})\Big|\mathbf{r}, \hat{\boldsymbol{\Theta}}\bigg] + \ln p(\boldsymbol{\Theta}) \\
&= -\frac{1}{\eta\sigma_v^2}\sum_l \sum_k E\bigg[\big|x_l(k) - s(k-l+1)h_{l-1}(k)\big|^2\Big|\mathbf{r}, \hat{\boldsymbol{\Theta}}\bigg] + \ln p(\boldsymbol{\Theta}) + C \\
&= \frac{1}{\eta\sigma_v^2}\sum_l \sum_k \Big(2\Re\Big\{E\big[x_l^*(k)s(k-l+1)|\mathbf{r}, \hat{\boldsymbol{\Theta}}\big]h_{l-1}(k)\Big\} \\
&\quad - |h_{l-1}(k)|^2 E\Big[|s(k-l+1)|^2\Big|\mathbf{r}, \hat{\boldsymbol{\Theta}}\Big]\Big) + \ln p(\boldsymbol{\Theta}) + [\text{terms independent of } \boldsymbol{\Theta}] \\
&= \frac{1}{\eta\sigma_v^2}\sum_l \sum_k \Bigg(2\Re\Big\{\sum_{j=1}^{J^{L+1}} \hat{x}_{l,j}^*(k)\big[\boldsymbol{\xi}_j\big]_l \varphi(k,j) h_{l-1}(k)\Big\} \\
&\quad - |h_{l-1}(k)|^2 \sum_{j=1}^{J^{L+1}} \Big|\big[\boldsymbol{\xi}_j\big]_l\Big|^2 \varphi(k,j)\Bigg) + \ln p(\boldsymbol{\Theta}) + [\text{terms independent of } \boldsymbol{\Theta}],
\end{aligned} \tag{272}$$

where $\hat{x}_{l,j}(k) \triangleq E\Big[x_l(k)\big|\mathbf{r}, \mathbf{s}(k) = \boldsymbol{\xi}_j, \hat{\boldsymbol{\Theta}}\Big]$ denotes the per-trellis-branch-based estimate of the pseudo-observation $x_l(k)$. Using the joint Gaussianity of the received signal sample $r(k)$ and pseudo-observations $\{x_l(k)\}_{l=1\ldots(L+1)}$ as well as their linear dependence as per (136), a straightforward manipulation yields (see, e.g., [72]) that $\hat{x}_{l,j}(k) = \big[\boldsymbol{\xi}_j\big]_l \hat{h}_{l-1}^{(i)}(k) + \eta\big(r(k) - \boldsymbol{\xi}_j^T \hat{\mathbf{h}}^{(i)}(k)\big)$. Using the definitions $\tilde{\mathbf{d}}_j(k) = \boldsymbol{\xi}_j\sqrt{\varphi(k,j)}$ and $\tilde{x}_{l,j}(k) = \hat{x}_{l,j}(k)\sqrt{\varphi^{(i)}(k,j)}$, the objective function of the BEM algorithm can further

be written as

$$\begin{aligned} Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}) &= \frac{1}{\eta\sigma_v^2}\sum_l\sum_k\left(2\Re\Big\{\sum_{j=1}^{J^{L+1}}\tilde{x}_{l,j}^*(k)\big[\tilde{\mathbf{d}}_j(k)\big]_l h_{l-1}(k)\Big\}\right. \\ &\quad \left. - |h_{l-1}(k)|^2\sum_{j=1}^{J^{L+1}}\Big|\big[\tilde{\mathbf{d}}_j(k)\big]_l\Big|^2\right) + \ln p(\boldsymbol{\Theta}) + [\text{terms independent of } \boldsymbol{\Theta}] \\ &= \ln p\Big(\{\tilde{\mathbf{x}}_l(k)\}_{k=1}^K, \big\{\big[\tilde{\mathbf{D}}(k)\big]_l\big\}_{k=1}^K, \boldsymbol{\Theta}\Big) + [\text{terms independent of } \boldsymbol{\Theta}]. \end{aligned} \tag{273}$$

Due to the joint Gaussianity of the sequences $\boldsymbol{\Theta}$ and $\{\tilde{\mathbf{x}}_l\}_1^K$, the M-step of the BEM algorithm can be written as

$$\hat{\mathbf{h}}(k) = E\Big[\mathbf{h}(k)\Big|\big\{\tilde{\mathbf{x}}\big\}_1^K, \big\{\tilde{\mathbf{D}}\big\}_1^K\Big], \quad k = 1, \ldots, K, \tag{274}$$

and the sequence estimate $\hat{\boldsymbol{\Theta}}$ is computed by the fixed-interval Kalman smoother operating on the state-space model given in (124) and (138). The branch APPs $\varphi(k, j)$ are obtained at the E-step by using the BCJR algorithm. □

