Synchronisatie in digitale communicatiesystemen: performantiegrenzen en praktische algoritmes

Synchronization in Digital Communication Systems: Performance Bounds and Practical Algorithms

Nele Noels

Promotoren: prof. dr. ir. M. Moeneclaey, prof. dr. ir. H. Steendam
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Elektrotechniek

Vakgroep Telecommunicatie en Informatieverwerking
Voorzitter: prof. dr. ir. H. Bruneel
Faculteit Ingenieurswetenschappen
Academiejaar 2008 - 2009

UNIVERSITEIT
GENT

ISBN 978-90-8578-293-3
NUR 959
Wettelijk depot: D/2009/10.500/51

Acknowledgement

I would like to dedicate this book to my three children, Anoek, Mathias and Dorien, who entered this world during the creation of my thesis. A warm hug goes to my husband Herman, whose love helped me to stay focussed over the past eight years.

On the professional level I would like to gratefully acknowledge the support of my supervisors, Prof. Marc Moeneclaey and Prof. Heidi Steendam. I would particularly like to thank Marc for his constant encouragement, inspirational guidance and keen interest in this study. Many thanks also to my present and former colleagues at the TELIN research department of the Ghent University, for their kind help and company every day. Last but not least, I would like to thank all the persons who are working at other universities in Belgium and abroad, and who were either directly or indirectly involved with my research.

Nele Noels - March 2009

Contents

Abbreviations

ACRB	Asymptotic Cramer-Rao Bound
APP	A Posteriori Probability
AWGN	Additive White Gaussian Noise
BCJR	Bahl, Cocke, Jelinek and Raviv
BB	BaseBand
BER	Bit Error Rate
BICM	Bit Interleaved Coded Modulation
BP	BandPass
CATV	Community Antenna Television
co	correct
cond	condition
CRB	Cramer-Rao Bound
DA	Data-Aided

dB	decibels
DD	Decision-Directed
DFT	Discrete Fourier Transform
EM	Expectation Maximization
FB	Feedback
FF	Feedforward
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix
HDD	Hard Decision-Directed
Hz	Hertz
i.i.d.	independent and identically distributed
ISI	Inter-Symbol-Interference
LDPC	Low-Density Parity-Check
MAP	Maximum A Posteriori
MAP-i	Maximum A Posteriori information-bit-by-information-bit
MAP-c	Maximum A Posteriori coded-bit-by-coded-bit
MCRB	Modified Cramer-Rao Bound
MFIM	Modified Fisher Information Matrix
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MVU	Minimum Variance Unbiased
NPA NCA	Non-Pilot-Aided Non-Code-Aided
PA CA	Pilot-Aided Code-Aided
PA NCA	Pilot-Aided Non-Code-Aided
PAM	Pulse Amplitude Modulation
pdf	power density function
PSK	Phase Shift Keying

QAM	Quadrature Amplitude Modulation
SDD	Soft Decision-Directed
SER	Symbol error rate
si	simplified
SNR	Signal-to-Noise Ratio
TCM	Trellis Coded Modulation

Symbols

#1	Assymetric burst structure
#2	Symmetric burst structure
$\frac{\partial}{\partial .}$	Partial derivative
$\rightarrow$	Going to
$\propto$	Equality up to an irrelevant scalin factor
!	Factorial operation
$\{\}^n$	n-Ary cartesian power of a set
$. \times .$	Cartesian product of the two sets or matrix dimension
$\star$	Convolution operator
$.^T$	Transpose
$.^*$	Complex conjugate
$.^H$	Conjugate Transpose

$\lfloor . \rfloor$	Floor function: maps a real number to the next smallest integer.
$.^{-1}$	Function or matrix inversion
$\lvert . \rvert$	Norm, complex amplitude or cardinality of a set
$\emptyset$	Empty set
$\in$	belongs to
$\subset$	subset of
$\mathbf{0}_{n \times m}$	All-zero matrix with n rows and m columns
$X(f)$	Fourier transform of a function $x(t)$
$\hat{x}$, $\hat{x}(\mathbf{r})$	Estimate of a parameter x
$x^{(n)}$	nth realization of a variable x
x_i	iTh component of a vector $\mathbf{x}$
$X_{i,j}$	(i,j)Th element of a matrix $\mathbf{X}$
$\sum_{\mathbf{x} \setminus x_i}$	Summation over all elements of a vector $\mathbf{x}$ accept x_i
$\mathcal{A}(.,.)$	Pilot symbol insertion rule
A_k	Actual value of the symbol $a(k)$
$\mathbf{a}$	Transmitted symbol sequence
$a(k)$	Symbol with index k in the burst
$a(k;\mathbf{b})$	kTh component of the symbol sequence that results from the information bit sequence $\mathbf{b}$ after applying channel encoding, bit-to-symbol mapping and pilot symbol insertion
$\mathbf{a}_d$	Data symbol sequence (indices of components take values in I_d)
B	(One-side) signal bandwidth
$\mathbf{b}$	Information bit sequence
$b(k)$	kTh information bit in the burst
C	Irrelevant constant
$\mathcal{C}(.)$	Encoding rule

CRB_i	Cramer-Rao bound related to the estimation of the ith useful parameter
CRB_ν	Cramer-Rao bound related to the estimation of ν
CRB_θ	Cramer-Rao bound related to the estimation of θ
$\mathbf{c}$	Coded bit sequence
$c(k)$	kTh coded bit in the burst
$\mathcal{D}(.)$	Mapping rule (on sequence level)
$\mathcal{D}_m(.)$	Bits-to-symbol mapping function (on symbol level)
$\mathbf{d}$	Data symbol sequence (indices of components range from 0 to N_d-1)
$d(k)$	kTh data symbol in the burst
d_M	Minimum Euclidean distance between the constellation points
$E_X[Y]$	Statistical expectation of Y with respect to the probability density function of X
E_s	Average symbol energy
E_b	Average energy per information bit
E_c	Average energy per coded bit
E_d	Average energy per data symbol
$\mathbf{e}(\mathbf{r})$	estimation error vector
$exp(.)$, $e^{\cdot}$	Exponential function
$e_\Theta(t)$	Error regarding the estimation of the instantaneous carrier phase shift
F	Carrier frequency offset
$F(x,y)$	Short-hand notation for the value of the exponential function at $\frac{E_s}{N_0}\left(2\Re\{x^*y\}-\lvert x\rvert^2\right)$
f_c	Carrier frequency
$f_{cod,i}(.,.)$	Code constraint exploited by a maximum a posteriori information-bit-by-information-bit detector
$f_{cod,c}(.,.)$	Code constraint exploited by a maximum a posteriori coded-bit-by-coded-bit detector
$f_{lh}(.)$	Likelihood of the kth data symbol as a function of the associated matched filter output sample

$f_{map}(.,.)$	Mapping constraint
$\mathbf{G}$	Generator matrix
$\mathbf{g}_i$	iTh generator sequence of convolutional code
$g(t)$	Nyquist pulse
$\mathbf{H}$	Parity check matrix
$\mathbf{I}$	Identity matrix
$\mathbf{I}_n$	Identity matrix of size n
$I[.]$	Indicator function
$\Im\{.\}$	Imaginary part of complex number
I_s	Set of symbol indices
I_p	Subset of pilot symbol indices
I_d	Subset of data symbol indices
$\mathbf{J}$	Fisher information matrix
$\mathbf{J}_d$	Data symbol contribution to the Fisher information matrix
$\mathbf{J}_p$	Pilot symbol contribution to the Fisher information matrix
$J_{\nu,\nu}$	First diagonal element of the Fisher information matrix related to the joint likelihood function of ν and θ
$J_{\theta,\theta}$	Second diagonal element of the Fisher information matrix related to the joint likelihood function of ν and θ
$J_{\nu,\theta}$	Off-diagonal elements of the Fisher information matrix related to the joint likelihood function of ν and θ
J_ν	Fisher information matrix related to the likelihood function of ν
J_θ	Fisher information matrix related to the likelihood function of θ
K	2π Divided by the symmetry angle of the constellation
k	Symbol or sample index
$L(\mathbf{X};\mathbf{Y})$	Likelihood function of the parameter vector $\mathbf{X}$, given the observation vector $\mathbf{Y}$
$\ell(\mathbf{X};\mathbf{Y})$	Log-likelihood function of the partameter vector $\mathbf{X}$, given the observation vector $\mathbf{Y}$

$\ell_i(\mathbf{X};.)$	Derivative of the log-likelihood function of the parameter vector $\mathbf{X}$ with respect to the ith component of $\mathbf{X}$
$\ell'(X;.)$	Derivative of the log-likelihood function of the scalar parameter X with respect to X
$\mathbf{M}$	Modified Fisher information matrix
M	Number of constellation points
$M_{\nu,\nu}$	First diagonal element of the modified Fisher information matrix related to the joint likelihood function of ν and θ
$M_{\theta,\theta}$	Second diagonal element of the modified Fisher information matrix related to the joint likelihood function of ν and θ
$M_{\nu,\theta}$	Off-diagonal elements of the modified Fisher information matrix related to the joint likelihood function of ν and θ
M_ν	Modified Fisher information matrix related to the likelihood function of ν
M_θ	Modified Fisher information matrix related to the likelihood function of θ
$MCRB_i$	Modified Cramer-Rao bound related to the estimation of the ith useful parameter
$MCRB_\nu$	Modified Cramer-Rao bound related to the estimation of ν
$MCRB_\theta$	Modified Cramer-Rao bound related to the estimation of θ
m	Number of bits per symbol
N	Number of parameters to be estimated, number of realizations to be generated
N_0	Noise power spectral density
N_b	Number of information bits per burst
N_c	Number of coded bits per burst
N_d	Number of data symbols per burst
N_p	Number of pilot symbols per burst
N_s	Number of symbols per burst
N_W	Weighted cardinality of the set of symbol indices
n	Symbol or sample index

n_M	Average number of constellation points at minimum distance
$\mathbf{P}$	Parity matrix
P	Predicate
$\mathcal{P}$	Average transmit power
$\Pr[.]$	Probability mass function of discrete-valued random variable
$\Pr[.\,\vert.]$	Conditional probability mass function of discrete-valued random variable
$\mathbf{p}$	Pilot symbol sequence
$p(.)$	Probability density function of continuous random variable
$p(.\,\vert.)$	Conditional probability density function of continuous random variable
$p(.;.)$	Parameterized probability density function of continuous random variable
$p(k)$	kTh pilot symbol in the burst
$Q(.)$	Q-function: area under the tail of the probability density function of a Gaussian random variable with zero mean and unit variance
$\mathbf{R}^e$	Error correlation matrix
$\Re\{.\}$	Real part of complex number
R_b	Information bit rate
R_c	Coded bit rate
R_d	Data symbol rate
R_s	Symbol rate
$\mathbf{r}$	Vector representation of the received baseband signal
$(\mathbf{r})_k$	kTh component of the observation vector
$\mathbf{r}^{(co)}$	Obervation vector resulting from the correct model
$\mathbf{r}^{(si)}$	Observation vector resulting from the simplified model
$r_{BB}(t)$	Received baseband signal
$r_{BP}(t)$	Received bandpass signal
S	Number of data symbols seperating two blocks of pilotsymbols

$\mathbf{S}(.)$	Matrix describing the linear transformation from the transmitted symbol sequence to the useful part of the observation vector
$\mathbf{s}(.,.)$	Useful part of the observation vector
$(\mathbf{s}(.,.))_k$	kTh component of the useful part of the observation vector
$s_{BB}(.)$	Transmitted baseband signal
$s_{BP}(.;.)$	Tranmitted bandpass signal
$\tilde{s}_{BP}(.;.,.)$	Useful part of received bandpass signal
T	Symbol period
T_{sample}	Nyquist sample period
t	Continuous time variable
t_0	Time instant at which the instantaneous phase shift is evaluated
$\mathbf{u}$	Useful parameter vector
$\mathcal{U}_i$	Domain of ith component of the useful parameter vector
$\mathbf{v}$	Nuisance parameter vector
$W(k)$	Weight function defined on the set of symbol indices
$\mathbf{w}$	Vector of complex-valued zero-mean additive white Gaussian noise samples
$(\mathbf{w})_k$	kTh component of the vector of complex-valued zero-mean additive white Gaussian noise samples
$w_{BB}(t)$	Complex-valued zero-mean additive white Gaussian noise
$w_{BP}(t)$	Real-valued zero-mean additive white Gaussian noise
$\mathbf{z}(.)$	Matched filter output sample sequence
$\mathbf{z}_i(.,.)$	Derivative of the matched filter output sample sequence with respect to ith component of the carrier synchronization parameter vector
$\mathbf{z}^{(co)}(.)$	Matched filter output sample sequence resulting from the correct observation model
$\mathbf{z}^{(si)}(.)$	Matched filter output sample sequence resulting from the simplified observation model
$z(k;.)$	Matched filter output sample with index k
$z^{(co)}(k;.)$	Matched filter output sample with index k resulting from the correct observation model

$z^{(si)}(k;.)$	Matched filter output sample with index k resulting from the simplified observation model
$z_i(k;.,.)$	Derivative of the matched filter output sample with index k with respect to ith component of the carrier synchronization parameter vector
$\Theta(t;.)$	Instantaneous carrier phase shift of the received signal vis-a-vis the receiver's local reference carrier
$\mathbf{\Psi}$	Carrier synchronization parameter vector
Ω	Signaling constellation
Ω_M	M-point signaling constellation
$\Omega_{M-\mathrm{PAM}}$	M-PAM constellation
$\Omega_{M-\mathrm{PSK}}$	M-PSK constellation
$\Omega_{M-\mathrm{QAM}}$	M-QAM constellation
$\gamma_p(.)$	Bijection from $\{0,1,\ldots,N_p-1\}$ to I_p that indicates which of the transmitted symbols are pilot symbols
$\gamma_p(.)$	Bijection from $\{0,1,\ldots,N_d-1\}$ to I_d that indicates which of the transmitted symbols are data symbols
δ_n	Kronecker delta
$\delta(.)$	Dirac impulse function
$\delta J_{\nu,\nu}$	Part of the first diagonal element of the Fisher information matrix related to the joint likelihood function of ν and θ that only occurs for the correct observation model
ϵ_{Sp}	Spacing ratio
$\zeta_{\mathcal{C}}$	Set of legitimate coded sequences
$\zeta_{\mathcal{D}}$	Set of legitimate data symbol sequences
θ	Phase shift at an arbitrary time instant
κ	Normalized time variable
κ_0	Normalized time instant at which the instantaneous phase shift is evaluated
$\kappa_{m,p}$	Mean of the symbol index over the subset of pilot symbol indices

$\kappa_{m,s}$	Mean of the symbol index over the set of indices
$\kappa_{m,W}$	Weighted average of the symbol index over the set of symbol indices
λ_p	Pilot symbol ratio
$\boldsymbol{\mu}(.)$	A posteriori average of the symbol sequence
$\mu(k;.)$	A posteriori average of the kth symbol
ν	Normalized frequency offset
ν_{max}	Maximum absolute value of the frequency offset
ρ	Code rate
σ^2	Statistical variance
σ_p^2	Variance of the symbol index over the subset op pilot symbol indices
σ_W^2	Weighted variance of the symbol index over the set of symbol indices
ω_i	iTh constellation point

Samenvatting

Communicatiekanalen moeten vaak signalen transporteren die afkomstig zijn van verschillende zenders. Zo wordt het elektromagnetisch spectrum bijvoorbeeld gebruikt voor allerlei verschillende toepassingen, gaande van draadloze gegevensoverdracht over radio- en televisiedistributie tot mobiele telefonie. Een van de standaardmethodes om interferentie tussen de afzonderlijke signalen te vermijden, is *draaggolfcommunicatie*. Het beschikbare frequentiespectrum wordt opgesplitst in een aantal frequentiebanden die elkaar niet overlappen, de zogenaamde *draaggolfkanalen*, en elke zender krijgt een verschillend draaggolfkanaal toegewezen. Om een signaal over een draaggolfkanaal te versturen, moet de frequentie-inhoud ervan gewoonlijk wel eerst worden verschoven naar een andere frequentieband die compatibel is met het betreffende draaggolfkanaal. Deze techniek, waarbij een signaal dus naar een andere frequentieband wordt gebracht, heet *modulatie*. Aan de ontvanger wordt het gemoduleerde signaal opnieuw verschoven naar de oorspronkelijke frequentieband (*gedemoduleerd*) met als doel het oorspronkelijke signaal te reconstrueren. Zowel het modulatieproces aan de zijde van de zender als het demodulatieproces aan de zijde van de ontvanger maakt gebruik van een lokaal gegenereerde sinusoïde (de *draaggolf*) waarvan de nominale frequentie overeenstemt met het centrum van het draaggolfkanaal. Om een betrouwbare reconstructie van het oorspronkelijke signaal mogelijk te maken, moet de draaggolf aan de ontvanger bovendien nauwkeurig

gesynchroniseerd worden met deze aan de zender: we spreken over *draaggolfsynchronisatie*. Daartoe heeft de ontvanger nood aan een *schatting* van een aantal *draaggolfsynchronisatieparameters,* zoals de frequentie en de fase van de draaggolf aan de zender.

Tijdens het transport over het communicatiekanaal wordt het verzonden signaal verstoord door allerlei ongunstige effecten die we waarnemen als *ruis*. Een belangrijke maat voor de signaalkwaliteit aan de ontvanger is de *signaal-ruisverhouding* (signal-to-noise ratio of SNR). De SNR is de vermogensverhouding tussen de nuttige informatie in het ontvangen signaal en de achtergrondruis. De aanwezigheid van ruis in het ontvangen signaal beperkt de mate waarin de verstuurde informatie (correct) kan worden gereconstrueerd uit het ontvangen gedemoduleerde signaal. In het geval van digitale communicatie bestaat deze informatie uit een sequentie van binaire getallen (binary digits of bits). Een goede maat voor de performantie van dergelijke systemen is dan ook de *bitfoutprobabiliteit* (bit error rate of BER). Dit is de waarschijnlijkheid dat een verzonden bit foutief wordt gereconstrueerd (we spreken over *bitdetectie*) aan de ontvanger. Hoe lager de BER, hoe betrouwbaarder de communicatie. Om een betrouwbaar transport van digitale informatie over een onbetrouwbaar kanaal mogelijk te maken, wordt *kanaalcodering* toegepast. Hierbij wordt op een gestructureerde manier redundantie toegevoegd aan de verzonden bitsequentie, waardoor foutdetectie en -correctie mogelijk wordt. De laatste tien jaar werden verschillende, uiterst krachtige en *iteratief decodeerbare* kanaalcoderingstechnieken ontworpen die in staat zijn een zeer lage BER te realiseren bij een zeer lage SNR (voorbeelden zijn onder andere turbo-codering, low-density parity-check (LDPC) codering en bit-interleaved-coded-modulation (BICM)). Een neveneffect van deze nieuwe ontwikkelingen is evenwel dat de draaggolfsynchronisatie nu moet gebeuren bij een SNR die lager is dan ooit tevoren. Dit stelt vanzelfsprekend hoge eisen aan het ontwerp van de draaggolfsynchronisatiestructuur en gaf ook een nieuwe impuls aan het onderzoek naar de fundamentele grenzen aan de nauwkeurigheid die met een dergelijke draaggolfsynchronisatiestructuur kan worden gerealiseerd.

Deze doctoraatsthesis onderzoekt performantiegrenzen en praktische algoritmen voor draaggolfsynchronisatie in digitale draaggolfcommunicatiesystemen. De klemtoon van het onderzoek ligt op signalen met een lage SNR. In tegenstelling tot de conventionele aanpak in de literatuur, vooronderstellen we niet dat de verstuurde informatiebits onderling onafhankelijk en gelijk verdeeld zijn. Dit laat ons toe om rekening te houden met de structuur die door de kanaalcodering wordt opgelegd aan de verstuurde bitsequentie. In plaats daarvan beschouwen we vier verschillende schattingsmodi: de pilot-aided code-aided modus, de data-aided modus, de pilot-aided non-code-aided modus en de non-pilot-aided non-code-aided modus. Elke schattingsmodus correspondeert met een andere (benaderende) hypothese over de massafunctie van de verzonden bitsequentie.

De eerste vier hoofdstukken van de tekst fungeren als basis voor hoofdstuk 5 en 6, waarin de fundamentele en originele bijdragen van mijn onderzoek worden beschreven. Na een volledige beschrijving van het bestudeerde communicatiesysteem, volgt er een theoretische bespreking van de detectie- en schattingsproblemen die zich voordoen in de ontvanger. Het concept van *gesynchroniseerde, maximum-a-posteriori (MAP)* bitdetectie wordt gehanteerd. Dit betekent dat de draaggolfsynchronisatieparameters worden geschat en gebruikt alsof het de werkelijke waarden betrof bij het bepalen van de informatiebits die de marginale a-posteriori bitwaarschijnlijkheid (a posteriori probability of APP) maximaliseren. Voor wat de schatting van de draaggolfsynchronisatieparameters betreft, zetten we vooral het belang van de *Cramer-Rao grens* (Cramer-Rao bound of CRB) en de *maximum-likelihood* (ML) schatter in de verf. In verband met de MAP bitdetectie besteden we hoofdzakelijk aandacht het efficiënt berekenen van de noodzakelijke marginale bit APPs. De techniek die daarvoor tegenwoordig meest wordt gebruikt, is het *som-product algoritme*. Vanuit het standpunt van de draaggolfsynchronisatie heeft dit algoritme de bijzonder interessante eigenschap dat, samen met de vereiste marginale bit APPs, ook de marginale symbool APPs (als bijproducten) worden berekend. Deze laatste APPs spelen een belangrijke rol bij het afleiden van CRBs en ML-gebaseerde schatters die rekening houden met de kanaalcodering.

Een eerste, belangrijk deel van het voorgestelde onderzoek heeft betrekking op het berekenen van werkelijke en gewijzigde CRBs voor het schatten van de draaggolfsynchronisatieparameters. We bekijken twee verschillende observatiemodellen: het correct continue-tijd model en een vereenvoudigd discretetijd model van het ontvangen signaal. Het laatste model wordt het meest gebruikt maar negeert wel een aantal gevolgen van een verschil in draaggolffrequentie tussen zender en ontvanger, zoals de onderdrukking van het nuttige signaal en de aanwezigheid van inter-symbool-interferentie. In het algemeen kunnen de gewijzigde CRBs gemakkelijk analytisch berekend worden. Het berekenen van de werkelijke CRBs is echter veel minder eenvoudig en kan slechts in een paar gevallen expliciet gebeuren. Bovendien geeft het gebruik van brute rekenkracht om het probleem op te lossen met een computer (zonder gebruik te maken van algoritmen en heuristieken om de berekeningen te versnellen) aanleiding tot een rekentijd die exponentieel toeneemt met het aantal verzonden bits. De uitdaging bestaat er dan ook in om een efficiëntere, gedeeltelijk analytische en gedeeltelijk numerieke, methode te ontwikkelen voor het bepalen van de CRBs. We tonen aan dat dit mogelijk is door rekening te houden met de specifieke structuur van het ontvangen signaal. Het belangrijkste resultaat is een algemene procedure om CRBs te berekenen met een rekencomplexiteit die lineair (in plaats van exponentieel) toeneemt met het aantal verzonden bits. Voor bepaalde schattingsmodi is een verdere vereenvoudiging van deze procedure mogelijk. Aan de hand van numerieke resultaten, verkregen volgens de voorgestelde procedure(s), wordt het effect van het observatiemodel en van de schattingsmodus op de CRBs besproken.

Een tweede, belangrijk deel van mijn doctoraatsonderzoek heeft betrekking op het ontwerp van een efficiënte draaggolfsynchronisatiestructuur voor ontvangers met een MAP bit detector die werkt volgens het som-product algoritme (zoals het geval is bij turbo-codes, LDPC codes en BICM). We bestuderen voornamelijk ML-gebaseerde algoritmes. Aan de hand van computersimulaties, vergelijken we hun performantie met de eerder verkregen CRBs. Het belangrijkste resultaat is een algemeen theoretisch raamwerk voor iteratieve draaggolfsynchronisatiestructuren die informatie uitwisselen tussen een MAP bit detector en de draaggolfsynchronisatieparameterschatter. Dit raamwerk gebruiken we als basis voor het ontwikkelen van een nieuw algoritme voor draaggolfsynchronisatie dat een goed compromis oplevert tussen nauwkeurigheid en rekencomplexiteit. Met behulp van computersimulaties tonen we aan dat, in het normale SNR werkingsgebied van communicatiesystemen die gebruik maken van een krachtige kanaalcodering, dit nieuwe algoritme in staat is beter te presteren dan de al bestaande methodes.

Het laatste hoofdstuk van de thesis vat de belangrijkste resultaten samen. Tenslotte, formuleer ik enkele opmerkingen en doe ik een aantal suggesties voor verder onderzoek.

Het gepresenteerde onderzoek resulteerde tot dusver in zes tijdschriftartikels [1–6] and veertien bijdragen op internationale wetenschappelijke conferenties [7–20]. Elf andere artikels (vier tijdschriftartikels and zeven conferentieartikels) [21–31] hebben betrekking op gerelateerd onderzoek dat buiten het kader van deze thesis valt.

Summary

Communication channels must often transfer signals from different transmitters. An obvious example is the use of electromagnetic waves, which carry a range of signals going from wireless data traffic over radio and television signals to cellular telephony. One of the standard methods to avoid interference between the signals of these separate transmitters is *bandpass communication*. The available frequency spectrum is divided into non-overlapping frequency bands (referred to as *bandpass channels*) and each transmitter is assigned to a different bandpass channel. The transmission of an information-bearing signal over a bandpass communication channel usually requires a shift of the range of frequencies contained in that signal to a frequency range that is compatible with the designated frequency band. This translation of the signal frequency range is referred to as *modulation*. At the receiver's end, the modulated signal is *demodulated* (i.e., frequency shifted back to the original frequency band) in order to recover the original signal. The modulation/demodulation process requires the presence of a locally generated sinusoidal *carrier signal* at both the transmitter's and the receiver's end of the bandpass communication system. To enable a reliable information transfer, it is also imperative that these two carrier signals are accurately *synchronized*. For this synchronization, the receiver needs an *estimate* of a number of carrier synchronization *parameters*, such as the frequency and the phase of the carrier signal at the transmitter.

While traveling over the communication channel, the transmitted signal is corrupted by a variety of possible mechanisms that are commonly categorized as *noise*. An important measure for the signal quality at the receiver is the *signal-to-noise ratio* (SNR). This is an engineering term for the power ratio between the meaningful information in the received signal and the background noise. Noise limits the ability to correctly recover the sent information from the received demodulated signal. In the case of *digital communication*, this information consists of a sequence of binary digits (bits). A quantitative criterion to measure the performance of such systems is therefore the *bit error rate* (BER). This is the probability that a bit is recovered (*detected*) erroneously at the receiver. The lower the BER, the more reliable the communication will be. Channel coding is applied in digital communication systems to better protect the transmitted information against the detrimental effects of the noise. The channel encoder introduces structured redundancy in the transmitted bit sequences. This makes it possible to detect and correct some of the occurred bit errors at the receiver. The last decade has seen the development of *iteratively decodable* channel codes of unprecedented power, with large coding gains that are capable of achieving a low BER at a very low SNR (examples are turbo codes, low-density parity-check (LDPC) codes and bit-interleaved coded-modulation (BICM)). A by-product of this development is that the carrier synchronization must now be performed at a SNR that is lower than ever before. Of course, this imposes high requirements on the carrier synchronizer design, but it also provided a new impulse to the research on fundamental bounds on the attainable performance of these carrier synchronizers.

This doctoral thesis investigates performance bounds and practical algorithms pertaining to carrier synchronization for digital bandpass communication systems. The focus lies on signals with a low SNR. As opposed to the conventional approach in the literature, we do not a-priori assume that the transmitted bits are independent and identically distributed. This allows us to take into account the structure that the channel code enforces upon the transmitted bit sequence. Instead, four different estimation modes are considered: the pilot-aided code-aided mode, the data-aided mode, the pilot-aided non-code-aided mode and the non-pilot-aided non-code-aided mode. Each one of these estimation modes corresponds to a different (approximate) hypothesis about the probability mass function of the transmitted bit sequence.

The first four chapters of the text serve as the basis for chapters 5 and 6, which contain the essential and original research contributions. After a comprehensive description of the digital bandpass communication system that is being considered, we provide a theoretical discussion of the detection and estimation problems encountered in the receiver. We adopt the concept of *synchronized maximum-a-posteriori* (MAP) bit detection, which involves estimating the carrier synchronization parameters and using these estimates as if they were the true values in order to determine the information bits that maximize

the marginal bit a posteriori probabilities (APP). As far as parameter estimation is concerned, we primarily highlight the importance of the *Cramer-Rao bound* (CRB) and the *maximum-likelihood* (ML) estimator. With regard to MAP bit detection, we specifically dig into the efficient computation of the required marginal bit APPs. The most frequently used method for this purpose is the *sum-product algorithm*. From a carrier synchronization perspective, this algorithm has the very interesting property of generating, together with the required marginal bit APPs, also (as by-products) the marginal symbol APPs. The latter APPs play an important role in the derivation of CRBs and ML-based estimation algorithms that take into account the channel code.

A first important part of the presented research is related to deriving true and modified Bayesian CRBs pertaining to carrier synchronization parameter estimation. Two different observation models are considered: the *correct* continuous-time model and a *simplified* discrete-time model of the received signal. The latter is commonly used but ignores the reduction of the useful signal and the occurrence of inter-symbol-interference caused by a difference in carrier frequency between the receiver and the transmitter. The computation of the MCRBs is generally straightforward and easily done analytically. The computation of the corresponding true CRBs is considerably more complicated and a complete *analytical* evaluation of CRBs is often infeasible. On the other hand, a brute-force *numerical* evaluation of the CRBs requires a computation time that increases exponentially with the number of transmitted bits. The challenge is thus to develop a more efficient, semi-analytical and semi-numerical, method to determine the CRBs. We show that this can be done by taking into account the specific structure of the received signal. The main result is the derivation of a general procedure for evaluating the CRBs with a complexity that is linear (and not exponential) in the number bits transmitted. Certain estimation modes still allow a further simplification of this procedure. Numerical results, generated according to the proposed procedure(s), are used to discuss the influence of the observation model and the estimation mode on the CRBs.

A second important part of my doctoral research is related to the design of the carrier synchronization section of receivers with a MAP bit detector operating according to the sum-product algorithm (as with turbo codes, LDPC codes and BICM). We mainly study algorithms that are very close approximations of the true ML estimator. We present simulation results pertaining to their performance, and compare this performance to the associated CRBs derived in the previous part of the thesis. The main result is a new theoretical framework for iterative synchronization structures that exchange information between the MAP bit detector and the synchronization parameter estimator. From this framework, we derive a new carrier synchronization algorithm with a good performance complexity trade-off. This new algorithm is shown to be capable of outperforming the conventional carrier synchronization algorithms at the normal operating SNR of systems with powerful channel codes.

Finally, the last chapter of this thesis summarizes the most important results. On top of that, I also formulate some remarks and I indicate a number of directions for future research.

The presented research has so far resulted in six journal articles [1–6] and fourteen contributions in international conference proceedings [7–20]. Apart from these, eleven published papers (four journal papers and seven conference papers) [21–31] provide additional findings in relation to research that falls outside the scope of this thesis.

1 Introduction

This doctoral thesis focuses on carrier synchronization for digital bandpass communication systems. More specifically, we have studied the estimation of the carrier synchronization parameters at the receiver (frequency and phase) and we have devoted special attention to communication channels that introduce a considerable level of noise when compared to the useful signal. The objective of our work was twofold. First, we wanted to evaluate the fundamental bounds on the performance of carrier parameter estimators. Secondly, we wanted to develop accurate carrier parameter estimation methods to approach these bounds.

The introductory chapter consists of two parts. Following this brief abstract, the next section will provide the reader with essential background information and also explains our motivation for the conducted research. An extended outline of this doctoral thesis can be found in section 1.2.

1.1 Background and Motivation

Any *communication system* can be viewed as a link between a *source* and a *sink*. Information is sent from the source and received at the sink. The communication system block diagram is shown in Fig. 1.1. The *transmitter* takes the

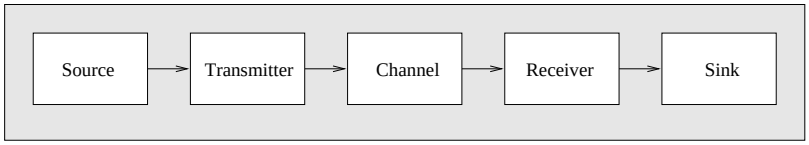

Figure 1.1: Key elements of a communication system.

information from the source and converts it into a form that is suitable for being transferred over the channel. The *channel* is the physical medium over which the actual transmission takes place. The *receiver* then captures the channel output and tries to recover the information that was sent by the source. Subsequently, an estimate of this information is passed to the sink as the received information.

Whatever the nature of the channel, the receiver essentially has to cope with the fact that the message (or information-bearing) signal is corrupted by a variety of possible mechanisms, commonly categorized as *noise*. Noise limits the ability to correctly identify the sent message signal and therefore limits information transfer. An important measure for the signal quality is the *signal-to-noise ratio* (SNR), which is an engineering term that expresses the power ratio between the meaningful information in the received signal and the background noise.

In the case of *digital communication* the information that is generated by the source is a sequence of binary digits (bits) that can either take the value "0" or "1". The probability that a bit will be received in error at the destination is called the *bit error rate* (BER). It provides a quantitative criterion to measure the performance of the system. The lower the BER, the more reliable the communication will be.

Channel coding is often used in digital communication systems to increase the reliability of the communication over the channel. Adding controlled redundancy to the information bit stream makes it possible to detect and correct some of the occurred bit errors. Of course some channel codes will be more effective than others and there will always remain error patterns which we cannot correct.

Communication channels often accommodate different signals from multiple transmitters. An obvious example are electromagnetic waves, which carry a broad variety of signals going from wireless data traffic over radio and television signals to cellular telephony. If the same medium is being used to transfer different signals, it is essential to avoid interference between the transmissions of these separate users. This can be done by resorting to *bandpass communication*. The available frequency spectrum is then divided into non-overlapping *frequency bands* and each transmitter is assigned to a different frequency band (each of which is specified by its own center frequency and bandwidth). It is crucial that each transmitter strictly observes the limits of its associated frequency band

since any transgression would create interference in the neighboring bands.

The transmission of a message signal over a bandpass communication channel usually requires a shift of the range of frequencies contained in the message signal to another frequency range that is compatible with the designated frequency band. This translation of the signal frequency range is accomplished by means of *carrier modulation*. Carrier modulation is defined as the process by which some characteristic of a carrier signal is varied in accordance with a modulating signal. In this document the message signal is referred to as the modulating signal or *baseband* signal, whereas the result of the modulation is referred to as the modulated signal or *bandpass* signal. The carrier signal is typically a sinusoid whose oscillation frequency and phase are respectively referred to as the *carrier frequency* and the *carrier phase*. At the receiver end, the modulated signal is demodulated (or frequency shifted from passband to baseband) to recover the original message signal. This demodulation requires the presence of a sinusoidal carrier signal generated at the receiver.

To enable a reliable detection of the transmitted information, it is imperative that the carrier signals at the transmitter and the receiver have almost exactly the same frequency and phase. However, as the carrier oscillators at the transmitter and receiver are operating independently, their frequency and phase are not the same, and the demodulation at the receiver is performed using erroneous carrier parameters. In order to cope with this problem the receiver is fitted with a *carrier synchronization* unit which has to perform two types of tasks. Firstly it has to *estimate* the *frequency offset* and the *phase shift* of the received signal vis-à-vis the local reference carrier. Once these carrier synchronization parameters have been estimated, the second task of the synchronization unit is to *correct* (rotate) the demodulated signal to compensate for these parameters. The corrected signal is then being processed further to recover the transmitted information.

The problem of estimating carrier synchronization parameters is present in all digital bandpass communication systems. Over the last 50 years a wide variety of synchronization algorithms has been developed to perform this task (see e.g. [32]). Over the last decade, however, several highly-effective channel codes have been developed, enabling reliable communication at very low signal-to-noise ratios. As a consequence, carrier synchronization must now be performed at lower SNRs than ever before, which also sets high requirements as to the synchronizers' design. The fact that traditional carrier synchronization techniques failed to cope with such high-noise environments has stimulated international research into more effective carrier synchronization structures. It is precisely in this context that our doctoral research is to be situated.

1.2 Outline

This thesis consists of seven chapters.

Following this introductory chapter, **chapter 2** gives a more complete source-to-sink description of the digital bandpass communication system that is being considered.

A theoretical discussion of the detection and estimation problems encountered in the receiver follows in **chapter 3**. This chapter highlights the importance of the *Cramer-Rao bound* (CRB), which is one of the most commonly used fundamental limits on the minimum achievable mean square estimation error (MSE). We also derive the concept of the synchronized receiver, which involves estimating the carrier frequency and phase as well as using these estimates as if they were the true synchronization parameters in order to recover the original information bits from the received demodulated signal.

Chapter 4 focuses on algorithms for *synchronized detection*, which means recovering the information bits assuming perfect knowledge of the carrier synchronization parameters. In particular it elaborates on the use of factor graphs and the sum-product algorithm and it presents the signal models that we use in our subsequent analysis.

The problem of carrier synchronization is discussed in chapters 5-6, which form the core contribution of this thesis. **Chapter 5** specifically concentrates on the derivation of Cramer-Rao lower bounds for carrier synchronization parameter estimation. A first important contribution of this chapter is an analytical expression which enables the efficient numerical evaluation of the CRB for a wide range of channel coded systems. Our results diverge in two important aspects from those in the existing literature. First, we do not assume the independence and identical distribution of the transmitted data symbols. This approach makes it possible to derive CRBs in the case of coded or pilot-symbol-aided transmissions. Second, we use the correct continuous-time model of the received signal. Where possible, we compare our results from the correct model with existing results obtained from a simplified discrete-time model of the received signal. The latter is commonly used but ignores the reduction of the useful signal and the occurrence of inter-symbol-interference caused by a nonzero frequency offset. The scientific contributions [1–4, 7–14] describe the original work related to this chapter.

Chapter 6 analyzes the performance of a number of practical feedforward carrier estimation algorithms and compares the resulting MSEs with the corresponding CRBs. The content of this chapter is based on [3–5, 11–17]. Our main contribution is a new theoretical framework for iterative synchronization structures that exchange information between the bit detector and the carrier synchronization parameter estimator. From this framework, we also derive a

new soft-decision-directed carrier synchronization algorithm with a good performance complexity trade-off. This new algorithm is shown to outperform the conventional carrier synchronization algorithms at the normal operating SNR of systems with powerful error correction.

Chapter 7 summarizes the main conclusions, adds some remarks and formulates ideas for future research.

2 System Description

In chapter 1 we briefly introduced the digital bandpass communication system. In this chapter we will give a more complete source-to-sink description of the system under consideration. The relevant notations and terminology are introduced. The model we present is a fairly standard one, mainly based on [33, 34] and it is depicted in Fig. 2.1. We recognize the five essential parts of any communication system:

- source (see section 2.1)
- transmitter (see section 2.2)
- channel (see section 2.3)
- receiver (see section 2.4)
- sink (see section 2.5)

The bit and symbol sequences that are processed by the system are written as row vectors (**b**, **c**, **d** and **a**).

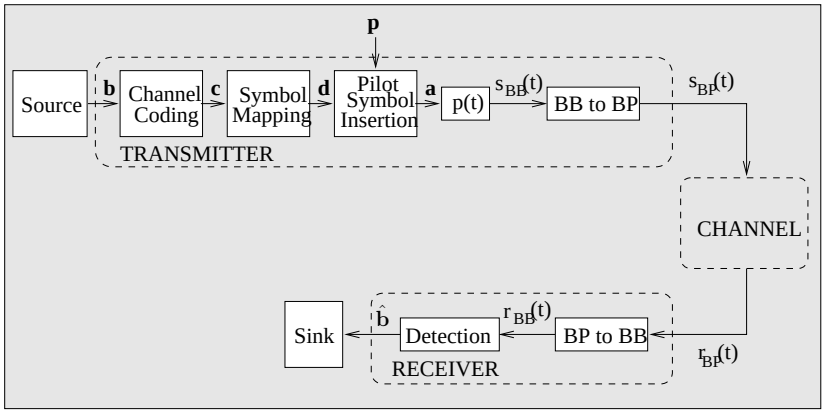

Figure 2.1: Block diagram of a digital bandpass communication system.

2.1 Source

The source of a digital bandpass communication system produces the (digitized) information that has to be communicated to the receiving sink. Its output is modeled as a random bit stream in which all bits are mutually independent and the probability of having a "1" is equal to the probability of having a "0". This reflects that the receiver has no prior knowledge about the information transmitted.

2.2 Transmitter

In many cases of practical interest, the information generated by the source is not transmitted continuously, but in *bursts*. The source bit stream is chopped into different sequences $\mathbf{b} = \begin{pmatrix} b(0) & b(1) & \dots & b(N_b-1) \end{pmatrix}$ with a length of N_b bits, and these so-called *information sequences* are transmitted one by one over the physical link. This type of burst communication is perhaps the most obvious approach towards sharing the transmission medium among multiple users that operate in a common frequency band: an overall time period can be divided into designated slots, with each of the transmitters being assigned a different time slot for transmission.

The transmitter of a bandpass communication system converts each information sequence $\mathbf{b}$ into a *bandpass* (BP) signal $s_{BP}(t)$ that is suitable for transmission over a frequency band of width $2B$ around a central frequency f_c. This conversion involves *channel coding*, *symbol mapping* and *pilot symbol insertion* followed by *pulse shaping* to produce the *baseband* (BB) signal $s_{BB}(t)$, and finally *up-conversion* of $s_{BB}(t)$ by modulating a sinusoidal *carrier* signal to obtain $s_{BP}(t)$.

The various functions of the transmitter are discussed in more detail below.

2.2.1 Channel Coding

Channel coding is often used in digital communication systems to protect the digital information against the detrimental effects (noise, interference, distortion, ...) of the channel. The channel encoder adds controlled redundancy to the information sequences. Each N_b-bit information sequence $\mathbf{b}$ is encoded into a unique *coded sequence* $\mathbf{c} = \begin{pmatrix} c(0) & c(1) & \dots & c(N_c-1) \end{pmatrix}$ containing $N_c(\geq N_b)$ bits, according to an encoding rule $\mathbf{c} = \mathcal{C}(\mathbf{b})$. Each coded bit is a function of a number of information bits. The information bits themselves may or may not be included in the coded sequence; codes that include the unmodified input in the output are called *systematic*, while those that do not are called *non-systematic*. In the systematic case, the coded sequence $\mathbf{c}$ can be separated into the N_b information bits $\mathbf{b}$ and $(N_c - N_b)$ so-called *parity-check bits*. The amount of redundancy introduced by encoding the information sequence is measured by the ratio N_c/N_b. The reciprocal of this ratio, namely, $\rho = N_b/N_c$, is called the *code rate*. Letting $\zeta_{\mathcal{C}}$ denote the set of legitimate coded sequences[1], the probability mass function of $\mathbf{c}$ is given by

$$\Pr[\mathbf{c} = \tilde{\mathbf{c}}] = \begin{cases} 2^{-N_b} & \text{, for } \tilde{\mathbf{c}} \in \zeta_{\mathcal{C}} \\ 0 & \text{, otherwise} \end{cases} . \tag{2.1}$$

where $\Pr[P]$ denotes the probability that proposition P is true. Equation (2.1) can be written more compactly by introducing the so-called *indicator function* $I[.]$, i.e., for a proposition P:

$$I[P] = \begin{cases} 0, \text{ if } P \text{ is false} \\ 1, \text{ if } P \text{ is true} \end{cases} \tag{2.2}$$

We obtain

$$\Pr[\mathbf{c} = \tilde{\mathbf{c}}] = 2^{-N_b} \cdot I[\tilde{\mathbf{c}} \in \zeta_{\mathcal{C}}] . \tag{2.3}$$

While nonlinear codes do exist, by far most codes used in practice (including the best codes today), are linear. A code $\mathcal{C}$ is linear when the modulo-2 sum of any two sequences in $\zeta_{\mathcal{C}}$ is another sequence in $\zeta_{\mathcal{C}}$. Using elementary modulo-2 matrix multiplication, the set $\zeta_{\mathcal{C}}$ of any (N_b,N_c) linear code can be represented both as

$$\zeta_{\mathcal{C}} = \left\{ \mathbf{c} \in \{0,1\}^{N_c} : \mathbf{H}\mathbf{c}^T = \mathbf{0}_{(N_c - N_b)} \right\} \tag{2.4}$$

and as

$$\zeta_{\mathcal{C}} = \left\{ \mathbf{b}\mathbf{G} : \mathbf{b} \in \{0,1\}^{N_b} \right\}, \tag{2.5}$$

where $\mathbf{0}_n$ is an all zero column vector of size n, $\mathbf{c}^T$ is the transpose of $\mathbf{c}$, and $\mathbf{H}$ and $\mathbf{G}$ are binary matrices of dimension $(N_c - N_b) \times N_c$ and $N_b \times N_c$, respectively. The matrix $\mathbf{H}$ in (2.4) is called the *parity-check matrix* of the linear

[1] The sequence $\mathbf{c}$ is a legitimate coded sequence if and only if there exists an information bit sequence $\mathbf{b}$ for which $\mathbf{c} = \mathcal{C}(\mathbf{b})$.

code and the matrix $\mathbf{G}$ in (2.5) is called the *generator matrix* of the linear code. Equation (2.5) may be interpreted as the encoding rule that maps the information sequence $\mathbf{b}$ into the corresponding coded sequence $\mathbf{c} = \mathbf{bG}$.