Proof of Theorem 2:

Proof. The constant envelope modulation (with the symbol amplitudes being normalized to one) entails that $\Big|\big[\boldsymbol{\xi}_j\big]_l\Big|^2 = 1$. In this case, the objective function of the BEM algorithm can be written as

$$\begin{aligned} Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}) &= \frac{1}{\eta\sigma_v^2}\sum_l\sum_k\left(2\Re\Big\{\sum_{j=1}^{J^{L+1}}\hat{x}_{l,j}^*(k)\big[\boldsymbol{\xi}_j\big]_l\varphi(k,j)h_{l-1}(k)\Big\} - |h_{l-1}(k)|^2\right) \\ &\quad + \ln p(\boldsymbol{\Theta}) + [\text{terms independent of } \boldsymbol{\Theta}] \\ &= \frac{1}{\eta\sigma_v^2}\sum_k\Big(2\Re\big\{\hat{\mathbf{x}}^H(k)\mathbf{h}(k)\big\} - \mathbf{h}^H(k)\mathbf{h}(k)\Big) \\ &\quad + \ln p(\boldsymbol{\Theta}) + [\text{terms independent of } \boldsymbol{\Theta}] \\ &= \ln p\big(\{\hat{\mathbf{x}}(k)\}_{k=1}^K, \boldsymbol{\Theta}\big) + [\text{terms independent of } \boldsymbol{\Theta}], \end{aligned} \tag{275}$$

where $\hat{\mathbf{x}}(k)$ is defined as in (146). Due to the joint Gaussianity of the sequences $\boldsymbol{\Theta}$ and $\{\hat{\mathbf{x}}\}_1^K$, the sequence estimate $\hat{\boldsymbol{\Theta}}$ is computed by the fixed-interval Kalman smoother operating on the state-space model given in (124) and (145). □

Appendix 3 On ARMA channel modeling and estimation

The statistics of the Rayleigh fading channel can efficiently be described by the means of the parametric model. The channel estimation in the context of the parametric modeling is essentially a three-step procedure where the first step is to select a suitable model. The second step is then to estimate the parameters of the assumed model, and the final step is to obtain the estimate of the channel impulse response (for coherent detection) or the power spectral density (for the configuration of the transmitter adaptivity) with the aid of the estimated parameters. In this appendix, the signal modeling and estimation issues are discussed in the case where the ARMA(p, q) channel hypermodel is employed.

When ARMA channel modeling ($q \geq p$) is used,[35] the dynamics of the received signal may be described with the following state-space model

$$\boldsymbol{f}(k) = \boldsymbol{A}\boldsymbol{f}(k-1) + \boldsymbol{G}\mathbf{w}(k-1) \tag{276}$$

$$r(k) = \boldsymbol{s}^T(k)\boldsymbol{B}\boldsymbol{f}(k) + v(k), \tag{277}$$

where the vector channel process $\boldsymbol{f}(k)$ is now defined as $\boldsymbol{f}(k) = \left[\mathbf{f}^T(k), \mathbf{f}^T(k-1), \ldots, \mathbf{f}^T(k-q)\right]^T$, the deterministic model matrix $\boldsymbol{A}$ is given as

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_p & \mathbf{0}_{L_{\text{ch}} \times (q-p+1)L_{\text{ch}}} \\ \mathbf{I}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}}} & \cdots & \mathbf{0}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}} \times (q-p+1)L_{\text{ch}}} \\ & \ddots & \mathbf{0}_{L_{\text{ch}}} & & \vdots \\ & \mathbf{0}_{L_{\text{ch}}} & \mathbf{I}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}}} & \mathbf{0}_{L_{\text{ch}} \times (q-p+1)L_{\text{ch}}} \end{bmatrix},$$

and the model matrix $\boldsymbol{B}$ includes the MA parameters and is defined as $\boldsymbol{B} = [\mathbf{B}_0, \ldots, \mathbf{B}_q]$, where $\mathbf{B}_0, \ldots, \mathbf{B}_q$ are $L_{\text{ch}} \times L_{\text{ch}}$ submatrices. The rest of the variables in (276) and (277) are as defined in Chapter 2.