A generator matrix in the form:

$$\mathbf{G} = [\mathbf{I}_{N_b} | \mathbf{P}], \tag{2.6}$$

where $\mathbf{I}_n$ is a $n \times n$ identity matrix and $\mathbf{P}$ is an $N_b \times (N_c - N_b)$ matrix, is said to be in a systematic form as it generates systematic coded sequences; $\mathbf{P}$ is known as the *parity matrix*. From the systematic form of the generator matrix of a linear code, the corresponding parity-check matrix $\mathbf{H}$ can be constructed as follows:

$$\mathbf{H} = \left[\mathbf{P}^T \left| \mathbf{I}_{(N_c - N_b)} \right.\right], \tag{2.7}$$

where $\mathbf{P}^T$ is the transposed of $\mathbf{P}$ and $\mathbf{P}^T$ is of order $(N_c - N_b) \times N_b$.

Some important types of linear channel codes are described next.

2.2.1.1 Low-Density Parity-Check Codes

Low-density parity-check (LDPC) codes [35, 36] are a class of *systematic* linear channel codes that have recently received a lot of attention in the technical literature. Their main advantage is that there exist approximate decoders with reasonable complexity that yield a very good performance (as will be explained in chapter 4).

The name LDPC comes from the property that every row and every column of their parity-check matrix $\mathbf{H}$ contains only a few "1"s in comparison to the amount of "0"s. This matrix is usually constructed at *random*, subject to sparsity constraints. If we transform the sparse parity-check matrix $\mathbf{H}$ into the systematic form $\mathbf{H}' = \left[\mathbf{P}^T | \mathbf{I}\right]$ from (2.7) by proper matrix premultiplication ($\mathbf{H}' = \mathbf{M} \cdot \mathbf{H}$), the generator matrix $\mathbf{G}$ can be calculated as $\mathbf{G} = [\mathbf{I} | \mathbf{P}]$ according to (2.6). It should however be noted that the sub-matrix $\mathbf{P}$, and therefore the generator matrix $\mathbf{G}$, is generally not sparse.

It is common practice to represent the parity-check matrix of an LDPC code by a *Tanner graph*. Such a graph consists of $N_c - N_b$ *check nodes* cn_i (the number of parity-check bits) and N_c *variable nodes* vn_j (the number of coded bits). Check node cn_i is connected to variable node vn_j if the (i, j)th element $H_{i,j}$ of $\mathbf{H}$ is a "1". As an example, Fig. 2.2 depicts the Tanner graph of an LDPC code with the following parity-check matrix:

$$\mathbf{H} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}}_{\text{variable nodes}} \begin{matrix} \\ \\ \\ \\ \updownarrow \\ \leftrightarrow \end{matrix} \begin{matrix} \\ \\ \\ \\ \text{check} \\ \text{nodes} \end{matrix} . \tag{2.8}$$

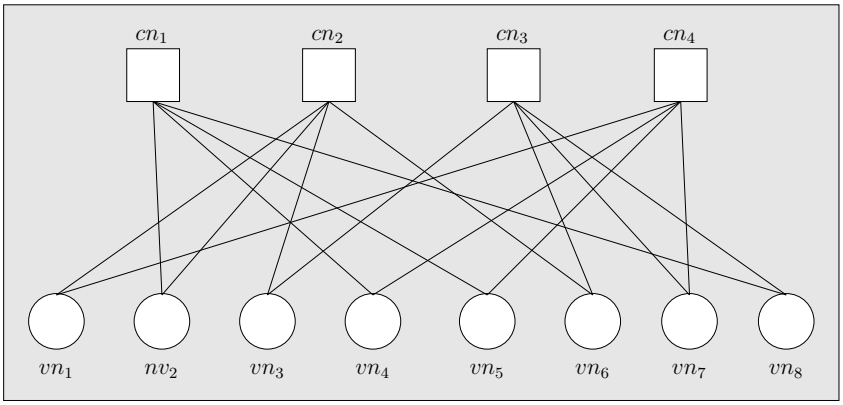

Figure 2.2: Tanner graph of the parity-check matrix in equation (2.8).

2.2.1.2 Convolutional Codes

Convolutional codes are another particular type of linear codes [33]. To keep it simple, we will stick to convolutional encoders with a rate $\rho = 1/\nu$ and memory κ, where ν and κ are positive integer-valued parameters. Such an encoder accepts one information bit at a time and outputs ν coded bits that depend on the value of the system *state*, that is, the value of the κ previous information bits. Given the current input bit and the current state, a ν-bit output sequence is generated and the state is updated. For this purpose, the encoder uses a shift register with κ stages (each stage being a one-bit storage device) and associated combinatorial logic that performs modulo-2 addition. We distinguish between *non-recursive* and *recursive* convolutional codes. Non-recursive convolutional codes use a linear feedforward shift register: the bits are fed towards the output and not back into the shift register. Recursive convolutional codes use a shift register with feedback: the content of the κ stages of the shift register is fed back to the input of the first stage. In common practice, recursive codes are usually systematic, whereas non-recursive ones are usually non-systematic.

The general block diagram of a non-recursive convolutional encoder is shown in Fig. 2.3. The κ stages of the shift register feed into ν modulo-2 adders via the links $\{g_{i,j} : i \in \{1, 2, .., \nu\},\ j \in \{1, 2, ..., \kappa\}\}$. The input to the first stage also feeds into the ν modulo-2 adders, via the links $\{g_{i,0} : i \in \{1, 2, .., \nu\}\}$. Not all these links are necessarily present. The quantities $\{g_{i,j}\}$ have value "1" when the associated link is present, and "0" otherwise. As the information sequence $\mathbf{b} = \begin{pmatrix} b(0) & b(1) & ... & b(N_b - 1) \end{pmatrix}$ enters the shift register, ν output sequences $\mathbf{c}^{(i)} = \begin{pmatrix} c^{(i)}(0) & c^{(i)}(1) & ... & c^{(i)}(N_b - 1) \end{pmatrix}$, $i = 1, 2, ..., \nu$, leave the modulo-2 adders. In mathematical terms, the correspondence between the information

bits $b(k)$ and coded bits $c^{(i)}(k)$ can be written as:

$$c^{(i)}(k) = \bigoplus_{n=0}^{min(k,\kappa)} b(k-n)\, g_{i,n}, \qquad 1 \le i \le \nu\,,\, 0 \le k \le N_b - 1,$$

where $\oplus$ denotes modulo-2 addition. On sequence level this becomes:

$$\mathbf{c}^{(i)} = \mathbf{b} \star \mathbf{g}_i, \qquad 1 \le i \le \nu,$$

where the operation $\star$ denotes the convolution operation and $\mathbf{g}_i = (g_{i,0} g_{i,1} ... g_{i,\kappa})$ are referred to as *generator sequences* of the convolutional code. A parallel-to-serial conversion of the decoder output sequences $\mathbf{c}^{(1)}$, $\mathbf{c}^{(2)}$, ..., $\mathbf{c}^{(\nu)}$ ultimately yields the coded sequence $\mathbf{c}$:

$$\mathbf{c} = \left(\; c(0) \quad c(1) \quad ... \quad c(N_c - 1) \;\right),$$

with

$$c(\nu i + (j-1)) = c^{(j)}(i) : \; 0 \le i \le N_b - 1,\, 1 \le j \le \nu.$$

The correspondence between the input sequence $\mathbf{b}$ and the coded sequences $\mathbf{c}$ can be expressed as $\mathbf{c} = \mathbf{bG}$, where $\mathbf{G}$ is an $N_b \times N_c$ generator matrix given by

$$\mathbf{G} = \begin{bmatrix} \mathcal{G}_0 & \mathcal{G}_1 & \mathcal{G}_2 & \cdots & \mathcal{G}_\kappa & & & & \\ & \mathcal{G}_0 & \mathcal{G}_1 & \mathcal{G}_2 & \cdots & \mathcal{G}_\kappa & & & \\ & & \mathcal{G}_0 & \mathcal{G}_1 & \mathcal{G}_2 & \cdots & \mathcal{G}_\kappa & & \\ & & & \ddots & \ddots & \ddots & \ldots & \ddots & \\ & & & & \mathcal{G}_0 & \mathcal{G}_1 & \mathcal{G}_2 & ... & \mathcal{G}_\kappa \end{bmatrix},$$

with

$$\mathcal{G}_\ell = \left[\; g_{1,\ell} \quad g_{2,\ell} \quad ... \quad g_{\nu,\ell} \;\right].$$

Assuming that $g_{1,0} = 1$, the non-recursive non-systematic convolutional encoder from Fig. 2.3 can be shown to be equivalent[2] to the recursive systematic convolutional encoder depicted in Fig. 2.4, whose first output sequence $\mathbf{c}^{(1)}$ equals the information sequence $\mathbf{b}$.

Unless specified otherwise, the shift registers in Fig. 2.3 and Fig. 2.4 start in the all-zero state, i.e., the state that corresponds to the situation where the content of all register stages is "0". In general there are 2^κ possible register states, each corresponding to a particular memory content. The *trellis* [37] is a useful means to track the serial progression through these states. As an example, the trellis of the non-recursive convolutional code with rate 1/2, memory 2, and generator sequences $\mathbf{g}_1 = (111)$ and $\mathbf{g}_2 = (101)$ (or $\mathbf{g}_1 = (7)_8$ and $\mathbf{g}_2 = (5)_8$ in octal notation) is depicted in Fig. 2.5. The diagram shows the transitions from one state to another as a function of the discrete time index k. Each value of k corresponds to a different so-called *trellis section*. The indication $b/c^{(1)}c^{(2)}$

[2]Two encoding rules $\mathcal{C}_1(.)$ and $\mathcal{C}_2(.)$ are said to be *equivalent* when they generate the same set of coded sequences, i.e., $\zeta_{\mathcal{C}_1} = \zeta_{\mathcal{C}_2}$.

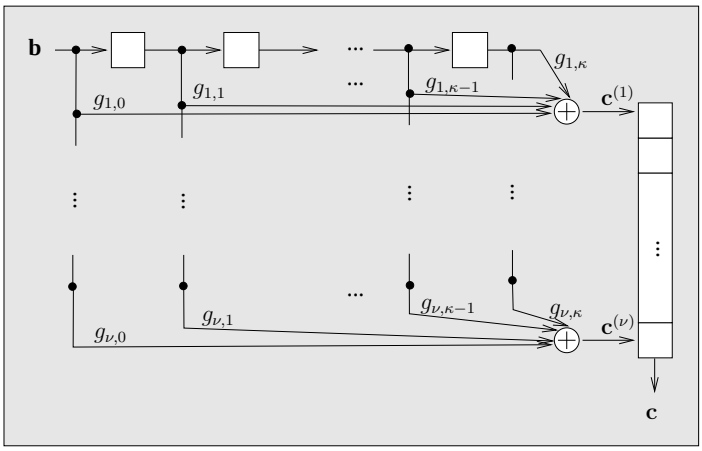

Figure 2.3: Non-recursive convolutional encoder with rate $1/\nu$ and memory κ.

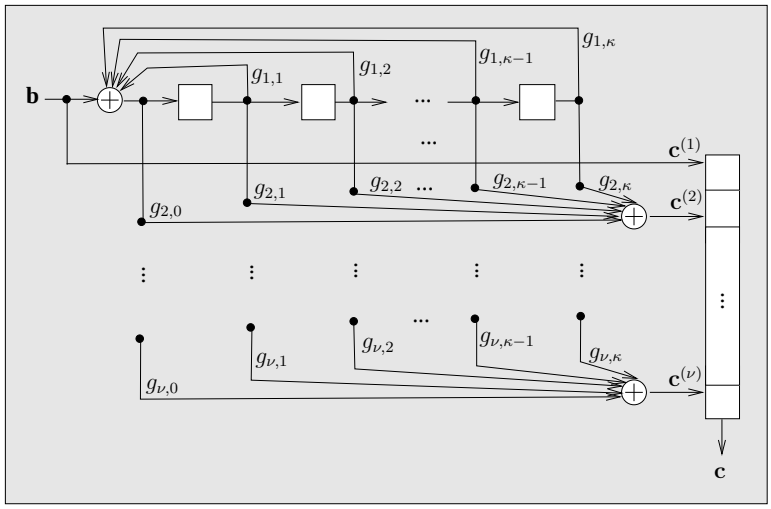

Figure 2.4: Equivalent recursive systematic convolutional encoder.

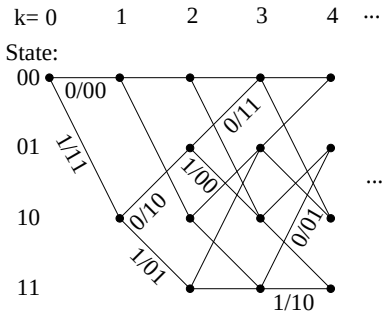

Figure 2.5: Trellis of non-recursive convolutional code with rate 1/2, memory 2, and generator sequences $\mathbf{g}_1 = (111)$ and $\mathbf{g}_2 = (101)$.

on top of each edge represents the values of the encoder input and output bits that are related to the corresponding state transition. For every coded sequence $\mathbf{c} \in \zeta_c$ there is a unique path through the trellis. The trellis is for a convolutional code what the Tanner graph is for an LDPC code: this graph not only provides a complete representation of the code, but also helps to describe the decoding algorithm as will be explained in chapter 4.

2.2.1.3 Turbo Codes

Simple codes are frequently used as constituent codes in concatenated coding schemes. The most recent development (early 1990s) in this field is *turbo encoding* [38], a scheme that combines x encoders and $(x-1)$ *interleavers* to produce a very high-performant channel code. Interleaving is the process of rearranging the ordering of a data sequence in a one to one deterministic format. The inverse of this process is referred to as deinterleaving, whereby the received sequence is restored to its original order.

Just as for LDPC codes, the optimal decoding of turbo codes is prohibitively complex. Practical turbo decoders use a suboptimal iterative decoding procedure that involves an exchange of information between the (simple) decoders of the constituent codes. For more details on the turbo decoding process the reader is referred to chapter 4.

In this thesis we will restrict our attention to the *parallel* concatenation of two rate 1/2 recursive systematic convolutional codes (see Fig. 2.6). In this case, the turbo encoder produces a sequence of coded bits that consists of three subsequences. The first sub-sequence contains the information bits. The second and the third subsequence contain the parity-check bits computed by respectively the first and the second convolutional encoder. The first encoder operates on the information bits (this is the first subsequence) whereas the second encoder operates on a known permutation of the information bits. The overall code rate is equal to 1/3. In some cases some of the parity-check bits are removed after encoding, in order to increase the code rate. This is referred to as *puncturing*.

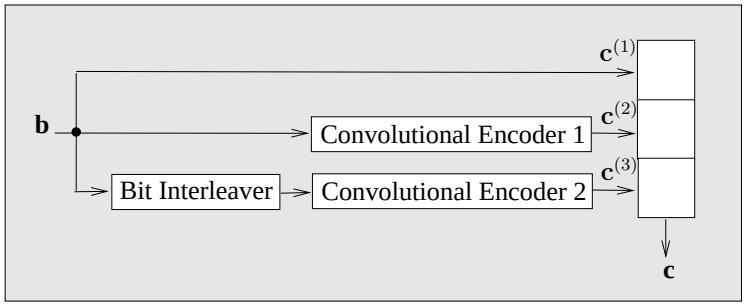

Figure 2.6: Parallel concatenated turbo encoder with rate 1/3 using convolutional codes.

While any real system must, of course, choose a particular interleaver, our computer simulations of turbo-coded systems will make use of *uniform random interleaving* [39] which is equivalent to averaging over all possible interleavers.

Of all practical channel codes known to date, turbo codes, together with LDPC codes, come closest to approaching the Shannon capacity limit [40], which represents the theoretical maximum information rate that can be transferred without errors over a noisy channel.

2.2.2 Symbol Mapping

A symbol mapper subsequently converts the coded sequence $\mathbf{c}$ to a sequence of N_d complex-valued *data symbols* $\mathbf{d} = \begin{pmatrix} d(0) & d(1) & \dots & d(N_d-1) \end{pmatrix}$. The bits at the input of the symbol mapper are first grouped in $N_d = N_c/m$ blocks of m bits; the kth block is denoted $\mathbf{c}_k = \begin{pmatrix} c_k(0) & c_k(1) & \dots & c_k(m-1) \end{pmatrix}$. Then, each $\mathbf{c}_k$, $k = 0, ..., N_d - 1$, is mapped to a complex number belonging to an M-point signaling constellation $\Omega_M = \{\omega_0, \omega_1, ..., \omega_{M-1}\}$ that meets the following conditions:

1. the constellation has a symmetry angle of $2\pi/K$ (K: integer larger than 1): when ω_i is a constellation point, then $\omega_i \exp(j2\pi/K)$ is also a constellation point;
2. $\frac{1}{M}\sum_{i=0}^{M-1} |\omega_i|^2 = 1$ (normalization).

The parameter M is commonly referred to as the order of the constellation. We write:

$$d(k) = \mathcal{D}_m(\mathbf{c}_k),$$

where $\mathcal{D}_m$ is a bijective time-invariant mapping function $\mathcal{D}_m : \{0,1\}^m \to \Omega_M$, with $M = 2^m$ that maps m bits to one constellation point. On sequence level, the overall mapping function will be denoted as $\mathcal{D}$ (i.e., without subscript m),

yielding $\mathbf{d} = \mathcal{D}(\mathbf{c})$. Denoting by $\zeta_{\mathcal{D}}$ the set of legitimate data symbol sequences[3], we have:

$$\Pr\left[\mathbf{d} = \tilde{\mathbf{d}}\right] = 2^{-N_b} \cdot I\left[\tilde{\mathbf{d}} \in \zeta_{\mathcal{D}}\right], \tag{2.9}$$

where $I[.]$ is the indicator function defined in (2.2). It is further assumed that the first- and second-order moments of the coded data symbols are the same as for uncoded symbols. Taking into account that uncoded transmission corresponds to $\mathcal{C}(\mathbf{b}) \equiv \mathbf{b}$ so that $\mathbf{d} = \mathcal{D}(\mathbf{c}) = \mathcal{D}(\mathbf{b})$, the uncoded symbols $d(k)$ are statistically independent and equiprobable. Hence, the assumption yields:

$$\begin{aligned} E[d(k)] &= 0, \\ E[d(k+n)\, d^*(k)] &= \delta_n, \\ E[d(k+n)\, d(k)] &= 0, \end{aligned}$$

for both coded and uncoded transmission. Here, $E[.]$ denotes averaging with respect to the a priori probability of the data symbols, $.^*$ is used to denote the complex conjugate and δ_k is the Kronecker delta: for $k \in \mathbb{Z}$; $\delta_0 = 1$, $\delta_{k\neq 0} = 0$. This property is exploited (either tacitly or explicitly) at many different places throughout this thesis, starting with the introduction of the notion "average symbol energy" in section 2.2.3. A motivation for this assumption is given in Appendix A on page 183.

The following signaling constellations are often used in practice:

- M-ary *Pulse Amplitude Modulation* (M-PAM) for which

  $$\Omega_M = \sqrt{\frac{3}{(M^2-1)}}\{\pm 1, \pm 3, ..., \pm(M-1)\}$$

  and the symmetry angle of the constellation is π.

- M-ary *Quadrature Amplitude Modulation* (M-QAM) for which

  $$\Omega_M = \left\{\omega : \Re\{\omega\}, \Im\{\omega\} \in \sqrt{\frac{3}{2(M-1)}}\left\{\pm 1, \pm 3, ..., \pm\left(\sqrt{M}-1\right)\right\}\right\},$$

  where $\Re\{.\}$ and $\Im\{.\}$ denote the real and the imaginary part of a complex number, and M is a power of 4. The symmetry angle of the constellation is $\pi/2$.

- M-ary *Phase Shift Keying* (M-PSK) for which

  $$\Omega_M = \left\{e^{j\frac{2\pi i}{M}} : i = 0, 1, ..., (M-1)\right\}.$$

  As $\Omega_{2-PSK} = \Omega_{2-PAM}$ and $\Omega_{4-PSK} = e^{j\frac{\pi}{4}}\Omega_{4-QAM}$ we need only to consider values of M larger than or equal to 8. The symmetry angle of the constellation in $2\pi/M$.

[3]The sequence $\mathbf{d}$ is a legitimate data symbol sequence if and only if there exists an information bit sequence $\mathbf{b}$ for which $\mathbf{d} = \mathcal{D}(\mathcal{C}(\mathbf{b}))$.

It should be noted that for a given signaling constellation Ω_M, a total of $M! = M \cdot (M-1) \cdot \cdots \cdot 1$ possible mapping functions $\mathcal{D}_m : \{0,1\}^{log_2 M} \rightarrow \Omega_M$ can be defined, and that the choice of the mapping function affects the overall system's performance. *Gray mapping*, where symbols at minimum Euclidean distance differ in exactly one bit, is widely known as optimal for minimizing the bit error rate in the case of uncoded transmission. In the case of coded transmission, choosing a proper mapping function is less straightforward. Over the last 30 years, a significant number of schemes have been proposed for channel coding with higher order ($M \geq 8$) constellations. One approach is to design coding and mapping functions jointly, so as to directly maximize the minimum Euclidean distance between legitimate data symbol sequences $\mathbf{d} \in \zeta_{\mathcal{D}}$; this is called *coded modulation*. One of the most often used coded modulation techniques is *trellis-coded modulation* (TCM) [41,42]. It combines the choice of a constellation with that of a convolutional code for the purpose of protecting the information bits without reducing the information bit rate or expanding the signal bandwidth. Another approach is to interconnect the channel encoder and the symbol mapper by means of a bit interleaver. Assume that two coded sequences $\mathbf{c}_1$ and $\mathbf{c}_2$ differ in Δ bits with $\Delta \ll N_c$. Thanks to the presence of the interleaver, these Δ bits that are different among the two coded sequences will most likely end up in Δ *distinct* data symbols. In other words, the transmitted data symbol sequences corresponding to $\mathbf{c}_1$ and $\mathbf{c}_2$ differ in Δ data symbols. This is referred to as *perfect interleaving*. Under the assumption of perfect interleaving, the mapping function can thus be optimized irrespective of the underlying encoding rule. This technique is commonly referred to as *bit interleaved coded modulation* (BICM) [43,44].