In this case, the vector of unknown parameters of the above state-space model is given as $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, where $\boldsymbol{\theta}_1 = \{\sigma_v^2, \boldsymbol{B}\}$ and $\boldsymbol{\theta}_2 = \{\mathbf{G}, \boldsymbol{A}\}$ include the parameters of the measurement equation and the state equation, respectively. Direct estimation of model parameters is not feasible. Instead, an iterative processing via the EM algorithm

[35] The case $q < p$ requires only a trivial modification of the case $q \geq p$ from the modeling and estimation points of view and, hence, it is not discussed here.

makes the estimation problem much easier to tackle. As in the AR modeling case, the objective function of the EM algorithm at the ith iteration can be decomposed as

$$\begin{aligned} Q\Big(\boldsymbol{\Theta}\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\Big) &= Q_1\left(\boldsymbol{\theta}_1\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\right) + Q_2\left(\boldsymbol{\theta}_2\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\right) \\ &= E\Big[\ln p(\mathbf{r}|\mathbf{s},\boldsymbol{\theta}_1,\mathbf{F})\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big] + E\Big[\ln p(\mathbf{F}|\boldsymbol{\theta}_2)\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big]. \end{aligned} \tag{278}$$

Although the maximization of (278) over the AR and MA parameters is now decoupled, the parameter estimators themselves are coupled through the conditional expectations (through the E-step).

The parameter estimator for $\boldsymbol{\theta}_2$ is the same as in the AR modeling case (see (159) and (160)). Thus, only the estimation of $\boldsymbol{\theta}_1$ need to be considered here. But the maximization of Q_1 directly over $\boldsymbol{B}$ becomes computationally unattractive. Instead, the space-alternating generalized EM (SAGE) algorithm [161] may provide a more feasible solution. The essence of the SAGE algorithm is that the specification of the complete data space, and, therefore, also the objective function, can be different for different parameters of the vector $\boldsymbol{\theta}_1$.

While the objective function Q_1 defined in (278) can be used for estimating σ_v^2 and the resulting estimator is a trivial modification of (161), given as

$$\begin{aligned} \widehat{\sigma_v^2}^{(i)} = \frac{1}{K}\sum_{k=1}^{K}\Big(|r(k)|^2 - 2\Re\big\{r^*(k)\mathbf{s}^T(k)\hat{\boldsymbol{B}}^{(i)}\hat{\boldsymbol{f}}^{(i)}(k|K)\big\} \\ + \mathbf{s}^T(k)\hat{\boldsymbol{B}}^{(i)}\boldsymbol{\Phi}_k^{(i)}\big(\hat{\boldsymbol{B}}^{(i)}\big)^H\mathbf{s}^*(k)\Big), \end{aligned} \tag{279}$$

the implementation of the M-step for the matrix $\boldsymbol{B}$ can be made easier by augmenting the observed data vector more than was done in (278) (by increasing the dimensionality of the missing data space). Specifically, similarly to (136), the received signal is decomposed into independent multipath components as follows:

$$r(k) = \sum_{l=1}^{L+1} s(k-l+1)\mathbf{b}_l^T\boldsymbol{f}(k) + \ddot{v}_l(k) = \sum_{l=1}^{L+1} x_l(k), \tag{280}$$

where $\mathbf{b}_l$ denotes the lth column of $\boldsymbol{B}^T$ and $\ddot{v}_l(k)$ denotes the zero-mean white Gaussian noise component whose correlation properties are given as

$$\frac{1}{2}E[\ddot{v}_i(k)\ddot{v}_j(k)] = \begin{cases} \frac{1}{L+1}\sigma_v^2, & i=j \\ 0, & i\neq j. \end{cases} \tag{281}$$

The objective function Q_1 will now be redefined for the purpose of estimating $\boldsymbol{B}$ as

$$Q_1\Big(\boldsymbol{\theta}_1\big|\hat{\boldsymbol{\Theta}}^{(i-1)}\Big) = E\Big[\ln p(\mathcal{X}|\mathbf{s},\boldsymbol{\theta}_1,\mathbf{F})\big|\mathbf{r},\mathbf{s},\hat{\boldsymbol{\Theta}}^{(i-1)}\Big], \tag{282}$$

where the complete data set is given as $\mathcal{X} = \big\{\{x_l(k)\}_{l=1,\ldots,L+1}^{k=1,\ldots,K}\big\}$. The incomplete data set is in this case given as $\mathcal{Y} = \{\mathbf{r}\}$, and there exists a many-to-one mapping from $\mathcal{X}$ to $\mathcal{Y}$ given by (280). Expanding (282) and maximizing it over $\boldsymbol{B}$ yields

$$\hat{\boldsymbol{B}}^{(i)} = \hat{\boldsymbol{B}}^{(i-1)} + \frac{1}{L+1}\sum_{k=1}^{K}\Big(r(k)\mathbf{s}^*(k)\big(\hat{\boldsymbol{f}}^{(i)}(k|K)\big)^H - \mathbf{s}^*(k)\mathbf{s}^T(k)\hat{\boldsymbol{B}}^{(i-1)}\boldsymbol{\Phi}_k^{(i)}\Big)\Big(\sum_{k=1}^{K}\boldsymbol{\Phi}_k^{(i)}\Big)^{-1}. \tag{283}$$