We will apply Gray mapping and BICM with Gray mapping in all our numerical simulations of uncoded and coded sytems.

2.2.3 Pilot Symbol Insertion

Very often a number (N_p) of *pilot symbols* with complex values $\mathbf{p} = (p(0)\ p(1)\ ...\ p(N_p-1))$ is inserted into the data symbol sequence $\mathbf{d}$, e.g., to aid synchronization. This yields an extended symbol sequence $\mathbf{a} = (a(k) : k \in I_s)$ of length $N_s = N_d + N_p$. The pilot symbol insertion rule $\mathbf{a} = \mathcal{A}(\mathbf{d}, \mathbf{p})$ is characterized by two bijections. The first bijection $\gamma_p(.)$ (with inverse $\gamma_p^{-1}(.)$) from the set of indices $\{0, 1, ..., N_p - 1\}$ to the set of indices $I_p \subset I_s$ indicates which of the symbols $a(k)$ are pilot symbols. The second bijection $\gamma_d(.)$ (with inverse $\gamma_d^{-1}(.)$) from the set of indices $\{0, 1, ..., N_d - 1\}$ to the set of indices $I_d = I_s \setminus I_p$ indicates which symbols $a(k)$ are data symbols:

$$\mathbf{a} = \mathcal{A}(\mathbf{d}, \mathbf{p}) = (a(k) : k \in I_s) \Leftrightarrow a(k) = \begin{cases} p\left(\gamma_p^{-1}(k)\right) & , k \in I_p \\ d\left(\gamma_d^{-1}(k)\right) & , k \in I_d \end{cases}. \quad (2.10)$$

For k belonging to the set I_p, $a(k)$ is a pilot symbol, i.e., the value $p\left(\gamma_p^{-1}(k)\right)$ of $a(k)$ is a priori known to the receiver. For k belonging to the set I_d, $a(k)$ is an unknown information-bearing data symbol $d\left(\gamma_d^{-1}(k)\right)$. Denoting by $\zeta_{\mathcal{A}}$ the

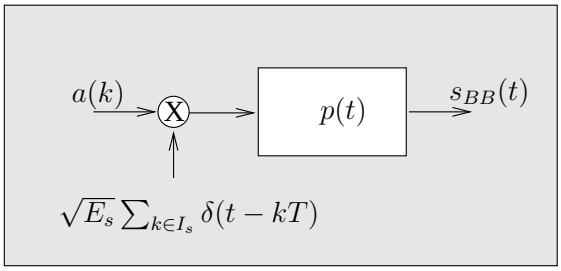

Figure 2.7: Baseband filtering at the transmitter.

set of legitimate symbol sequences[4], we have:

$$\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}\right]=2^{-N_b}\cdot I\left[\tilde{\mathbf{a}}\in\zeta_{\mathcal{A}}\right]. \tag{2.11}$$

Pilot symbols will be regarded as random variables that are statistically independent of all other symbols and have an a priori probability mass function given by $\Pr\left[a\left(k\right)=\tilde{a}\right]=I\left[\tilde{a}=p\left(\gamma_p^{-1}\left(k\right)\right)\right]$ for $k\in I_p$, where $I\left[.\right]$ is the indicator function defined in (2.2).

The ratio N_p/N_s will be referred to as the *pilot symbol ratio* λ_p. Since pilot symbols reduce both spectral efficiency (i.e., the number of information bits per second that can be transmitted per Hertz of bandwidth) and power efficiency (i.e., the percentage of the available transmit power that is used for the transmission of information-bearing data symbols), it is desirable to limit the number of transmitted pilot symbols. We will therefore assume that $\lambda_p\ll 1$.

To make sure that all symbols have the same average symbol energy, it is also assumed that $E\left[\left|a\left(k\right)\right|^2\right]=1$ for $k\in I_s$. As far as the data symbols are concerned (i.e., $k\in I_d$), this evidently holds for uncoded PAM, uncoded QAM, uncoded PSK and coded PSK; we show in Appendix A on page 183 that it also holds for many practical coded PAM and coded QAM modulations. It further implies that all pilot symbols are assumed to have values that are located on the unit circle of the complex plane, i.e., $\left|p\left(k\right)\right|^2=1$ for $k=0,\,1,\,...,\,N_p-1$.

We use both 4-QAM (4-PSK) pilot symbols and 4-QAM (4-PSK) data symbols in all the simulations of systems with pilot symbols.

2.2.4 Pulse Shaping

The symbol sequence $\mathbf{a}$ is applied to a baseband transmit filter with transfer function $P\left(f\right)$ and impulse response $p\left(t\right)$, yielding a continuous-time baseband signal $s_{BB}\left(t\right)$. The (one-side) bandwidth of the transmit filter is equal to B: $P\left(f\right)=0$ for $\left|f\right|>B$. In Fig. 2.7, this transition from the discrete-time to the continuous-time domain is modeled by applying to the filter a sequence of

[4]The sequence $\mathbf{a}$ is a legitimate symbol sequence if and only if there exists an information bit sequence $\mathbf{b}$ for which $\mathbf{a}=\mathcal{A}\left(\mathcal{D}\left(\mathcal{C}\left(\mathbf{b}\right)\right),\mathbf{p}\right)$.

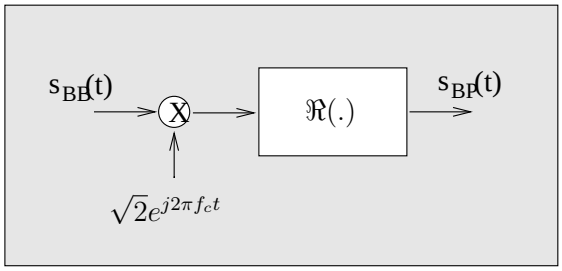

Figure 2.8: Up-conversion at the transmitter.

Dirac impulses at a rate $R_s = 1/T$, with weights $\{\sqrt{E_s}a(k)\}$. This weighted sequence of Dirac impulses results from multiplying the symbols $\{a(k)\}$ with the transmit clock signal, consisting of a periodic sequence of Dirac impulses multiplied by $\sqrt{E_s}$ and occurring at instances $\{kT\}$. This yields:

$$s_{BB}(t;\mathbf{a}) = \sqrt{E_s}\sum_{k\in I_s} a(k)\, p(t-kT), \tag{2.12}$$

where the quantity E_s denotes the average symbol energy. The conversion (2.12) from $\mathbf{a}$ to $s_{BB}(t)$ is referred to as *linear digital modulation*, because the signal $s_{BB}(t)$ is linear in the transmitted symbols $a(k)$.

The average energy per information bit E_b, per coded bit E_c and per data symbol E_d are related to the average symbol energy E_s by $E_b = E_s\frac{N_s}{N_b} = E_s\frac{N_s}{\rho m N_d}$, $E_c = E_s\frac{N_s}{N_c} = E_s\frac{N_s}{mN_d}$ and $E_d = E_s\frac{N_s}{N_d}$. The information bit rate R_b, the coded bit rate R_c and the data symbol rate R_d are related to the symbol rate $R_s = 1/T$ by $R_b = R_s\frac{N_b}{N_s} = R_s\frac{\rho m N_d}{N_s}$, $R_c = R_s\frac{N_c}{N_s} = R_s\frac{mN_d}{N_s}$ and $R_d = R_s\frac{N_d}{N_s}$. The average transmit power $\mathcal{P}$ is given by $\mathcal{P} = E_bR_b = E_cR_c = E_dR_d = E_sR_s$.

In this thesis the transmit filter is assumed to be a *real-valued* and *even symmetric unit-energy square-root Nyquist filter* with respect to the time interval T (*symbol period*), i.e.,

$$p(t) = p(-t) \in \mathbb{R},\ \forall t \tag{2.13}$$

and

$$g(kT) \quad = \quad \int_{-\infty}^{\infty} p\left(\tilde{t}\right) p\left(\tilde{t}-kT\right) \mathrm{d}\tilde{t} = \delta_k, \tag{2.14}$$

where $g(t)$ is the Nyquist pulse with Fourier transform $G(f)$.

2.2.5 Up-conversion

The communication system is assigned a bandpass channel (i.e., a frequency band of width $2B$ around a central frequency $f_c \gg B$), so the baseband signal $s_{BB}(t;\mathbf{a})$ must be up-converted or frequency-translated by means of a sinusoidal

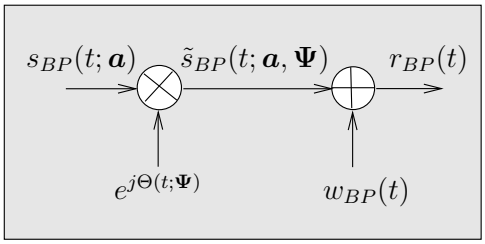

Figure 2.9: AWGN channel with phase and frequency uncertainty.

waveform, called the *carrier*. In Fig. 2.8 the resulting bandpass signal $s_{BP}(t;\mathbf{a})$ is modeled as:

$$s_{BP}(t;\mathbf{a}) = \sqrt{2}\Re\left\{s_{BB}(t;\mathbf{a})\, e^{j2\pi f_c t}\right\}. \tag{2.15}$$

The frequency content of $s_{BP}(t;\mathbf{a})$ is concentrated near the *carrier frequency* f_c. Due to the factor $\sqrt{2}$ in (2.15), the bandpass signal $s_{BP}(t;\mathbf{a})$ has the same power as the baseband signal $s_{BB}(t;\mathbf{a})$, and consequently the energy per symbol is the same for both signals.

2.3 Channel

The channel model is a description of how the physical transmission medium alters the signal $s_{BP}(t)$ that is being transmitted. In this thesis, we will assume an *additive white Gaussian noise* (AWGN) channel, as depicted in Fig. 2.9. This is a convenient model for communication over wired channels when the signal bandwidth is small as compared to the channel bandwidth, so that the channel transfer function is essentially constant over the signal bandwidth (e.g., transmission of a 1 MHz bandwidth signal over a coaxial Community Antenna Television or CATV network) and for communication over wireless channels when there is a dominating line-of-sight path (as in satellite communication and radio relay systems). The signal at the input of the receiver is described as:

$$r_{BP}(t) \;=\; \tilde{s}_{BP}(t;\mathbf{a},\boldsymbol{\Psi}) + w_{BP}(t), \tag{2.16}$$

with $w_{BP}(t)$ real-valued AWGN with zero mean and power spectral density $N_0/2$ and $\tilde{s}_{BP}(t;\mathbf{a},\boldsymbol{\Psi})$ given by

$$\tilde{s}_{BP}(t;\mathbf{a},\boldsymbol{\Psi}) = \sqrt{2}\Re\left\{\sqrt{E_s}\sum_{k\in I_s} a(k)\, p(t-kT)\, e^{j(2\pi f_c t+\Theta(t;\boldsymbol{\Psi}))}\right\},$$

where $\Theta(t;\boldsymbol{\Psi})$ reflects the absence of accurate carrier information at the receiver; $\Theta(t;\boldsymbol{\Psi})$ stands for the instantaneous carrier phase shift of the received signal vis-a-vis the receiver's local reference carrier, and is assumed to be fully specified by a parameter vector $\boldsymbol{\Psi}$.

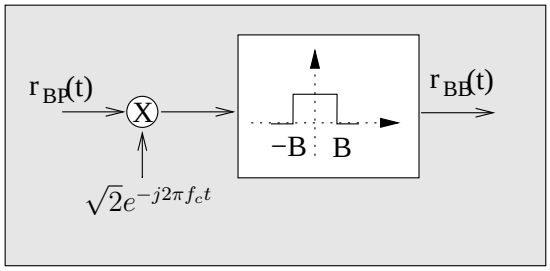

Figure 2.10: Down-conversion at the receiver.

2.4 Receiver

In the receiver of a digital bandpass communication system, the received bandpass signal $r_{BP}(t)$ is first down-converted to obtain its complex baseband equivalent. As indicated in Fig. 2.10, the down-conversion is accomplished by multiplying $r_{BP}(t)$ with a local oscillator signal with frequency f_c, followed by low-pass filtering to constrain the bandwidth to $[-B, B]$. The resulting baseband signal $r_{BB}(t)$ is given by:

$$r_{BB}(t) = \tilde{s}_{BB}(t; \mathbf{a}, \mathbf{\Psi}) + w_{BB}(t), \tag{2.17}$$

where

$$\tilde{s}_{BB}(t; \mathbf{a}, \mathbf{\Psi}) = \sqrt{E_s} \sum_{k \in I_s} a(k)\, p(t - kT)\, e^{j\Theta(t;\mathbf{\Psi})}$$

and $w_{BB}(t)$ is complex-valued zero-mean AWGN with independent real and imaginary parts each having a power spectral density equal to $N_0/2$ over the frequency interval $[-B, B]$.

Further baseband processing then aims at producing, from the baseband signal $r_{BB}(t)$ from (2.17), an estimate $\hat{b}(k)$ for each of the N_b bits of the information sequence $\mathbf{b} = \begin{pmatrix} b(0) & b(1) & \ldots & b(N_b - 1) \end{pmatrix}$. The process of generating these estimates is referred to as *bit detection*.

2.5 Sink

After bit detection, the sink accepts the output sequence $\hat{\mathbf{b}}$ from the receiver and, knowing the digitization method used at the source, attempts to reconstruct the original information.

2.6 Conclusion and Remarks

In this chapter we have presented the system model: source, transmitter, channel, receiver and sink. A very important part of the receiver that has only been

touched upon briefly is the baseband processing with the ultimate goal of recovering (or detecting) the original information bit sequence. More details on the detection process are given in chapter 4. But first, we will review a number of elements from the Bayesian detection and estimation theory in chapter 3.

3

Detection and Estimation Theory

As explained in the previous chapter, the receiver of a digital bandpass communication system contains a bit detector whose first task consists in estimating the transmitted information bits from the incoming signal. This kind of bit detection is only one of the many applications of the *detection and estimation theory,* which aims at extracting the value of a set of parameters $\mathbf{u}$ from a set of noisy observations $\mathbf{r}$.

After investigating the *maximum a posteriori* decision and estimation criteria (section 3.3 and section 3.4.1), which are the foundation of many practical detection and estimation algorithms, we will take a close look at the *Cramer-Rao bound* (section 3.4.2 and section 3.4.3), one of the most often used fundamental limits on the minimum achievable mean square estimation error.

3.1 Problem Formulation

We observe the following detection or estimation problem.

A set of N random parameters u_1, u_2, ..., u_N is denoted by a column vector $\mathbf{u} = \begin{pmatrix} u_1 & u_2 & ... & u_N \end{pmatrix}^T$. Each of the parameters is defined over some domain $\mathcal{U}_i$, so that $\mathbf{u}$ is defined over $\mathcal{U}_1 \times ... \times \mathcal{U}_N$. As a result of a particular realization of $\mathbf{u}$, a random vector variable $\mathbf{r}$ is observed. The probabilistic dependence of $\mathbf{r}$ on $\mathbf{u}$ is described by the *conditional* probability density function $p(\mathbf{r}|\mathbf{u})$. When viewed as function of $\mathbf{u}$, $p(\mathbf{r}|\mathbf{u})$ is called the *likelihood function* $L(\mathbf{u};\mathbf{r})$ of $\mathbf{u}$, while its logarithm $\ln p(\mathbf{r}|\mathbf{u})$ is called the *log-likelihood function* $\ell(\mathbf{u};\mathbf{r})$ of $\mathbf{u}$. From the observation $\mathbf{r}$, we want to recover the value of $\mathbf{u}$. We denote the recovered value as $\hat{\mathbf{u}}(\mathbf{r}) = \begin{pmatrix} \hat{u}_1(\mathbf{r}) & \hat{u}_2(\mathbf{r}) & ... & \hat{u}_N(\mathbf{r}) \end{pmatrix}^T$. Depending on whether $\mathbf{u}$ is discrete-valued or continuous-valued, the quantities $\hat{u}_i(\mathbf{r})$ are referred to as *decisions* or *estimates*, while the process of generating these quantities is called *detection* or *estimation*. The mapping of the observation into a decision or an estimate is denoted as the decision or estimation *rule*.

When the observation $\mathbf{r}$ depends not only on the parameters $\mathbf{u}$ to be estimated but also on a *nuisance* vector parameter $\mathbf{v}$ (which is assumed to be independent of $\mathbf{u}$), the likelihood function of $\mathbf{u}$ is obtained by averaging the conditional distribution $p(\mathbf{r}|\mathbf{v},\mathbf{u})$ of $\mathbf{r}$ given $(\mathbf{v},\mathbf{u})$ over the a priori distribution $p(\mathbf{v})$ of the nuisance vector parameter $\mathbf{v}$, i.e.,

$$L(\mathbf{u};\mathbf{r}) = p(\mathbf{r}|\mathbf{u}) = E_{\mathbf{v}}\left[p(\mathbf{r}|\mathbf{v},\mathbf{u})\right]. \tag{3.1}$$

The corresponding log-likelihood function of $\mathbf{u}$ is then given by:

$$\ell(\mathbf{u};\mathbf{r}) = \ln p(\mathbf{r}|\mathbf{u}) = \ln\left(E_{\mathbf{v}}\left[p(\mathbf{r}|\mathbf{v},\mathbf{u})\right]\right). \tag{3.2}$$

We refer to $p(\mathbf{r}|\mathbf{v},\mathbf{u})$ as the joint likelihood function $L(\mathbf{v},\mathbf{u};\mathbf{r})$ of $\mathbf{v}$ and $\mathbf{u}$, as $p(\mathbf{r}|\mathbf{v},\mathbf{u})$ is relevant to the joint estimation of $\mathbf{v}$ and $\mathbf{u}$. A special case of interest occurs when the conditional probability density function $p(\mathbf{r}|\mathbf{v},\mathbf{u})$ is Gaussian (see further).

3.2 Bayesian and Standard Approach

There exists a Bayesian and a standard non-Bayesian approach to the detection and the estimation of the wanted parameters $\mathbf{u}$. The *Bayesian* detection and estimation theory assumes that the parameters $\mathbf{u}$ are *random* variables with known joint a priori probability density function $p(\mathbf{u}) = p(u_1, u_2, ..., u_N)$. A special case of interest occurs when the a priori distribution $p(\mathbf{u})$ is uniform over $\mathcal{U}_1 \times ... \times \mathcal{U}_N$ (see further). The key characteristic of the Bayesian approach is that all parameters (including the vector $\mathbf{u}$) are treated as random variables and all uncertainties are expressed as probabilities. This gives Bayesian analysis an attractive uniformity and coherence in comparison to the standard approach which treats the wanted parameters $\mathbf{u}$ as *deterministic* unknowns while the nuisance parameters $\mathbf{v}$ are usually modeled as random variables.

3.3 Bayesian Detection

We will first examine the Bayesian detection problem. In this case the components of the wanted parameter $\mathbf{u}$ are *discrete-valued* random variables. In what follows, we will also assume that the components of the nuisance parameter $\mathbf{v}$ are *continuous-valued* random variables. If this were not to hold for all components of $\mathbf{v}$, the derivation below is still valid, provided that a summation (rather than integration) over the discrete-valued components of $\mathbf{v}$ is carried out.

3.3.1 Maximum a posteriori Detection

If a parameter is detected erroneously, i.e., $\hat{u}_i(\mathbf{r}) \neq u_i$, a *detection error* occurs. The probability of having a detection error, i.e., $\Pr[\hat{u}_i(\mathbf{r}) \neq u_i]$, is a quantitative measure for the performance of the detector. Optimal detection, which minimizes the detection error $\Pr[\hat{u}_i(\mathbf{r}) \neq u_i]$ for $i = 0, 1, ..., N-1$, is achieved by *maximum a posteriori* (MAP) decision of the individual parameters u_i. The corresponding MAP detector maximizes the a posteriori probability (APP) $p(u_i|\mathbf{r})$ with respect to u_i [34]:

$$\hat{u}_{i,\mathrm{MAP}}(\mathbf{r}) = \arg\max_{u_i \in \mathcal{U}_i} p(u_i|\mathbf{r}), \quad i = 0, ..., N-1. \tag{3.3}$$

In principle, the APPs $p(u_i|\mathbf{r})$ involved in (3.3) can be obtained as marginals of the joint APP $p(\mathbf{v}, \mathbf{u}|\mathbf{r})$, i.e., according to:

$$p(u_i|\mathbf{r}) = \sum_{\mathbf{u} \backslash u_i} \int_{\mathbf{v}} p(\mathbf{v}, \mathbf{u}|\mathbf{r}) \, \mathrm{d}\mathbf{v}, \tag{3.4}$$

$$= \sum_{\mathbf{u} \backslash u_i} \int_{\mathbf{v}} p(\mathbf{u}|\mathbf{r}, \mathbf{v}) \, p(\mathbf{v}|\mathbf{r}) \, \mathrm{d}\mathbf{v}, \tag{3.5}$$

where $\mathbf{u} \backslash u_i$ denotes that the summation ranges over all parameters $\mathbf{u}$, except u_i. The joint APP $p(\mathbf{v}, \mathbf{u}|\mathbf{r})$ can be computed from the joint likelihood function $L(\mathbf{v}, \mathbf{u}; \mathbf{r}) = p(\mathbf{r}|\mathbf{v}, \mathbf{u})$, the a priori distributions $p(\mathbf{u})$ and $p(\mathbf{v})$, and Bayes rule (i.e., $p(Y|X)\,p(X) = p(X|Y)\,p(Y)$). We obtain:

$$p(\mathbf{v}, \mathbf{u}|\mathbf{r}) \propto L(\mathbf{v}, \mathbf{u}; \mathbf{r}) \, p(\mathbf{u}) \, p(\mathbf{v}), \tag{3.6}$$

where "$\propto$" denotes equality up to an irrelevant scaling factor not depending on $(\mathbf{v}, \mathbf{u})$. Since distributions are normalized, we can easily determine this scaling factor. Unfortunately, the integral in (3.4) over all possible values of the continuous parameters $\mathbf{v}$ is commonly too complex to evaluate. In addition, the summation in (3.4) may contain a huge number of terms. An elegant solution to this problem is to make use of a two-stage recovery structure. First, an estimate $\hat{\mathbf{v}}(\mathbf{r})$ of the continuous-valued nuisance parameter vector $\mathbf{v}$ is derived from the observation $\mathbf{r}$. Then, the wanted parameters u_i are detected assuming ideal knowledge of the nuisance parameters, i.e., assuming that $\mathbf{v} \equiv \hat{\mathbf{v}}(\mathbf{r})$. Stated

more formally, this involves approximating in (3.5) $p(\mathbf{v}|\mathbf{r})$ by $\delta(\mathbf{v}-\hat{\mathbf{v}}(\mathbf{r}))$, where $\delta(.)$ is the multidimensional Dirac delta function. This yields:

$$\hat{u}_i(\mathbf{r}) \quad = \quad \arg\max_{u_i \in \mathcal{U}_i} p(u_i|\mathbf{r}, \hat{\mathbf{v}}(\mathbf{r})), \quad k = 0, ..., N-1, \tag{3.7}$$

with

$$p(u_i|\mathbf{r}, \hat{\mathbf{v}}(\mathbf{r})) = \sum_{\mathbf{u}\backslash u_i} p(\mathbf{u}|\mathbf{r}, \hat{\mathbf{v}}(\mathbf{r})) \tag{3.8}$$

and

$$p(\mathbf{u}|\mathbf{r}, \hat{\mathbf{v}}(\mathbf{r})) \propto L(\hat{\mathbf{v}}(\mathbf{r}), \mathbf{u}; \mathbf{r}) p(\mathbf{u}), \tag{3.9}$$

where "$\propto$" denotes equality up to an irrelevant scaling factor not depending on $\mathbf{u}$.

Let $|\mathcal{U}_j|$ denote the cardinality of the domain alphabet $\mathcal{U}_j$ of the jth parameter u_j. The summation in (3.8) contains $\prod_{j\neq i \wedge j=1}^{N} |\mathcal{U}_j|$ terms, one for each possible combination of the parameters $\mathbf{u}\backslash u_i$. This would imply that the computational complexity associated with the evaluation of the APPs $p(u_i|\mathbf{r}, \hat{\mathbf{v}}(\mathbf{r}))$ increases exponentially with the number of wanted parameters N. In many practical scenarios, however, the required APPs can be computed more efficiently, either exactly or approximately, by applying the *sum-product algorithm* to a *factor graph* representing a suitable factorization of the joint APP $p(\mathbf{u}|\mathbf{r}, \hat{\mathbf{v}}(\mathbf{r}))$ from (3.9) [45].

3.3.2 Factor Graphs and the Sum-Product Algorithm

A factor graph is a diagram that represents the factorization of a function of several variables

$$f(x_1, ..., x_N) = \prod_j f_j(X_j), \tag{3.10}$$

where X_j is a subset of $\mathbf{x} = (x_1, ..., x_N)$. It consists of nodes, edges and half-edges (the latter are connected to only one node). There is a unique node for every factor and a unique (half-)edge for every variable. Node f_j is connected to variable x_n when $x_n \in X_j$.[1] Factor graphs not only allow us to visualize a function, they also provide an efficient way to compute all its *marginals*, $g_i(x_i) = \sum_{\mathbf{x}\backslash x_i} f(x_1, x_2, \ldots, x_N)$[2], by means of the sum-product algorithm. This is a message-passing algorithm, where messages are computed in the nodes of a graph and passed along its edges. A message along a given edge is a function of the variable corresponding to that edge. Messages can at all times be interpreted as probability mass functions (or probability density functions, for continuous variables). The marginal $g_i(x_i)$ is given by the product of the

[1] Implicit in this definition is the assumption that none of the variables appears in more than two factors. We will illustrate in chapter 4 how this seemingly severe restriction can be easily lifted by introducing additional 'dummy' variables that are 'clones' of the original variables.

[2] In the case of continuous valued variables, summations must be replaced with integrals over the corresponding domain.

two messages (one in each direction) along the edge corresponding to x_i. For more details on how messages in nodes are computed we refer to [46].

The application of the sum-product algorithm on a graph that corresponds to a tree is straightforward and yields the exact marginals. When the graph contains cycles (loops), the sum-product algorithm becomes an iterative procedure that, after convergence, yields only an *approximation* of the marginals. However, when the cycles in the graph are large, the resulting marginals turn out to be quite accurate. Hence, the accuracy of the sum-product algorithm is directly related to the minimum cycle length (i.e., the *girth*) of the graph.

While discrete variables can usually be handled very efficiently by the sum-product algorithm, continuous variables often lead to intractable integrals. On top of that, the messages over the corresponding edges will be probability density functions rather than probability mass functions (the latter have convenient vector representations). Therefore, the question arises how these messages should be represented. This immediately explains why, from a computational point of view, the framework of sum-product algorithm and factor graphs is attractive for the computation of the APPs from (3.8) that assume $\mathbf{v} = \hat{\mathbf{v}}(\mathbf{r})$, but much less for the computation of the correct APPs from (3.4) that involve integration over $\mathbf{v}$.

3.4 Bayesian Estimation

We will now consider the problem of Bayesian parameter estimation. In this case the components of the wanted parameter $\mathbf{u}$ are *continuous-valued* random variables. We also assume that the components of $\mathbf{u}$ are real-valued. When estimating complex-valued parameters, their real and imaginary parts are stacked into a real-valued vector $\mathbf{u}$.

3.4.1 Bayesian Estimation Algorithms

For further use, we define the *estimation error* vector $\mathbf{e}(\mathbf{r})$ as:

$$\mathbf{e}(\mathbf{r}) = \begin{pmatrix} u_1 - \hat{u}_1(\mathbf{r}) \\ u_2 - \hat{u}_2(\mathbf{r}) \\ \vdots \\ u_N - \hat{u}_N(\mathbf{r}) \end{pmatrix} = \mathbf{u} - \hat{\mathbf{u}}(\mathbf{r}), \tag{3.11}$$

with the i-th component of (3.11) denoting the estimation error $e_i(\mathbf{r}) = u_i - \hat{u}_i(\mathbf{r})$ related to the i-th parameter u_i. We also define the error correlation matrix $\mathbf{R}^{\mathbf{e}}$ as:

$$\mathbf{R}^{\mathbf{e}} = E_{\mathbf{r},\mathbf{u}}\left[\mathbf{e}(\mathbf{r})\,\mathbf{e}^T(\mathbf{r})\right], \tag{3.12}$$

where the expectation $E_{\mathbf{r},\mathbf{u}}[.]$ denotes averaging over the joint probability density function $p(\mathbf{r},\mathbf{u}) = p(\mathbf{r}|\mathbf{u})\,p(\mathbf{u})$ of the random vector variables $\mathbf{r}$ and $\mathbf{u}$. The elements of $\mathbf{R}^{\mathbf{e}}$ are

$$R^{\mathbf{e}}_{i,j} = E_{\mathbf{r},\mathbf{u}}\left[e_i(\mathbf{r})\,e_j(\mathbf{r})\right]. \tag{3.13}$$

Note that $\mathbf{R^e}$ is a symmetrical matrix. The diagonal element $R^{\mathbf{e}}_{i,i}$ represents the *mean square error* (MSE) $E_{\mathbf{r},\mathbf{u}}\left[\left(e_i\left(\mathbf{r}\right)\right)^2\right]$ and the off-diagonal elements $(i \neq j)$ are the cross correlations of the estimation errors.

Estimation accuracy is usually measured by the MSE. The estimate of $\mathbf{u}$ which minimizes the MSEs $E_{\mathbf{r},\mathbf{u}}\left[\left(e_i\left(\mathbf{r}\right)\right)^2\right]$ for $i = 1, 2, ..., N$ is referred to as the *minimum mean square error* (MMSE) estimate $\hat{\mathbf{u}}_{MMSE}\left(\mathbf{r}\right)$. In [47], it is shown that the MMSE estimation rule reduces to

$$\hat{\mathbf{u}}_{MMSE}\left(\mathbf{r}\right) = \int_{-\infty}^{\infty} \mathbf{u}\, p\left(\mathbf{u}\,|\mathbf{r}\right) \mathrm{d}\mathbf{u}, \tag{3.14}$$

which indicates that $\hat{\mathbf{u}}_{MMSE}\left(\mathbf{r}\right)$ equals the mean of the a posteriori distribution $p\left(\mathbf{u}\,|\mathbf{r}\right)$, i.e., the *a posteriori mean* of $\mathbf{u}$. In the same way as in section 3.3, the a posteriori probability density function $p\left(\mathbf{u}\,|\mathbf{r}\right)$ of $\mathbf{u}$ given $\mathbf{r}$ can be obtained as a marginal of the joint APP $p\left(\mathbf{v},\mathbf{u}\,|\mathbf{r}\right)$ which, in turn, can be computed from (3.6). In this case the framework of sum-product algorithm and factor graphs is not the appropriate way to compute $p\left(\mathbf{u}\,|\mathbf{r}\right)$ from $p\left(\mathbf{v},\mathbf{u}\,|\mathbf{r}\right)$ because $\mathbf{u}$ is continuous-valued. We obtain:

$$p\left(\mathbf{u}\,|\mathbf{r}\right) = \sum_{\mathbf{v}} p\left(\mathbf{v},\mathbf{u}\,|\mathbf{r}\right), \tag{3.15}$$

$$\propto \sum_{\mathbf{v}} L\left(\mathbf{v},\mathbf{u};\mathbf{r}\right) p\left(\mathbf{u}\right) p\left(\mathbf{v}\right), \tag{3.16}$$

where the summation ranges over all nuisance parameters $\mathbf{v}$ and "$\propto$" denotes equality up to an irrelevant scaling factor not depending on $\mathbf{u}$. Without loss of generality it was assumed that all nuisance parameters $\mathbf{v}$ are discrete-valued (in case of any continuous-valued components of $\mathbf{v}$ the corresponding summation must be replaced by an integration). In general, the computation of the MMSE estimate is rather complex, because the required averaging of the a posteriori density $p\left(\mathbf{u}\,|\mathbf{r}\right)$ in (3.14); in addition, the summation in (3.15)-(3.16) might contain a huge number of terms.

In order to avoid the computational complexity associated with the MMSE estimate, it is more convenient to opt for the estimate that maximizes the a posteriori probability $p\left(\mathbf{u}\,|\mathbf{r}\right)$. The resulting estimate is called the *maximum a posteriori* (MAP) estimate, which we denote as $\hat{\mathbf{u}}_{MAP}\left(\mathbf{r}\right)$. We obtain:

$$\hat{\mathbf{u}}_{MAP}\left(\mathbf{r}\right) = \arg\max_{\mathbf{u}} p\left(\mathbf{u}\,|\mathbf{r}\right). \tag{3.17}$$

Because the logarithm is a monotonic function the MAP estimate also corresponds to the maximum of $\ln p\left(\mathbf{u}\,|\mathbf{r}\right)$. This is often easier to compute, especially when $L\left(\mathbf{v},\mathbf{u};\mathbf{r}\right) = p\left(\mathbf{r}\,|\mathbf{v},\mathbf{u}\right)$ is Gaussian. If the maximum is interior (i.e., if the maximum of $p\left(\mathbf{u}\,|\mathbf{r}\right)$ belongs to the domain $\mathcal{U}_1 \times ... \times \mathcal{U}_N$ of $\mathbf{u}$) and $\frac{\partial p(\mathbf{u}|\mathbf{r})}{\partial u_i}$, $i = 1, 2, ..., N$ exists at the maximum, then a necessary, but not sufficient, condition is obtained by stating that the derivative of $\ln p\left(\mathbf{u}\,|\mathbf{r}\right)$ with respect to

each parameter u_i, $i = 1, 2, ..., N$ is zero for $\mathbf{u} = \hat{\mathbf{u}}_{MAP}(\mathbf{r})$. This yields a set of N simultaneous equations referred to as the *MAP equations*:

$$\left.\frac{\partial \ln p(\mathbf{u}|\mathbf{r})}{\partial u_i}\right|_{\mathbf{u}=\hat{\mathbf{u}}_{MAP}(\mathbf{r})} = 0, \quad i = 1, 2, ..., N. \tag{3.18}$$

In many cases of practical interest the MAP and MMSE estimates turn out to be equal [47]. However, technically it is usually easier to compute the MAP solutions than to find the a posteriori mean. As the summation in (3.16) gives rise to a computational burden that is exponential in the size of $\mathbf{v}$, MAP estimation as defined in (3.17) cannot be used for large nuisance parameter vectors, in which case you inevitably have to resort to further approximation techniques.

For a uniform a priori probability density function $p(\mathbf{u})$, the MAP estimation rule from (3.17) reduces to the *maximum likelihood* (ML) estimation rule, which is defined as:

$$\hat{\mathbf{u}}_{ML}(\mathbf{r}) = \arg\max_{\mathbf{u}} L(\mathbf{u};\mathbf{r}) = \arg\max_{\mathbf{u}} \ell(\mathbf{u};\mathbf{r}), \tag{3.19}$$

where $\ell(\mathbf{u};\mathbf{r}) = \ln L(\mathbf{u};\mathbf{r}) = \ln p(\mathbf{r}|\mathbf{u})$. In the same way as for the MAP estimate, a necessary (but not sufficient) condition on the ML estimate can be derived; the resulting *ML equations* are:

$$\left.\frac{\partial \ell(\mathbf{u};\mathbf{r})}{\partial u_i}\right|_{\mathbf{u}=\hat{\mathbf{u}}_{ML}(\mathbf{r})} = 0, \quad i = 1, 2, ..., N. \tag{3.20}$$

3.4.2 Bayesian Cramer-Rao Bound

The Bayesian *Cramer-Rao bound* (CRB) is a lower bound on the mean square error related to the estimation of a random continuous parameter. It results from the following inequality, derived by Van Trees [47]:

$$\mathbf{R}^{\mathbf{e}} - \mathbf{J}^{-1} \geq 0, \tag{3.21}$$

which holds if the a priori probability density function $p(\mathbf{u})$ of $\mathbf{u}$ is zero at the boundaries of its support $\mathcal{U}_1 \times ... \times \mathcal{U}_N$ ("*weak unbiasedness condition*"), in addition to some regularity conditions (details and proof see [48]). The notation $\mathbf{A} \geq 0$ in (3.21) indicates that $\mathbf{A}$ is a positive semi-definite matrix (i.e., $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$, irrespective of $\mathbf{x}$), $\mathbf{R}^{\mathbf{e}}$ is defined as in (3.12) and $\mathbf{J}^{-1}$ is the inverse of the Bayesian *Fisher Information Matrix* (FIM) $\mathbf{J}$. The (i,j)-th element of $\mathbf{J}$ is given by:

$$J_{i,j} = -E_{\mathbf{r},\mathbf{u}}\left[\frac{\partial^2 \ell(\mathbf{u};\mathbf{r})}{\partial u_i \partial u_j} + \frac{\partial^2 \ln p(\mathbf{u})}{\partial u_i \partial u_j}\right], \tag{3.22}$$

$$= E_{\mathbf{r},\mathbf{u}}\left[\frac{\partial \ell(\mathbf{u};\mathbf{r})}{\partial u_i}\frac{\partial \ell(\mathbf{u};\mathbf{r})}{\partial u_j} + \frac{\partial \ln p(\mathbf{u})}{\partial u_i}\frac{\partial \ln p(\mathbf{u})}{\partial u_j}\right], \tag{3.23}$$

where the expectation $E_{\mathbf{r},\mathbf{u}}[.]$ is with respect to $p(\mathbf{r},\mathbf{u})$, which is obtained as $p(\mathbf{r}|\mathbf{u})\,p(\mathbf{u})$.

The weak unbiasedness condition is not necessarily fulfilled. This is for example the case when the parameters u_i are uniformly distributed over finite observation intervals $\mathcal{U}_i$ as in frequency offset and phase shift estimation. On the other hand, the Bayesian Cramer-Rao inequality (3.21) holds at *high* signal-to-noise ratio (SNR) for *any* regular estimation problem, also if the weak unbiasedness condition is not fulfilled (proof see [48]).

Note that $\mathbf{J}$ is a symmetrical matrix consisting of two parts. The terms in (3.22) and (3.23) involving $\ell(\mathbf{u};\mathbf{r})$ relate to the information obtained from the observation $\mathbf{r}$. The terms in (3.22) and (3.23) involving $p(\mathbf{u})$ relate to the a priori information on $\mathbf{u}$. It is easily verified that the latter terms disappear when $\mathbf{u}$ is uniformly distributed over $\mathcal{U}_1 \times ... \times \mathcal{U}_N$, in which case the Bayesian FIM reduces to:

$$J_{i,j} = \int \left[\frac{\partial \ell(\mathbf{u};\mathbf{r})}{\partial u_i} \frac{\partial \ell(\mathbf{u};\mathbf{r})}{\partial u_j} \right] p(\mathbf{r}|\mathbf{u})\, p(\mathbf{u})\, \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{u}. \tag{3.24}$$

It easily follows from (3.21) that

$$E_{\mathbf{r},\mathbf{u}} \left[(e_i(\mathbf{r}))^2 \right] \geq CRB_i, \tag{3.25}$$

where the expectation $E_{\mathbf{r},\mathbf{u}}[.]$ is with respect to $p(\mathbf{r},\mathbf{u})$, and CRB_i is the ith diagonal element of the inverse of the FIM $\mathbf{J}$. The right-hand side of (3.25) is also known as the Bayesian Cramer-Rao bound: it imposes a fundamental lower bound on the MSE of *any* estimator of the ith parameter u_i.

The off-diagonal elements of the Bayesian FIM are an indication of the parameter coupling. When the (i,j)-th element of the Bayesian FIM is equal to zero, the parameters u_i and u_j are said to be *decoupled*, meaning that CRB_i is not affected by the inaccuracy in estimating u_j, and vice versa.

Any estimate that equals the Bayesian CRB in accuracy is called an *efficient* estimate. In [47], it has been shown that the *a posteriori* probability density function $p(\mathbf{u}|\mathbf{r})$ of $\mathbf{u}$ must be Gaussian for all $\mathbf{r}$ in order for an efficient estimate to exist; in addition, if an efficient estimate exists, it is $\hat{\mathbf{u}}_{MAP}(\mathbf{r})$. Because the MMSE estimate can not have a larger error this tells us that $\hat{\mathbf{u}}_{MAP}(\mathbf{r}) = \hat{\mathbf{u}}_{MMSE}(\mathbf{r})$ whenever an efficient estimate exists. When an efficient estimate does not exist we do not know how closely the MSE, using either $\hat{\mathbf{u}}_{MMSE}(\mathbf{r})$ or $\hat{\mathbf{u}}_{MAP}(\mathbf{r})$, approaches the lower bound. On the other hand, it can be shown that the MAP estimate is always *asymptotically* efficient, i.e., for a large enough number of observations (and a fixed value of the SNR); or for a large enough SNR (and a given number of observations) [47,48]. The latter property provides some additional motivation for using the MAP estimate even when an efficient estimate does not exist.

In the case where two random variables u_1 and u_2 are to be estimated from the observation $\mathbf{r}$, *two* CRBs on the MSE $E_{\mathbf{r},\mathbf{u}} \left[(e_1(\mathbf{r}))^2 \right]$ arise from (3.22). The first $\left(\mathbf{J}^{-1}\right)_{1,1}$ is obtained by considering the first diagonal element of the inverse of the 2×2 FIM related to the log-likelihood function

$\ell(u_1, u_2; \mathbf{r}) = \ln p(\mathbf{r}|u_1, u_2)$ of (u_1,u_2). The second is the inverse of the scalar FIM related to the marginal log-likelihood function $\ell(u_1; \mathbf{r}) = \ln p(\mathbf{r}|u_1)$ of u_1. It has been shown in [49] that the bound using $\ell(u_1)$ is always greater than or equal to the bound provided by $\left(\mathbf{J}^{-1}\right)_{1,1}$, and thus more tight. It should be noted, however, that the computational complexity associated with the evaluation of the corresponding CRBs is higher for $\ell(u_1; \mathbf{r})$ than for $\ell(u_1, u_2; \mathbf{r})$ because obtaining the latter likelihood function requires averaging over u_2. Similar remarks can be made in the N-dimensional case.

We will now consider the computation of the Bayesian FIM in the presence of nuisance parameters. Taking (3.1) into account we obtain from (3.24):

$$J_{i,j} = \int \left[\frac{\partial \ell(\mathbf{u}; \mathbf{r})}{\partial u_i} \frac{\partial \ell(\mathbf{u}; \mathbf{r})}{\partial u_j}\right] p(\mathbf{r}|\mathbf{v}, \mathbf{u}) p(\mathbf{u}) p(\mathbf{v}) \, \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}. \quad (3.26)$$

Differentiating (3.2) with respect to the ith component of $\mathbf{u}$ further yields:

$$\frac{\partial \ell(\mathbf{u}; \mathbf{r})}{\partial u_i} = \frac{1}{p(\mathbf{r}|\mathbf{u})} E_{\mathbf{v}}\left[\frac{\partial p(\mathbf{r}|\mathbf{v}, \mathbf{u})}{\partial u_i}\right], \quad (3.27)$$

$$= \int_{\mathbf{v}} \left(\frac{p(\mathbf{v})}{p(\mathbf{r}|\mathbf{u})} \frac{\partial p(\mathbf{r}|\mathbf{v}, \mathbf{u})}{\partial u_i}\right) \mathrm{d}\mathbf{v}, \quad (3.28)$$

where the integration is over all components of the nuisance parameter vector $\mathbf{v}$. When $\mathbf{v}$ contains discrete-valued components the corresponding integral in (3.28) and in subsequent equations has to be replaced with a summation. By taking into account that

$$\frac{\partial p(\mathbf{r}|\mathbf{v}, \mathbf{u})}{\partial u_i} = p(\mathbf{r}|\mathbf{v}, \mathbf{u}) \frac{\partial \ln p(\mathbf{r}|\mathbf{v}, \mathbf{u})}{\partial u_i}$$

and (Bayes rule)

$$\frac{p(\mathbf{v}) p(\mathbf{r}|\mathbf{v}, \mathbf{u})}{p(\mathbf{r}|\mathbf{u})} = p(\mathbf{v}|\mathbf{r}, \mathbf{u}) \quad (3.29)$$

the derivative of the log-likelihood function from (3.27) can be put into the following form:

$$\frac{\partial \ell(\mathbf{u}; \mathbf{r})}{\partial u_i} = \int_{\mathbf{v}} \left(p(\mathbf{v}|\mathbf{r}, \mathbf{u}) \frac{\partial \ln p(\mathbf{r}|\mathbf{v}, \mathbf{u})}{\partial u_i}\right) \mathrm{d}\mathbf{v}, \quad (3.30)$$

$$= E_{\mathbf{v}}\left[\frac{\partial \ell(\mathbf{v}, \mathbf{u}; \mathbf{r})}{\partial u_i} |\mathbf{r}, \mathbf{u}\right], \quad (3.31)$$

where the expectation $E_{\mathbf{v}}[.|\mathbf{r}, \mathbf{u}]$ denotes averaging with respect to the a posteriori probability $p(\mathbf{v}|\mathbf{r}, \mathbf{u})$ of the nuisance parameter vector $\mathbf{v}$, given $\mathbf{r}$ and $\mathbf{u}$. Substitution of (3.31) into (3.26) yields an exact expression of the Bayesian FIM in terms of the a posteriori average (over $\mathbf{v}$, for a given $\mathbf{r}$ and $\mathbf{u}$) of the derivatives (with respect to the components of $\mathbf{u}$) of the joint log-likelihood function $\ell(\mathbf{v}, \mathbf{u}; \mathbf{r})$ of $\mathbf{v}$ and $\mathbf{u}$. When $p(\mathbf{r}|\mathbf{v}, \mathbf{u})$ is Gaussian, the logarithm

$\ell(\mathbf{v},\mathbf{u};\mathbf{r}) = \ln p(\mathbf{r}|\mathbf{v},\mathbf{u})$ is readily available in closed-form. The feasibility of the evaluation of the Bayesian FIM hence entirely depends on the computational complexity associated with the calculation of the APPs $p(\mathbf{v}|\mathbf{r},\mathbf{u})$ in (3.30).