A common feature, however, is that by further augmenting the data vector, the convergence speed of the resulting EM algorithm will become slower [161].

The operation of the above ARMA parameter estimator is illustrated with a simple example, where a frequency selective channel with three independently fading, equal power channel taps is employed. In Fig. 41, the estimated scattering function is overlaid with the true scattering function when the ARMA(2,2) parametric channel model was assumed and the length of the known data record was $K = 400$. The estimated ARMA parameters were obtained via the SAGE algorithm (159),(160),(279),(283), and the estimates were averaged over 100 independent realizations. In the simulations, a correct initialization for the SAGE algorithm was assumed.

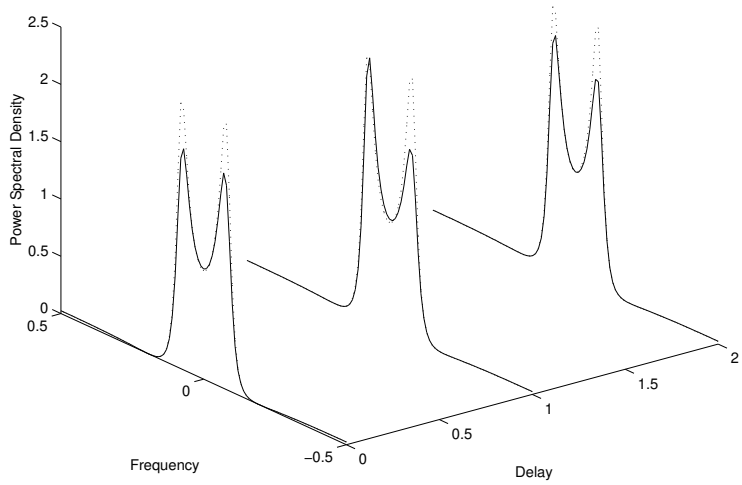

Fig 41. Scattering function estimates via the SAGE algorithm. The true power spectral density is given by the dotted lines.

A recursive estimator for the AR parameters is given in (176). Similarly, a recursive estimator for the MA parameters can be obtained from (174) by applying the *stochastic gradient approximation* technique. Specifically, when using the objective function (282), a recursive estimator with the tracking capability for the matrix $\boldsymbol{B}$ can be expressed as

$$\hat{\boldsymbol{B}}(k) = \hat{\boldsymbol{B}}(k-1) + \lambda_B \Big(r(k)\mathbf{s}^*(k)\hat{\boldsymbol{f}}^H(k|k) - \mathbf{s}^*(k)\mathbf{s}^T(k)\hat{\boldsymbol{B}}(k-1)\boldsymbol{\Phi}_{k|k} \Big), \quad (284)$$

where λ_B is a step-size parameter. When $p > q$, matrices $\boldsymbol{\Phi}_{k-1|k}$ and $\boldsymbol{\Psi}_{k|k}$ (needed for the recursive AR estimator) are obtained from the one-step Kalman smoother, whereas the matrix $\boldsymbol{\Phi}_{k|k}$ is obtained from the Kalman filtering equations. When $p \leq q$, it can be easily shown that the specific structure of the model matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ allows to compute the matrices $\boldsymbol{\Phi}_{k-1|k}$ and $\boldsymbol{\Psi}_{k|k}$ by the standard Kalman filtering as well. Therefore, in this specific case, the recursive estimators for the AR and MA parameters are effectively cross-coupled with the standard Kalman filter and no matrix inversions are needed at any phase of the estimation process.