3.4.3 Modified Bayesian Cramer-Rao Bound

In order to avoid the computational complexity caused by the nuisance parameters, a *modified* Bayesian FIM (MFIM) $\mathbf{M}$ has been derived, for which the following holds (see [50, 51]):

$$\mathbf{R}^{\mathbf{e}} - \mathbf{M}^{-1} \geq \mathbf{R}^{\mathbf{e}} - \mathbf{J}^{-1} \geq 0. \tag{3.32}$$

The (i,j)-th element of $\mathbf{M}$ is given by:

$$M_{i,j} = -E_{\mathbf{r},\mathbf{v},\mathbf{u}}\left[\frac{\partial^2 \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_i \partial u_j} + \frac{\partial^2 \ln p(\mathbf{v},\mathbf{u})}{\partial u_i \partial u_j}\right], \tag{3.33}$$

$$= E_{\mathbf{r},\mathbf{v},\mathbf{u}}\left[\frac{\partial \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_i}\frac{\partial \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_j} + \frac{\partial \ln p(\mathbf{v},\mathbf{u})}{\partial u_i}\frac{\partial \ln p(\mathbf{v},\mathbf{u})}{\partial u_j}\right], \tag{3.34}$$

where $E_{\mathbf{r},\mathbf{v},\mathbf{u}}[.]$ denotes averaging with respect to $p(\mathbf{r},\mathbf{v},\mathbf{u})$, which is obtained as $p(\mathbf{r}|\mathbf{v},\mathbf{u})\,p(\mathbf{v},\mathbf{u})$. Note the similarities between (3.33) and (3.22), but while the FIM (3.22) is expressed in terms of the log-likelihood function $\ell(\mathbf{u};\mathbf{r})$ of $\mathbf{u}$, the MFIM is expressed in terms of the joint log-likelihood function $\ell(\mathbf{v},\mathbf{u};\mathbf{r})$ of $\mathbf{v}$ and $\mathbf{u}$, which is simpler to compute in the presence of nuisance parameters $\mathbf{v}$.

It easily follows from (3.32) that the MSE regarding the estimation of u_i is lower bounded by the so-called *modified* Bayesian CRB (MCRB), which is in general looser than the true Bayesian CRB:

$$E\left[(e_i(\mathbf{r}))^2\right] \geq CRB_i \geq MCRB_i, \tag{3.35}$$

where the expectation $E[.]$ is with respect to $p(\mathbf{r},\mathbf{u})$, CRB_i is the ith diagonal element of the inverse of the Bayesian FIM $\mathbf{J}$ from section 3.4.2, and $MCRB_i$ is the ith diagonal element of the inverse of the Bayesian MFIM $\mathbf{M}$.

When $p(\mathbf{r}|\mathbf{v},\mathbf{u})$ is Gaussian and $p(\mathbf{u})$ is uniform, the elements of the MFIM from (3.33)-(3.34) are much simpler than the elements of the FIM from (3.22)-(3.23) as far as analytical evaluation is concerned. To illustrate this, let us consider a complex-valued observation vector $\mathbf{r} = (r(0)\,r(1)\,...r(N_r-1))$, given by:

$$\mathbf{r} = \mathbf{s}(\mathbf{v},\mathbf{u}) + \mathbf{w}, \tag{3.36}$$

where $\mathbf{s}(.,.)$ is a known vector function and $\mathbf{w} = (w(0)\,w(1)\,...w(N_r-1))$ is a complex-valued white Gaussian noise vector with zero mean and covariance matrix $E\left[\mathbf{w}^H\mathbf{w}\right] = 2\sigma^2\mathbf{I}_{N_r}$, where $\mathbf{I}_k$ represents a $k \times k$ identity matrix. The a priori distribution $\mathbf{u}$ is assumed to be uniform. Taking also into account that $\mathbf{v}$

is independent of $\mathbf{u}$, the terms involving $p(\mathbf{v},\mathbf{u})$ in (3.33) and (3.34) disappear, and the (i,j)-th element of the Bayesian MFIM can be obtained as

$$M_{i,j} = \int \left[\frac{\partial \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_i} \frac{\partial \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_j} \right] p(\mathbf{r}|\mathbf{v},\mathbf{u})\, p(\mathbf{v},\mathbf{u})\, \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}. \quad (3.37)$$

The joint log-likelihood function $\ell(\mathbf{v},\mathbf{u};\mathbf{r})$ of $\mathbf{v}$ and $\mathbf{u}$ resulting from (3.36) is given by:

$$\ell(\mathbf{v},\mathbf{u};\mathbf{r}) = -\frac{1}{2\sigma^2} |\mathbf{r} - \mathbf{s}(\mathbf{v},\mathbf{u})|^2. \quad (3.38)$$

Differentiating (3.38) with respect to u_i yields:

$$\frac{\partial \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_i} = -\frac{1}{\sigma^2} \Re \left\{ (\mathbf{r} - \mathbf{s}(\mathbf{v},\mathbf{u})) \left(\frac{\partial \mathbf{s}(\mathbf{v},\mathbf{u})}{\partial u_i} \right)^H \right\}. \quad (3.39)$$

Substitution of (3.39) into (3.37) yields the MFIM. For given values of $\mathbf{u}$ and $\mathbf{v}$, the quantity $(\mathbf{r} - \mathbf{s}(\mathbf{v},\mathbf{u}))$ is a Gaussian random vector variable with zero mean and covariance matrix $2\sigma^2 \mathbf{I}_{N_r}$, such that

$$\int \left[\frac{\partial \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_i} \frac{\partial \ell(\mathbf{v},\mathbf{u};\mathbf{r})}{\partial u_j} \right] p(\mathbf{r}|\mathbf{v},\mathbf{u})\, \mathrm{d}\mathbf{r} = \frac{1}{\sigma^2} \Re \left\{ \left(\frac{\partial \mathbf{s}(\mathbf{v},\mathbf{u})}{\partial u_i} \right) \left(\frac{\partial \mathbf{s}(\mathbf{v},\mathbf{u})}{\partial u_j} \right)^H \right\}$$

and

$$M_{i,j} = \frac{1}{\sigma^2} \int \Re \left\{ \left(\frac{\partial \mathbf{s}(\mathbf{v},\mathbf{u})}{\partial u_i} \right) \left(\frac{\partial \mathbf{s}(\mathbf{v},\mathbf{u})}{\partial u_j} \right)^H \right\} p(\mathbf{v},\mathbf{u})\, \mathrm{d}\mathbf{v}\mathrm{d}\mathbf{u}, \quad (3.40)$$

$$= \frac{1}{\sigma^2} E_{\mathbf{v},\mathbf{u}} \left[\Re \left\{ \left(\frac{\partial \mathbf{s}(\mathbf{v},\mathbf{u})}{\partial u_i} \right) \left(\frac{\partial \mathbf{s}(\mathbf{v},\mathbf{u})}{\partial u_j} \right)^H \right\} \right]. \quad (3.41)$$

In (3.41), $E_{\mathbf{v},\mathbf{u}}[.]$ denotes averaging with respect to $p(\mathbf{v},\mathbf{u})$, i.e., to all parameters (both useful and nuisance).

3.5 Standard Estimation

This section briefly addresses the problem of standard (non-Bayesian) parameter estimation. In this case, the components of the wanted parameters $\mathbf{u}$ are considered to be continuous-valued and *nonrandom* (or *deterministic*).

The error correlation matrix that applies to the estimation of a deterministic vector parameter $\mathbf{u}$ is given by:

$$\mathbf{R}^{\mathbf{e}}(\mathbf{u}) = E_{\mathbf{r}|\mathbf{u}} \left[\mathbf{e}(\mathbf{r})\, \mathbf{e}^T(\mathbf{r}) \right],$$

with elements

$$R^{\mathbf{e}}_{i,j}(\mathbf{u}) = E_{\mathbf{r};\mathbf{u}} \left[e_i(\mathbf{r})\, e_j(\mathbf{r}) \right]. \quad (3.42)$$

In the above, the estimation error vector $\mathbf{e}(\mathbf{r})$ is defined as in (3.11) and the expectation $E_{\mathbf{r};\mathbf{u}}[.]$ denotes averaging over the conditional probability density function $p(\mathbf{r};\mathbf{u})$ of $\mathbf{r}$ for a given value of the deterministic parameter $\mathbf{u}$.

The ultimate goal in the standard estimation theory is the minimization of the estimator MSEs $R^{\mathbf{e}}_{i,i}(\mathbf{u})$ *for every value of* $\mathbf{u}$ and each $i = 1, 2, ..., N$. Unfortunately, $R^{\mathbf{e}}_{i,i}(\mathbf{u})$ is usually a function of $\mathbf{u}$, so that the MMSE estimator will be generally unrealizable in the standard framework. As an alternative to the direct minimization of the MSE, one usually focuses on *unbiased* estimators. These are estimators for which the average of the estimation error vector $\mathbf{e}(\mathbf{r})$ with respect to the conditional probability density function $p(\mathbf{r};\mathbf{u})$ equals zero. The MSEs of unbiased estimators coincide with their variances and the estimator that minimizes the variance subject to the unbiasedness constraint, i.e., $\int_{\mathbf{r}} \mathbf{e}(\mathbf{r})\, p(\mathbf{r};\mathbf{u})\, d\mathbf{r} = 0$ for every $\mathbf{u}$, is usually referred to as the *minimum variance unbiased* (MVU) estimator. Although the existence of the MVU estimator is not guaranteed, the MVU criterion is the primary estimation criterion in the standard approach to parameter estimation. One of the main reasons that the MVU formulation is so widely adopted is the fact that the ML estimation rule from (3.19) is known to approximately provide the MVU estimator under mild regularity conditions [47].

There also exists a standard version of the inequality from (3.21) that applies to deterministic parameters [47]:

$$\mathbf{R}^{\mathbf{e}}(\mathbf{u}) - \mathbf{J}^{-1}(\mathbf{u}) \geq 0. \tag{3.43}$$

This non-Bayesian inequality merely applies to unbiased estimators, but also holds for biased estimators in the limit of an infinite number of observations (i.e., as $N \to \infty$) [48]. In (3.43), $\mathbf{R}^{\mathbf{e}}(\mathbf{u})$ is the error correlation matrix as interpreted in the standard sense (see above) and $\mathbf{J}(\mathbf{u})$ denotes the standard FIM:

$$J_{i,j}(\mathbf{u}) = -E_{\mathbf{r};\mathbf{u}}\left[\frac{\partial^2 \ell(\mathbf{u};\mathbf{r})}{\partial u_i \partial u_j}\right], \tag{3.44}$$

$$= E_{\mathbf{r};\mathbf{u}}\left[\frac{\partial \ell(\mathbf{u};\mathbf{r})}{\partial u_i}\frac{\partial \ell(\mathbf{u};\mathbf{r})}{\partial u_j}\right], \tag{3.45}$$

with $E_{\mathbf{r};\mathbf{u}}[.]$ denoting averaging over the conditional probability density function $p(\mathbf{r};\mathbf{u})$. Note the similarity between (3.24) and (3.45). Usually it makes no sense to compare Bayesian FIMs with standard FIMs since the underlying experiments are inherently different. However, it follows directly from (3.24) and (3.45) that the Bayesian FIM for a *uniformly distributed* $\mathbf{u}$ is the *same as* the standard FIM $\mathbf{J}(\mathbf{u})$, provided that $\mathbf{J}(\mathbf{u})$ is actually *independent of* $\mathbf{u}$.

3.6 Conclusions and Remarks

In this chapter we have reviewed a number of elements from the detection and estimation theory that will prove to be essential for the main contributions

of this thesis (chapters 5 and 6). The analysis in section 3.3 constitutes the basis for the practical two-stage receiver structure that is commonly used in digital bandpass communication systems. First, an estimate $\hat{\mathbf{\Psi}}$ of the carrier parameter vector $\mathbf{\Psi}$ is derived from the received baseband signal $r_{BB}(t)$ (see (2.17)). This is referred to as *carrier synchronization*. Then the information bits $b(k)$ are detected, assuming ideal knowledge of the synchronization parameters, i.e., assuming that $\mathbf{\Psi} = \hat{\mathbf{\Psi}}$. This is referred to as *synchronized detection*. While synchronized detection is elaborated on in chapter 4, carrier synchronization will be further discussed in the chapters 5 and 6 of this thesis.

4 Synchronized Detection

In this chapter, we will apply the theoretical analysis from section 3.3 to the problem of detecting the information bits $b(k)$ from the baseband observation $r_{BB}(t)$ of a linearly modulated signal that is corrupted by additive white Gaussian noise (AWGN).

The outline is as follows. We start by presenting the *correct* observation model in section 4.1. Next, we introduce the concept of *synchronized detection* in section 4.2. The case of uncoded transmission is discussed in section 4.3, while coded transmission is subsequently dealt with in section 4.4. Finally, section 4.5 gives results for a *simplified* discrete-time observation model that is commonly used but ignores the reduction of the useful signal and the inter-symbol-interference (ISI) caused by a nonzero frequency offset.

4.1 Correct Observation Model

We recall from (2.17) that $r_{BB}(t)$ is given by:

$$r_{BB}(t) = \sqrt{E_s} \sum_{k \in I_s} a(k)\, p(t - kT)\, e^{j\Theta(t;\mathbf{\Psi})} + w_{BB}(t) \tag{4.1}$$

where

- E_s denotes the average symbol energy;
- $\mathbf{a} = (a(k) : k \in I_s)$ is a row-vector consisting of the N_s transmitted symbols. The symbol indices belong to the set I_s. For k belonging to the set of indices $I_p \subset I_s$, $a(k)$ is a pilot symbol, i.e., a priori known to the receiver. For k belonging to the set of indices $I_d = I_s \backslash I_p$ the symbol $a(k)$ is an unknown data symbol. There are $N_p \leq N_s$ pilot symbols and $N_d = N_s - N_p$ data symbols. The latter result from the encoding and mapping of N_b information bits $\mathbf{b} = (b(0)\ b(1)\ ...\ b(N_b - 1))$. The information bits are modeled as independent random variables with $\Pr[b(k) = 0] = \Pr[b(k) = 1] = 1/2$. The encoding rule, the mapping rule and the pilot symbol insertion rule are defined as $\mathbf{c} = \mathcal{C}(\mathbf{b})$, $\mathbf{d} = \mathcal{D}(\mathbf{c})$ or $d(k) = \mathcal{D}_m(\mathbf{c}_k)$, and $\mathbf{a} = \mathcal{A}(\mathbf{d}, \mathbf{p})$ respectively. Here, $\mathbf{c} = (c(0)\ c(1)\ ...\ c(N_c - 1))$ denotes a sequence of N_c coded bits with sub-vectors $\mathbf{c}_i = (c(im)\ c(im+1)\ ...\ c(im+m-1))$, $\mathbf{d} = (d(0)\ d(1)\ ...\ d(N_d - 1))$ denotes a sequence of N_d data symbols and $\mathbf{p} = (p(0)\ p(1)\ ...\ p(N_p - 1))$ denotes a sequence N_p of pilot symbols. We use an M-point signaling constellation which is denoted $\Omega = \{\omega_0, \omega_1, ..., \omega_{M-1}\}$, with $M = 2^m$.
- T is the symbol period;
- $\Theta(t; \mathbf{\Psi})$ is the instantaneous carrier phase shift of the received signal vis-a-vis the receiver's local reference carrier, and is fully specified by the parameter vector $\mathbf{\Psi}$;
- $p(t)$ is the transmit pulse with bandwidth B;
- $w_{BB}(t)$ is complex-valued zero-mean AWGN with independent real and imaginary parts each having a power spectral density equal to $N_0/2$ over the frequency interval $[-B, B]$.

4.2 Synchronized MAP Bit Detection

Following the reasoning from section 3.3, the synchronized bit detector uses an estimate $\hat{\mathbf{\Psi}}$ of the synchronization parameter vector $\mathbf{\Psi}$ and assumes that this estimate equals the correct value of $\mathbf{\Psi}$, i.e.,

$$\hat{\mathbf{\Psi}} \equiv \mathbf{\Psi}, \tag{4.2}$$

when producing, from $r_{BB}(t)$ and $\hat{\mathbf{\Psi}}$, a detected information bit sequence $\hat{\mathbf{b}} = (\hat{b}(0)\ \hat{b}(1)\ ...\ \hat{b}(N_b - 1))$.

The probability $\Pr\left[\hat{b}(k) \neq b(k)\right]$ that an information bit $b(k)$ will be received in error at the destination is called the *bit error rate* (BER). It provides a quantitative criterion to measure the performance of the system: the lower the BER, the more reliable the communication will be.

When (4.2) holds, the optimal bit detector in the sense of minimum BER is the maximum a posteriori *information-bit-by-information-bit* (MAP-i) detector, which (directly) computes the MAP decisions $\hat{b}(k)$ for each of the N_b bits of the information sequence $\mathbf{b}$. This type of detector is typically used for uncoded transmission, for convolutional codes and for turbo codes with convolutional constituent codes but not for LDPC codes. LDPC codes commonly use a MAP *coded-bit-by-coded-bit* (MAP-c) detector; basically, this is because sparse parity-check matrices typically result in a simple implementation of this type of detector. A MAP-c detector computes the MAP estimate $\hat{c}(k)$ for each of the N_c bits of the coded sequence $\mathbf{c}$ (as will be explained in section 4.4). The receiver then further extracts the estimate $\hat{\mathbf{b}}$ of $\mathbf{b}$ from $\hat{\mathbf{c}} = (\hat{c}(0)\ \hat{c}(1)\ ...\ \hat{c}(N_c-1))$ by a post-processing routine that reverses the processing of the encoding. This post-processing is trivial for a systematic code, since

$$\mathbf{b} = \mathbf{c} \begin{pmatrix} \mathbf{I}_{N_b} \\ \mathbf{0}_{(N_c-N_b)\times N_b} \end{pmatrix},$$

where $\mathbf{I}_n$ is a $n \times n$ identity matrix and $\mathbf{0}_{n\times m}$ is an $n \times m$ all-zero matrix. In this case, the MAP-i detector and the MAP-c detector are equivalent. In most other cases, MAP-c detection does not result in the minimum BER when (4.2) holds.

Let us first consider the MAP-i detector. Taking in (3.7)-(3.9) $\mathbf{u} = \mathbf{b}$ and $\mathbf{v} = \mathbf{\Psi}$ yields:

$$\hat{b}(k) \quad = \quad \arg\max_{b(k)} p\left(b(k) \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right), \tag{4.3}$$

with

$$p\left(b(k) \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right) = \sum_{\mathbf{b}\setminus b(k)} p\left(\mathbf{b} \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right) \tag{4.4}$$

and

$$p\left(\mathbf{b} \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right) \propto p\left(\mathbf{r} \left| \hat{\mathbf{\Psi}}, \mathbf{b} \right.\right) p(\mathbf{b}), \tag{4.5}$$

where $\mathbf{r}$ denotes a vector representation of the signal $r_{BB}(t)$ from (4.1) (e.g., $\mathbf{r}$ consists of samples $r_{BB}(kT_{sample})$ with $2BT_{sample} \leq 1$).

It follows directly from (4.1) that the conditional probability density function $p\left(\mathbf{r} \left| \hat{\mathbf{\Psi}}, \mathbf{b} \right.\right)$ of $\mathbf{r}$ given $\mathbf{b}$ and $\hat{\mathbf{\Psi}}$ is given by:

$$p\left(\mathbf{r} \left| \hat{\mathbf{\Psi}}, \mathbf{b} \right.\right) \quad = C \quad \prod_{k\in I_d} F\left(a(k;\mathbf{b}), z\left(k; \mathbf{r}, \hat{\mathbf{\Psi}}\right)\right), \tag{4.6}$$

where C is a constant not depending on $\mathbf{b}$, $a(k;\mathbf{b})$ denotes the kth component of the symbol sequence $\mathbf{a}$ that corresponds to the bit sequence $\mathbf{b}$ (i.e., $\mathbf{a} = \mathcal{A}(\mathcal{D}(\mathcal{C}(\mathbf{b})), \mathbf{p})$), the function $F(.,.)$ is defined as:

$$F(x,y) = exp\left(\frac{E_s}{N_0}\left(2\Re(x^* y) - |x|^2\right)\right) \tag{4.7}$$

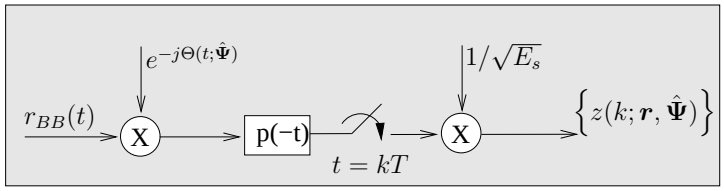

Figure 4.1: Matched filter output sample generation.

and the quantities $z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)$ are given by:

$$z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right) = \frac{1}{\sqrt{E_s}}\int_{-\infty}^{\infty} r_{BB}(t)\, p(t-kT)\, e^{-j\Theta\left(t;\hat{\mathbf{\Psi}}\right)}\mathrm{d}t. \tag{4.8}$$

As indicated in Fig. 4.1, the quantity $z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)$ from (4.8) is obtained by applying to $r_{BB}(t)$ an instantaneous rotation of $-\Theta\left(t;\hat{\mathbf{\Psi}}\right)$ rad, feeding the result to a filter $p(-t)$ *'matched'* to the transmit pulse $p(t)$, sampling the matched filter output at instant kT, and scaling the resulting samples down by a factor $\sqrt{E_s}$.

The conceptual block diagram of the resulting procedure is shown in Fig. 4.2. Once an estimate $\hat{\mathbf{\Psi}}$ of the parameter vector $\mathbf{\Psi}$ has been generated, the matched filter output samples $z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)$, $k \in I_d$ are produced and used to compute the required marginal bit APPs $p\left(b(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$ according to (4.4)-(4.5). Finally, a decision device produces the bit estimates $\hat{b}(k)$ according to (4.3). The multiplications with $e^{-j\Theta\left(t;\hat{\mathbf{\Psi}}\right)}$ in Fig. 4.2 constitutes the carrier phase shift correction made by the receiver, so as to compensate for the fact that the demodulation uses the carrier signal $e^{-j2\pi f_c t}$ rather than the ideal carrier signal $e^{-j(2\pi f_c t+\Theta(t;\mathbf{\Psi}))}$ (see chapter 2, section 2.4).

Introducing the estimation error $e_\Theta(t) = \Theta(t;\mathbf{\Psi}) - \Theta\left(t;\hat{\mathbf{\Psi}}\right)$, the quantities $z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)$ can be decomposed as:

$$z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right) = \sum_n a(n)\,\mathcal{I}(k,n) + w(k), \tag{4.9}$$

where

$$\mathcal{I}(k,n) = \int_{-\infty}^{\infty} p(t-nT)\, p(t-kT)\, e^{je_\Theta(t)}\mathrm{d}t \tag{4.10}$$

and $\mathbf{w} = (w(k) : k \in I_s)$ is a sequence of independent and identically distributed (i.i.d.) complex-valued white Gaussian random variables with i.i.d. real and imaginary parts, each having a power spectral density equal to $N_0/2E_s$. Assuming that the estimation error $e_\Theta(t)$ remains essentially constant over a time

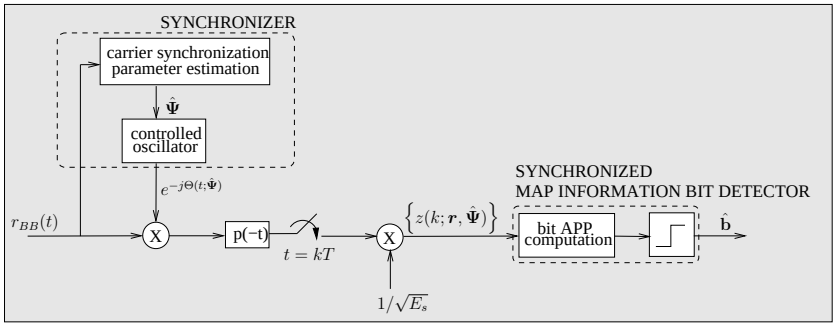

Figure 4.2: Receiver with synchronized MAP-i detector, resulting from the correct continuous-time observation model.

interval equal to the duration of the pulse $p(t)$ (which amounts to a few symbol intervals), we can safely replace in (4.10) $e_\Theta(t)$ by $e_\Theta(kT)$, which yields

$$\mathcal{I}(k,n) = e^{je_\Theta(kT)}\delta(k-n)$$

and

$$z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right) \quad = \quad a(k)\,e^{je_\Theta(kT)} + w(k)\,. \tag{4.11}$$

We observe that the performance of the detector is affected by the estimation errors $e_\Theta(kT)$, $k \in I_d$.

The summation in (4.4) contains 2^{N_b-1} terms, one for each of the 2^{N_b-1} possible combinations of (N_b-1) information bits. This would imply that the computational complexity of the "bit APP computation" function of the synchronized detector in Fig. 4.2 increases exponentially with the information sequence length N_b. However, as indicated in chapter 3, the marginal bit APP $p\left(b(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$ can be computed more efficiently, either exactly or approximately, by applying the *sum-product algorithm* to a *factor graph* representing a suitable factorization of the joint bit APP $p\left(\mathbf{b}\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$ from (4.5) [45].

It is obvious that the bit APPs $p\left(b(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$ are marginals of the joint APP $p\left(\mathbf{b}\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$ from (4.5), but the latter function is not easily factorized. Therefore, we introduce additional variables (grouped into a vector, say $\mathbf{x} = \{x_l\}$), in such a way that $p\left(\mathbf{b},\mathbf{x}\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$ leads to a convenient factor graph representation. Applying the sum-product algorithm on this graph then yields the required bit APPs $p\left(b(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$, as well as a posteriori probabilities of the additional variables $p\left(x(l)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right)$.

Let us introduce the vector variables $\mathbf{c} = \mathcal{C}(\mathbf{b})$ with sub-vectors $\mathbf{c}_k$ and $\mathbf{d}$ with components $d(k) = \mathcal{D}_m(\mathbf{c}_k)$. Taking into account the index bijec-

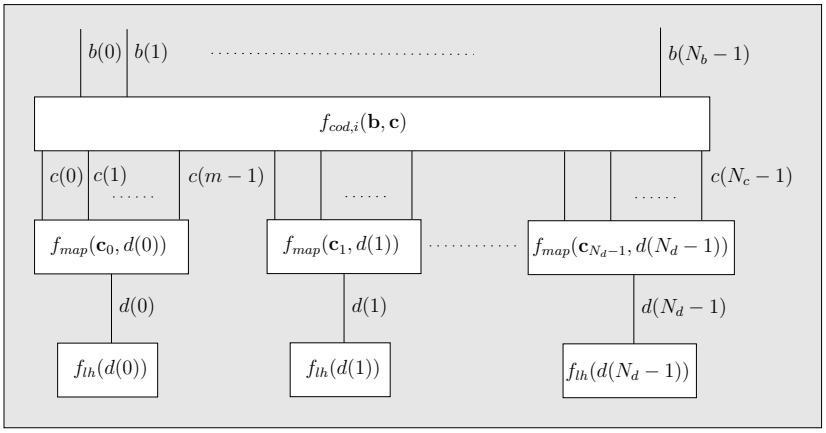

Figure 4.3: Factor graph of a synchronized MAP-i detector, for the case of a generic channel code and M-ary symbol mapping ($m = log_2 M$).

tions $\gamma_d(.)$ and $\gamma_p(.)$ from (2.10), which control the pilot symbol insertion rule $\mathcal{A}(\mathbf{d}, \mathbf{p})$, we obtain:

$$p\left(\mathbf{b}, \mathbf{c}, \mathbf{d} \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right) \quad \propto \quad f_{cod,i}(\mathbf{b}, \mathbf{c}) \prod_{k=0}^{N_d-1} f_{map}(\mathbf{c}_k, d(k)) f_{lh}(d(k)), \quad (4.12)$$

where we have dropped all factors not depending on $(\mathbf{b}, \mathbf{c}, \mathbf{d})$. The left-most factor enforces the code constraint $f_{cod,i}(\mathbf{b}, \mathbf{c}) = I[\mathbf{c} = \mathcal{C}(\mathbf{b})]$ to the information bits $\mathbf{b}$ and the coded bits $\mathbf{c}$, the factors marked $f_{map}(.,.)$ enforce the mapping constraints $f_{map}(\mathbf{c}_k, d(k)) = I[d(k) = \mathcal{D}_m(\mathbf{c}_k)]$ to the coded bits and the data symbols, and the right-most factors correspond to $f_{lh}(d(k)) = F\left(d(k), z\left(\gamma_d(k); \mathbf{r}, \hat{\mathbf{\Psi}}\right)\right)$, which is the likelihood of the data symbols $d(k)$ as a function of the associated (scaled an rotated) matched filter output sample $z(.;.,.)$. The factor graph representing the factorization from (4.12) is shown in Fig. 4.3. Only matched filter outputs corresponding to data symbol positions are exploited, as pilot symbols provide no information about the information bits.

A similar discussion for the MAP-c detector leads to:

$$\hat{c}(k) \quad = \quad \arg\max_{c(k)} p\left(c(k) \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right), \quad (4.13)$$

where the coded bit APPs $p\left(c(k) \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right)$ are marginals of the joint APP $p\left(\mathbf{c} \left| \mathbf{r}, \hat{\mathbf{\Psi}} \right.\right)$ which can be factorized as:

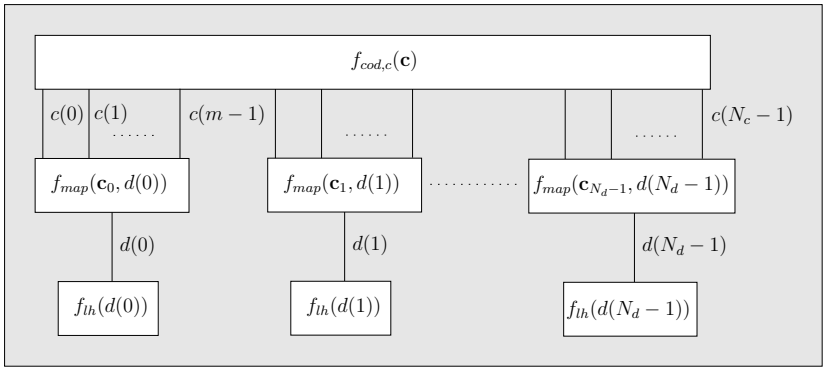

Figure 4.4: Factor graph of a synchronized MAP-c detector, for the case of a generic channel code and M-ary symbol mapping ($m = log_2 M$).

$$p\left(\mathbf{c}, \mathbf{d} \middle| \mathbf{r}, \hat{\mathbf{\Psi}}\right) \quad \propto \quad f_{cod,c}\left(\mathbf{c}\right) \prod_{k=0}^{N_d-1} f_{map}\left(\mathbf{c}_k, d\left(k\right)\right) f_{lh}\left(d\left(k\right)\right), \qquad (4.14)$$

where we have introduced N_d additional variables $\mathbf{d} = (d\,(0)\; d\,(1)\; ...\; d\,(N_d-1))$ and "$\propto$" denotes equality up to an irrelevant factor not depending on $(\mathbf{c}, \mathbf{d})$. The left-most factor in (4.14) enforces the code constraint $f_{cod,c}\left(\mathbf{c}\right) = I[\mathbf{c} \in \zeta_{\mathcal{C}}]$ to the coded bits $\mathbf{c}$, whereas the factors marked $f_{map}\left(.,.\right)$ and $f_{lh}\left(.\right)$, are defined as in (4.12). The factor graph representing the factorization from (4.14) is shown in Fig. 4.4 and the conceptual block diagram of the resulting receiver is shown in Fig. 4.5.

Cycles are present between the mapper nodes and the code constraint nodes in Fig. 4.3 and Fig. 4.4 when $m > 1$. The sum-product algorithm applied to these factor graphs will therefore generally be iterative and suboptimal.

The computational efficiency of many advanced decoding algorithms originates from a further decomposition of the code constraint nodes based on a suitable factorization of $f_{cod,i}\left(\mathbf{b}, \mathbf{c}\right) = I[\mathbf{c} = \mathcal{C}\left(\mathbf{b}\right)]$ or $f_{cod,c}\left(\mathbf{c}\right) = I[\mathbf{c} \in \zeta_{\mathcal{C}}]$, through the introduction of more additional variables.

4.3 Uncoded Transmission

In the special case of uncoded transmission, the information sequence $\mathbf{b}$ and the coded sequence $\mathbf{c}$ are the same, so that the MAP-i detector and the MAP-c detector yield the same result. The code constraint $I\left[\mathbf{c} = C\left(\mathbf{b}\right)\right]$ reduces to

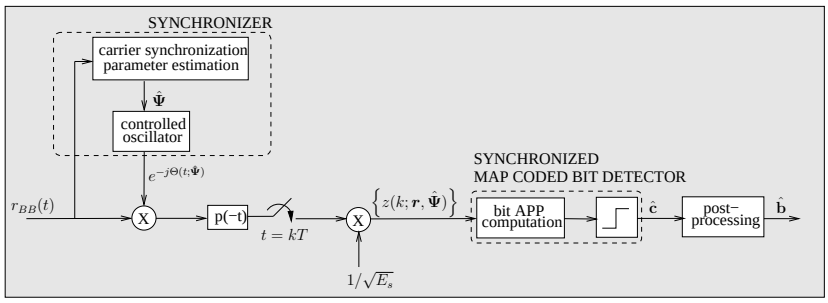

Figure 4.5: Receiver with synchronized MAP-c detector, resulting from the correct continuous-time observation model.

$I[\mathbf{c}=\mathbf{b}]$, or, equivalently,

$$f_{cod,i}(\mathbf{b},\mathbf{c}) = \prod_{k=0}^{N_b-1} I[c(k)=b(k)]. \tag{4.15}$$

Taking (4.15) into account, the factor graph of the MAP-i detector, representing the factorization (4.12), can be further decomposed as is shown in Fig. 4.6. As there are no cycles, application of the sum-product algorithm to this graph yields the exact marginal bit and symbol APPs. We obtain

$$\begin{aligned} &p\left(b(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right) \\ &\propto \textstyle\sum_{\mathbf{b}_{\lfloor\frac{k}{m}\rfloor}\backslash b(k)} F\left(\mathcal{D}_k\left(\mathbf{b}_{\lfloor\frac{k}{m}\rfloor}\right), z\left(\gamma_d\left(\lfloor\frac{k}{m}\rfloor\right);\mathbf{r},\hat{\mathbf{\Psi}}\right)\right) \end{aligned} \tag{4.16}$$

and (as $d\left(\gamma_d^{-1}(k)\right) = a(k)$ for $k \in I_d$)

$$\begin{aligned} &p\left(a(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right) \\ &\propto F\left(a(k), z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)\right),\ k \in I_d. \end{aligned} \tag{4.17}$$

In (4.16), $\mathbf{b}_i = (b(im)\ b(im+1)\ ...\ b(im+m-1))$ and $\lfloor\frac{k}{m}\rfloor = \text{floor}\left(\frac{k}{m}\right)$ denotes the index (before pilot symbol insertion) of the data symbol containing the kth information bit $b(k)$; $m = log_2 M$ denotes the number of bits per data symbol.

Assuming Gray mapping and high SNR, the BER resulting from (4.3),(4.16) is well approximated by [33]:

$$BER = \frac{SER}{m}, \tag{4.18}$$

where SER denotes the symbol error rate (SER) resulting from MAP data symbol detection:

$$SER \quad = \quad \Pr\left[\hat{a}(k) \neq a(k) :\ k \in I_d\right], \tag{4.19}$$

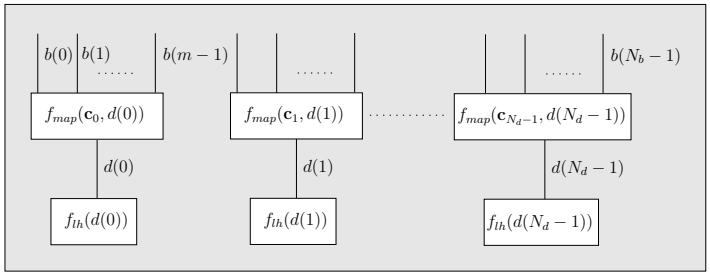

Figure 4.6: Factor graph of a synchronized MAP-i detector for the case of uncoded transmission with M-ary symbol mapping.

with

$$\hat{a}\left(k\right) \quad = \quad \arg\max_{a(k)} p\left(a\left(k\right)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right), \;\; k \in I_d. \tag{4.20}$$

The SER (4.19) is well approximated by [33]:

$$SER \quad = \quad n_M Q\left(\sqrt{\frac{E_s}{2N_0}\left(d_M\right)^2}\right), \tag{4.21}$$

where d_M and n_M denote the minimum Euclidean distance between the constellation points (normalized constellation) and the average number of constellation points at minimum distance, respectively. The function $Q\left(.\right)$ in (4.21) is defined as:

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-\frac{v^2}{2}} \mathrm{d}v \tag{4.22}$$

and represents the area under the tail of the probability density function of a Gaussian random variable with zero mean and unit variance. For M-PAM, M-QAM and M-PSK constellations, one obtains:

$$M-PAM: \;\; n_M = \frac{2\left(M-1\right)}{M},\, d_M = \sqrt{\frac{12}{M^2-1}}, \tag{4.23}$$

$$M-QAM: \, n_M = \frac{4\left(\sqrt{M}-1\right)}{\sqrt{M}},\, d_M = \sqrt{\frac{6}{M-1}}, \tag{4.24}$$

$$M-PSK \;\left(M>2\right): \quad n_M = 2,\, d_M = \sin\frac{\pi}{M}. \tag{4.25}$$

The quantity n_M ranges from $n_M = 1$ (for 2-PAM) to $n_M = 4$ (for an infinitely large QAM constellation). The SER from (4.21) is therefore mainly determined by the value of $E_s d_M^2/N_0$. In Fig. 4.7 we have depicted the uncoded SER from (4.21) as a function of the SNR E_s/N_0 for 2-, 4- and 8-PAM, 4-, 16-

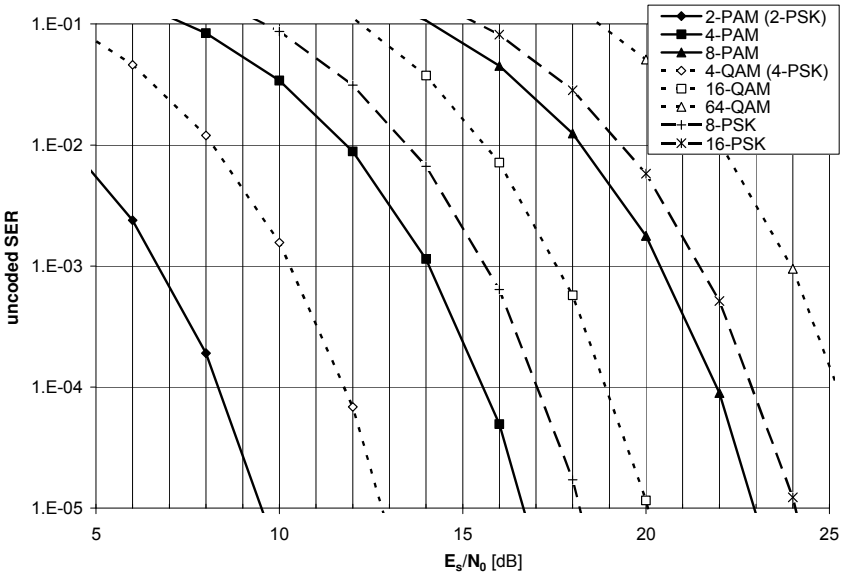

Figure 4.7: Uncoded SER as a function of E_s/N_0 for 2-, 4- and 8-PAM, 4-, 16- and 64-QAM, and 8- and 16-PSK signal constellations.

and 64-QAM, and 8- and 16-PSK. For a given constellation type (PAM, QAM or PSK), the minimum Euclidean distance between constellation points d_M decreases and therefore the BER increases when the number of constellation points M increases.

4.4 Coded Transmission

There is obviously more than one way to decompose the code constraint nodes in Fig. 4.3 and Fig. 4.4 by means of a factorization of $f_{cod,i}(\mathbf{b}, \mathbf{c}) = I[\mathbf{c} = \mathcal{C}(\mathbf{b})]$ or $f_{cod,c}(\mathbf{c}) = I[\mathbf{c} \in \zeta_{\mathcal{C}}]$ through the introduction of additional variables (or edges) [45]. We show in Appendix B on page 187 how different factor graphs can be constructed from

1. the generator matrix $\mathbf{G}$ of the code;
2. the parity-check matrix $\mathbf{H}$ of the code;
3. the trellis of the code.

The matrix-based methods (1. and 2.) yield factor graphs that contain as many additional *binary*-valued variables as there are "1"-elements in the producing matrix; in general, these graphs have cycles. Graphs resulting from generator

matrices represent a factorization of $f_{cod,i}(\mathbf{b},\mathbf{c}) = I[\mathbf{c} = \mathcal{C}(\mathbf{b})]$, whereas graphs resulting from parity-check matrices represent a factorization of $f_{cod,c}(\mathbf{c}) = I[\mathbf{c} \in \zeta_{\mathcal{C}}]$. The latter kind of graphs are commonly known as *Tanner graphs*. The trellis-based method (3.) yields *cycleless* factor graphs; the number of additional variables equals the number of sections in the producing trellis and each additional variable may take as many values as there are states in the trellis. Graphs resulting from the trellis [37] of a convolutional code represent a factorization of $f_{cod,i}(\mathbf{b},\mathbf{c}) = I[\mathbf{c} = \mathcal{C}(\mathbf{b})]$, whereas graphs resulting from the trellis [52] that has been derived from the parity-check matrix of an arbitrary linear code represent a factorization of $f_{cod,c}(\mathbf{c}) = I[\mathbf{c} \in \zeta_{\mathcal{C}}]$.

Each of these factor graphs results in a different sum-product algorithm, or equivalently, in a different way to approximate the marginal bit APPs

$$p\left(b(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right),\ p\left(c(k)\left|\mathbf{r},\hat{\mathbf{\Psi}}\right.\right),$$

that occur in the MAP bit detection rules from (4.3) and (4.13). Choosing a suitable factor graph is therefore a matter of trading off the accuracy of the resulting marginal bit APPs against the computational complexity associated with their computation. While the accuracy of the sum-product algorithm is determined by the minimum cycle length (i.e., the girth) of the graph, the complexity of the procedure primarily depends on the number of edges and the size of the variable alphabets in the graph. Some practical examples:

- The standard detection algorithm for LDPC codes is the MAP-c detection algorithm and results from replacing in Fig. 4.4 the code constraint node by the Tanner graph of the code. The defining property of an LDPC code is that the parity-check matrix is sparse: the number of additional variables in the Tanner graph is small. In order to improve their decoding performance, LDPC codes are designed to exhibit a large girth.

- The well-known *BCJR* detection algorithm for convolutional codes [53], named after its inventors Bahl, Cocke, Jelinek and Raviv, is a MAP-i detection algorithm and results from replacing in Fig. 4.3 the code constraint node by the cycleless trellis-based graph. Convolutional codes are typically constructed using a small number of shift register stages: the size of the variable alphabets in the trellis-based graph is feasible.