Appendix 4 A proof of Lemma 4 of Chapter 6

Proof of Lemma 4:

Proof. Using (221) and suppressing the iteration indices, we obtain that

$$
\begin{aligned}
&\int_0^{2\pi} q_\phi(\phi)\ln p(\mathbf{r}|\mathbf{s},\phi)d\phi \\
&= -\frac{1}{\sigma_v^2}\int_0^{2\pi} p(\phi|\mathbf{r},\bar{\mathbf{s}})\Big(\big\|\mathbf{r}-\mathbf{s}e^{j\phi}\big\|^2\Big)d\phi + C \\
&= -\frac{1}{\sigma_v^2}\big(\mathbf{r}^H\mathbf{r}+\mathbf{s}^H\mathbf{s}\big) + A\,\frac{2}{\sigma_v^2}\,\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{2}{\sigma_v^2}\Re\left\{\mathbf{r}^H\bar{\mathbf{s}}e^{j\phi}\right\}}\Re\big\{\mathbf{r}^H\mathbf{s}e^{j\phi}\big\}d\phi + C \\
&= -\frac{1}{\sigma_v^2}\big(\mathbf{r}^H\mathbf{r}+\mathbf{s}^H\mathbf{s}\big) + A\,\frac{2}{\sigma_v^2}\,\frac{1}{2\pi} \\
&\quad\times\int_0^{2\pi} e^{\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\cos(\arg(\mathbf{r}^H\bar{\mathbf{s}})+\phi)}\big|\mathbf{r}^H\mathbf{s}\big|\cos\big(\arg(\mathbf{r}^H\mathbf{s})+\phi\big)d\phi + C \\
&= -\frac{1}{\sigma_v^2}\big(\mathbf{r}^H\mathbf{r}+\mathbf{s}^H\mathbf{s}\big) + A\,\frac{2}{\sigma_v^2}\,\frac{1}{2\pi} \\
&\quad\times\int_0^{2\pi} e^{\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\cos(\phi)}\big|\mathbf{r}^H\mathbf{s}\big|\cos\big(\arg(\mathbf{r}^H\mathbf{s})-\arg(\mathbf{r}^H\bar{\mathbf{s}})+\phi\big)d\phi + C \\
&= -\frac{1}{\sigma_v^2}\big(\mathbf{r}^H\mathbf{r}+\mathbf{s}^H\mathbf{s}\big) + A\,\frac{2}{\sigma_v^2}\,\big|\mathbf{r}^H\mathbf{s}\big|\cos\big(\arg(\mathbf{r}^H\mathbf{s})-\arg(\mathbf{r}^H\bar{\mathbf{s}})\big) \\
&\quad\times\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\cos(\phi)}\cos(\phi)d\phi - A\,\frac{2}{\sigma_v^2}\,\big|\mathbf{r}^H\mathbf{s}\big|\sin\big(\arg(\mathbf{r}^H\mathbf{s})-\arg(\mathbf{r}^H\bar{\mathbf{s}})\big) \\
&\quad\times\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\cos(\phi)}\sin(\phi)d\phi + C,
\end{aligned}
\tag{285}
$$

where C is a constant independent of ϕ and $\mathbf{s}$, and

$$
A = \frac{1}{p(\mathbf{r}|\bar{\mathbf{s}})}\,\frac{1}{(\pi\sigma_v^2)^K}e^{-\frac{1}{\sigma_v^2}(\mathbf{r}^H\mathbf{r}+\bar{\mathbf{s}}^H\bar{\mathbf{s}})} \equiv \frac{1}{I_0\big(\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\big)}. \tag{286}
$$

Inserting (286) into (285) and using the fact that

$$
\int_0^{2\pi} e^{\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\cos(\phi)}\sin(\phi)d\phi = 0 \tag{287}
$$

yields

$$
\int_0^{2\pi} q_\phi(\phi)\ln p(\mathbf{r}|\mathbf{s},\phi)d\phi = -\frac{1}{\sigma_v^2}\Big(\mathbf{r}^H\mathbf{r} - 2\zeta\,\Re\big\{\mathbf{r}^H\mathbf{s}\,e^{-j\arg(\mathbf{r}^H\bar{\mathbf{s}})}\big\} + \mathbf{s}^H\mathbf{s}\Big),
$$

where $\zeta = \frac{I_1\left(\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\right)}{I_0\left(\frac{2}{\sigma_v^2}|\mathbf{r}^H\bar{\mathbf{s}}|\right)}$. □

Book orders:
OULU UNIVERSITY PRESS
P.O. Box 8200, FI-90014
University of Oulu, Finland

Distributed by
OULU UNIVERSITY LIBRARY
P.O. Box 7500, FI-90014
University of Oulu, Finland

UNIVERSITY OF OULU P.O. Box 7500 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-951-42-8685-8 (Paperback)
ISBN 978-951-42-8686-5 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)