- The standard detection algorithm for turbo codes follows from applying the sum-product algorithm to the concatenated factor graph consisting of x sub-graphs (one for each of the x constituent codes), which are interconnected by $(x-1)$ random interleavers. When these sub-graphs represent a factorization of $f_{cod,i}(\mathbf{b},\mathbf{c}) = I[\mathbf{c} = \mathcal{C}(\mathbf{b})]$ ($f_{cod,c}(\mathbf{c}) = I[\mathbf{c} \in \zeta_{\mathcal{C}}]$) the resulting detector is a MAP-i detector (MAP-c detector). As an example, Fig. 4.8 shows the factor graph of the parallel concatenated turbo code of Fig. 2.6, where $I\left[\mathbf{c}^{(2)} = \mathcal{C}_{conv1}(\mathbf{b})\right]$ and $I\left[\mathbf{c}^{(3)} = \mathcal{C}_{conv2}(\mathbf{b}')\right]$ denote the encoding rules of the two constituent convolutional encoders (convolutional encoder 1 and 2 respectively) and $\mathbf{b}'$ represents the interleaved

version of the information bit sequence **b**. By replacing in Fig. 4.3 the code constraint node by the factor graph shown in Fig. 2.6 we obtain a MAP-i detector.

Fig. 4.9 shows the BER (dashed lines) and the SER (solid lines) resulting from the standard decoding techniques for turbo or LDPC coded 2-PSK (2-PAM) and 4-PSK (4-QAM), assuming perfect synchronization (i.e., $\hat{\mathbf{\Psi}} \equiv \mathbf{\Psi}$), as a function of E_s/N_0. The following channel codes are considered:

- A turbo (T) code that consists of the parallel concatenation of two identical 16-state rate-1/2 recursive systematic convolutional codes with generator sequences $(37)_8$ and $(21)_8$ (in octal notation) through a uniform random interleaver of length $I = N_b$; the parity bits are punctured to obtain an overall rate of r.
- A systematic rate 1/2 LDPC code with a block length of $2N_b = 1000$ bits.

For the sake of comparison the result for uncoded (U) 2-PSK and 4-PSK is also displayed. We observe that the normal operating SNR of the perfectly synchronized receiver (say, BER $< 10^{-3}$) is significantly lower in the case of turbo or LDPC coded transmission than in the case of uncoded transmission.

4.5 Simplified Observation Model

When the time variations of the carrier phase shift $\Theta(t; \mathbf{\Psi})$ are slow in comparison to the symbol rate $1/T$, it is not unusual that the observed baseband signal $r_{BB}(t)$ from (4.1) is matched-filtered and sampled *before* a carrier phase shift correction is applied. As indicated in Fig. 4.10, the discrete-time observation at the output of the matched filter then consists of the samples $\mathbf{x} = (x(k) : k \in I_s)$, which are given by:

$$x(k) = \int_{-\infty}^{\infty} r_{BB}(t)\, p(t - kT)\, \mathrm{d}t. \tag{4.26}$$

Taking into account (4.1), this yields:

$$x(k) = \sqrt{E_s} \sum_{n \in I_s} a(n)\, \mathcal{I}(k, n) + w(k), \tag{4.27}$$

where

$$\mathcal{I}(k, n) = \int_{-\infty}^{\infty} p(t - nT)\, p(t - kT)\, e^{j\Theta(t;\mathbf{\Psi})} \mathrm{d}t \tag{4.28}$$

and $\mathbf{w} = (w(k) : k \in I_s)$ is a sequence of independent and identically distributed (i.i.d.) complex-valued white Gaussian random variables with i.i.d. real and imaginary parts, each having a power spectral density equal to $N_0/2$. Due to the uncompensated phase shift $\Theta(t; \mathbf{\Psi})$, $\mathcal{I}(k, n)$ does not equal $\delta(k - n)$: this results in a power loss (because $|\mathcal{I}(k, k)| \leq 1$) and also in the introduction of

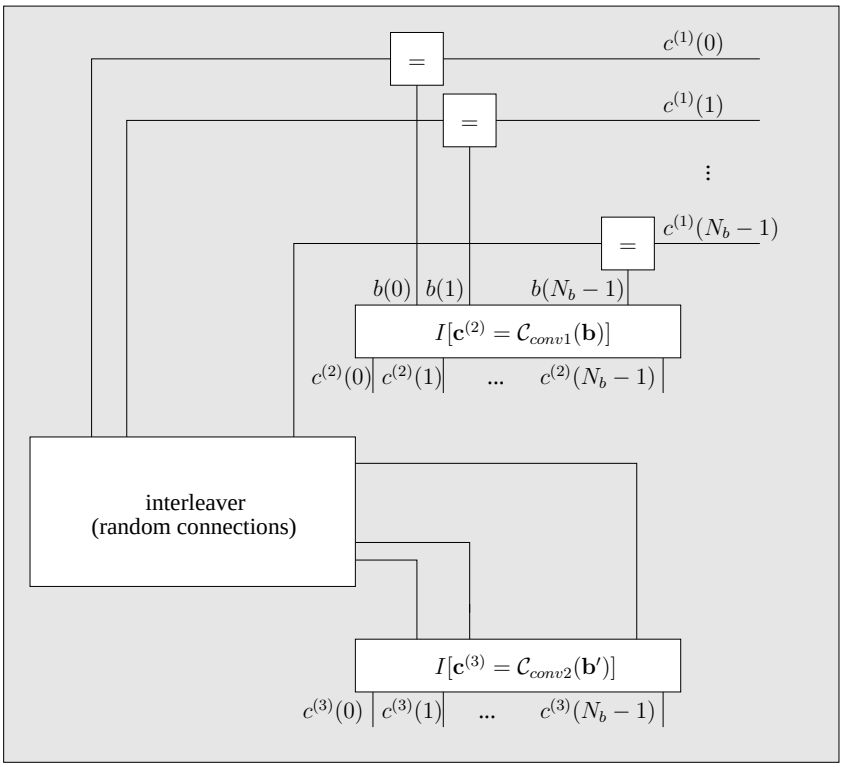

Figure 4.8: Decomposition of the code constraint node for the case of turbo coded transmission.

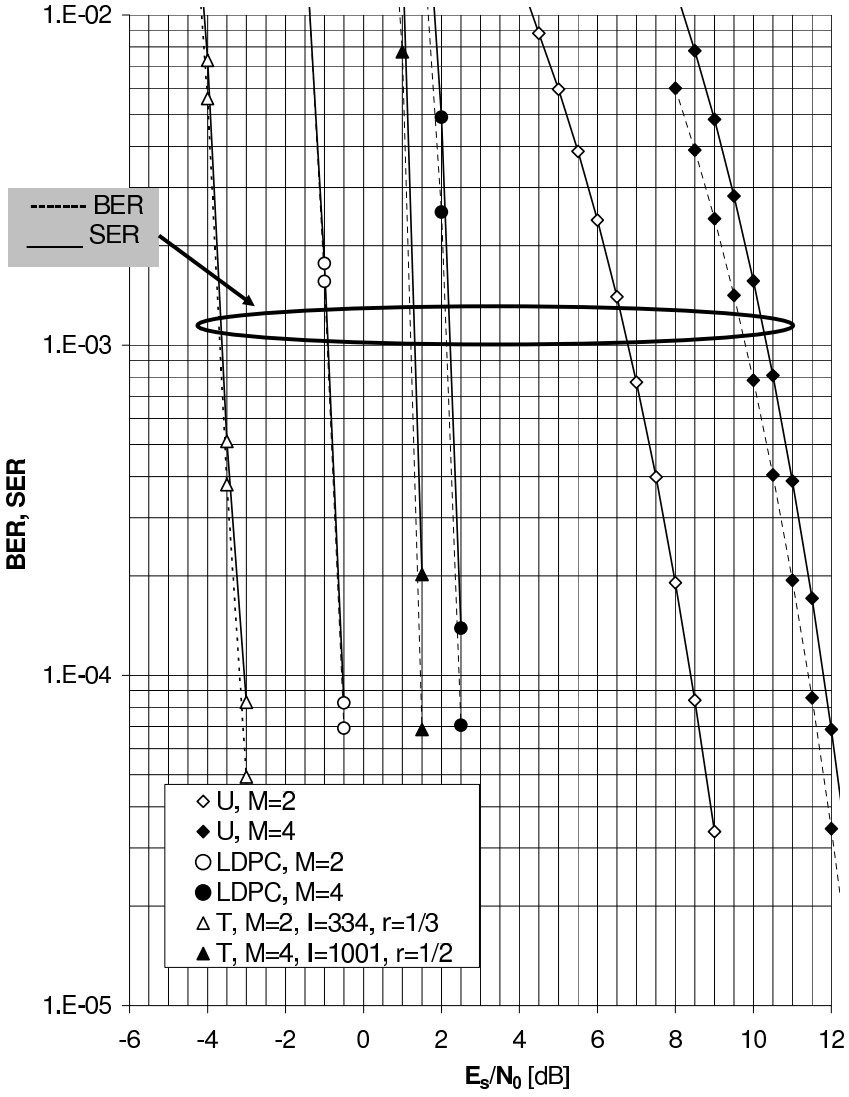

Figure 4.9: BER and SER as obtained with MAP bit detection of uncoded (U), turbo (T) coded or LDPC coded M-PSK ($M \in \{2, 4\}$).

inter-symbol-interference (ISI) at the matched filter output (because $\mathcal{I}(k,n) \neq 0$ for $k \neq n$).

A commonly used simplified (si) observation model $\mathbf{r}^{(si)} = \left(r^{(si)}(k) : k \in I_s\right)$ results from the assumption that the time variations of the carrier phase shift $\Theta(t;\mathbf{\Psi})$ in (4.1) are so small that $e^{j\Theta(t;\mathbf{\Psi})}$ does not change noticeably over the duration of the pulse $p(t)$. In this case, $e^{j\Theta(t;\mathbf{\Psi})}$ in (4.28) is well approximated by $e^{j\Theta(kT;\mathbf{\Psi})}$. This yields:

$$r^{(si)}(k) = \sqrt{E_s}a(k)\,e^{j\Theta(kT;\mathbf{\Psi})} + w(k). \tag{4.29}$$

Similarly as for the correct observation model, a MAP-i detector and a MAP-c detector can now be derived from (4.29). For the MAP-i detector, we obtain:

$$\hat{b}(k) = \arg\max_{b(k)} p\left(b(k)\left|\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right.\right), \tag{4.30}$$

with

$$p\left(b(k)\left|\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right.\right) = \sum_{\mathbf{b}\backslash b(k)} p\left(\mathbf{b}\left|\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right.\right) \tag{4.31}$$

and

$$p\left(\mathbf{b}\left|\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right.\right) \propto p\left(\mathbf{r}^{(si)}\right|\hat{\mathbf{\Psi}}, \mathbf{b}\left.\right) p(\mathbf{b}). \tag{4.32}$$

It follows directly from (4.29) that the conditional probability density function $p\left(\mathbf{r}^{(si)}\right|\mathbf{\Psi}, \mathbf{b}\left.\right)$ of $\mathbf{r}^{(si)}$ given $\mathbf{b}$ and $\hat{\mathbf{\Psi}}$, is given by:

$$p\left(\mathbf{r}^{(si)}\right|\hat{\mathbf{\Psi}}, \mathbf{b}\left.\right) = C \prod_{k\in I_d} F\left(a(k;\mathbf{b}), z^{(si)}\left(k;\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right)\right), \tag{4.33}$$

where C is a constant not depending on $\mathbf{b}$, $a(k;\mathbf{b})$ denotes the kth component of the symbol sequence $\mathbf{a}$ that corresponds to $\mathbf{b}$ (i.e., $\mathbf{a} = \mathcal{A}(\mathcal{D}(\mathcal{C}(\mathbf{b})), \mathbf{p})$), the function $F(.,.)$ is defined as in (4.7), and $z^{(si)}\left(k;\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right)$ is given by:

$$z^{(si)}\left(k;\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right) = \frac{1}{\sqrt{E_s}} r^{(si)}(k)\, e^{-j\Theta\left(kT;\hat{\mathbf{\Psi}}\right)}. \tag{4.34}$$

Note the similarity between the expressions for $p\left(\mathbf{r}^{(si)}\left|\hat{\mathbf{\Psi}}, \mathbf{b}\right.\right)$ from (4.33) and $p\left(\mathbf{r}\left|\hat{\mathbf{\Psi}}, \mathbf{b}\right.\right)$ from (4.6).

The conceptual block diagram of this procedure is shown in Fig. 4.11. As far as the computation of marginal bit APPs is concerned, we only need to replace the correct matched filter output samples $z\left(k;\mathbf{r}, \hat{\mathbf{\Psi}}\right)$ from (4.8) with their simplified equivalents $z^{(si)}\left(k;\mathbf{r}^{(si)}, \hat{\mathbf{\Psi}}\right)$ from (4.34). It should be noted that the latter quantities yield the same decomposition as the correct matched

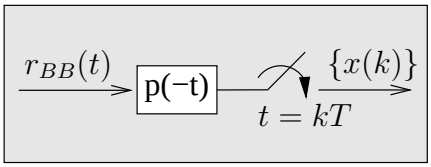

Figure 4.10: Alternative matched filter output sample generation.

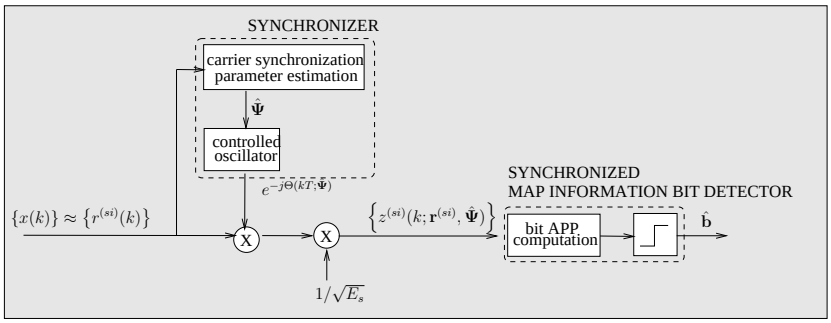

Figure 4.11: Receiver with synchronized MAP-i detector, resulting from the simplified continuous-time observation model.

filter output samples $z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)$ when the time variations of the estimation error $e_\Theta(t) = \Theta(t;\mathbf{\Psi}) - \Theta\left(t;\hat{\mathbf{\Psi}}\right)$ are slow as compared to the symbol rate $1/T$:

$$z^{(si)}\left(k;\mathbf{r}^{(si)},\hat{\mathbf{\Psi}}\right) = a(k)\, e^{je_\Theta(kT)} + w(k), \tag{4.35}$$

with $\mathbf{w} = (w(k) : k \in I_s)$ a sequence of i.i.d. complex-valued white Gaussian random variables with i.i.d. real and imaginary parts, each having a power spectral density equal to $N_0/2E_s$.

The above analysis remains valid for the MAP-c detector provided that we replace $\left\{\hat{b}(k), b(k), \mathbf{b}\right\}$ with $\{\hat{c}(k), c(k), \mathbf{c}\}$ and let $a(k;\mathbf{c})$ denote the kth component of the symbol sequence $\mathbf{a}$ that corresponds to the coded bit sequence $\mathbf{c}$ (i.e., $\mathbf{a} = \mathcal{A}(\mathcal{D}(\mathcal{C}(\mathbf{b})), \mathbf{p})$).

It is important to realize that the observations $\mathbf{r}$ from (4.1) and $\mathbf{r}^{(si)}$ from (4.29) are not equivalent for a time-varying phase shift $\Theta(t;\mathbf{\Psi})$, and that therefore estimating $\mathbf{\Psi}$ and $\mathbf{b}$ (MAP-i detector) or $\mathbf{c}$ (MAP-c detector) from $\mathbf{r}^{(si)}$ (see Fig. 4.11 or Fig. 4.12) instead of $r_{BB}(t)$ (see Fig. 4.2 or Fig. 4.5) is sub-optimum.

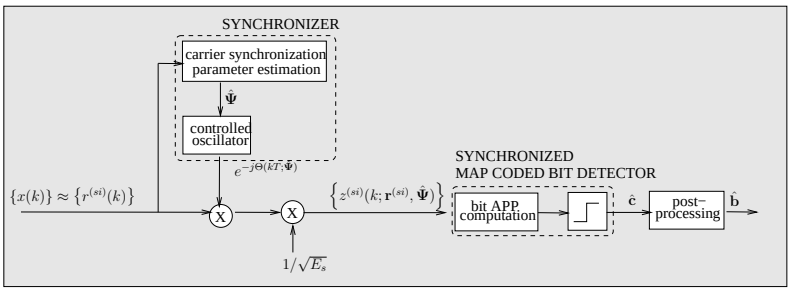

Figure 4.12: Receiver with synchronized MAP-c detector, resulting from the simplified continuous-time observation model.

4.6 Conclusion and Remarks

In this chapter we discussed the synchronized detection of information bits in digital bandpass communication systems.

We specifically focused on the description of the baseband observation model of the received signal. Two different observation models were considered: the correct continuous-time observation model (4.1),(4.8) and a commonly used simplified discrete-time observation model (4.29),(4.34).

Special attention is drawn to the fact that the application of the sum-product algorithm to the factor graph of the joint bit APPs $p\left(\mathbf{b}\left|\mathbf{r},\hat{\boldsymbol{\Psi}}\right.\right)$ and $p\left(\mathbf{c}\left|\mathbf{r},\hat{\boldsymbol{\Psi}}\right.\right)$ not only yields the (approximate) marginal bit APPs $p\left(b\left(k\right)\left|\mathbf{r},\hat{\boldsymbol{\Psi}}\right.\right)$, $k=0,1,...,N_b-1$ and/or $p\left(c\left(k\right)\left|\mathbf{r},\hat{\boldsymbol{\Psi}}\right.\right)$, $k=0,1,...,N_c-1$, but also (as by-products) the marginal data symbol APPs $p\left(d\left(k\right)\left|\mathbf{r},\hat{\boldsymbol{\Psi}}\right.\right)$, $k=0,1,...,N_d-1$. We will come back on this when evaluating *Cramer-Rao bounds* for coded transmission and when discussing superior *code-aided* estimation techniques that iterate between data detection and parameter estimation.

5

CRBs for Frequency and Phase Estimation

In this chapter we will derive true and modified Bayesian Cramer-Rao bounds (CRB) pertaining to carrier synchronization parameter estimation.

Section 5.1 gives the general problem formulation. Section 5.2 briefly reviews some related results from the literature. The correct and the simplified observation models (introduced in chapter 4) are recapitulated in section 5.3. In section 5.4, four different estimation modes are introduced. Analytical and numerical results pertaining to the modified CRBs are provided in section 5.5 and section 5.6, while section 5.7 and section 5.8 are devoted to the true CRBs.

5.1 Problem Formulation

Cramer-Rao bounds (CRBs) provide a fundamental lower bound on the achievable mean square estimation error, and as such serve as a useful benchmark for practical parameter estimators. Modified CRBs (MCRBs) are also lower bounds on the achievable mean square estimation error, but are in general looser (i.e.,

lower) than the corresponding (true) CRBs.

We recall from chapter 3 that the Bayesian CRBs and MCRBs are the diagonal elements of the inverse of the (true and modified) Bayesian *Fisher information matrices* (FIM), which in turn are related to a conditional probability density function (pdf) of the type $p(\mathbf{r}|\mathbf{u})$. Here, $\mathbf{r}$ is a vector that is referred to as the observation vector, whereas $\mathbf{u}$ is a vector comprising one or more continuous parameters that affect the statistics of $\mathbf{r}$. When viewed as a function of $\mathbf{u}$, $p(\mathbf{r}|\mathbf{u})$ is called the *likelihood function* of $\mathbf{u}$ and is denoted $L(\mathbf{u};\mathbf{r})$. Evaluating the FIM related to $L(\mathbf{u};\mathbf{r})$ is mathematically quite difficult when the statistics of $\mathbf{r}$ depend not only on $\mathbf{u}$, but also on a number of *nuisance* parameters, i.e., unknown parameters (continuous or discrete) other than $\mathbf{u}$, and denoted by a vector $\mathbf{v}$. The modified FIM (MFIM) related to $L(\mathbf{u};\mathbf{r})$ can in general be evaluated much easier than the true FIM related to $L(\mathbf{u};\mathbf{r})$.

We will consider the transmission of a linearly modulated signal over a band-limited additive white Gaussian noise (AWGN) channel. The transmitted signal is obtained by applying the symbol sequence $\{a(k)\}$ to a square-root Nyquist transmit filter. We also assume that the carrier phase shift of the received signal vis-a-vis the receiver's local reference carrier is linear in time, i.e.,

$$\Theta(t;\mathbf{\Psi}) = 2\pi F(t-t_0) + \theta = 2\pi\nu(\kappa-\kappa_0) + \theta. \tag{5.1}$$

In (5.1), F is the carrier frequency offset and $\theta = \Theta(t_0;\mathbf{\Psi})$ is the phase shift at an arbitrary time instant t_0. In order to deal with dimensionless quantities, we introduce the normalized frequency offset $\nu = FT$ and the normalized time variables $\kappa = t/T$ and $\kappa_0 = t_0/T$. The quantities ν and θ are uniformly distributed in the intervals $[-\nu_{max}, \nu_{max}]$ and $[-\pi, \pi]$, respectively. The column vector $\mathbf{\Psi}$ collects the unknown parameters ν and θ:

$$\mathbf{\Psi} = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix} = \begin{pmatrix} \nu \\ \theta \end{pmatrix}. \tag{5.2}$$

The complex baseband representation $r_{BB}(t)$ of the received noisy signal can be written as (see (2.17) and (4.1)):

$$r_{BB}(t) = s_{BB}(t;\mathbf{a},\mathbf{\Psi}) + w_{BB}(t), \tag{5.3}$$

where

$$s_{BB}(t;\mathbf{a},\mathbf{\Psi}) \quad = \quad \sqrt{E_s}\sum_{k\in I_s} a(k)\, p(t-kT)\, e^{j\Theta(t;\mathbf{\Psi})} \tag{5.4}$$

and the quantities $\mathbf{a}$, T, $p(t)$, E_s and $w_{BB}(t)$ are defined as follows:

- $\mathbf{a} = (a(k) : k \in I_s)$ is a row vector consisting of the N_s transmitted symbols. The symbol indices k belong to a set I_s of N_s consecutive integers. The vector $\mathbf{a}$ contains N_p known pilot symbols and N_d unknown random data symbols. For k belonging to the set of indices $I_p \subset I_s$, $a(k)$ is a pilot symbol with $|a(k)|^2 = 1$. For k belonging to the set

of indices $I_d = I_s \backslash I_p$, $a(k)$ is a data symbol with $E[a(k)] = 0$ and $E[a^*(k)\, a(l)] = \delta_{k-l}$. The data symbols belong to an M-point signaling constellation $\Omega = \{\omega_0, \omega_1, ..., \omega_{M-1}\}$ and result from the encoding and mapping of a sequence of N_b information bits. Further details about $\mathbf{a}$ will be provided in section 5.4.

- T is the symbol interval; $1/T$ denotes the symbol rate;
- $p(t)$ is a unit energy transmit pulse, with bandwidth B, that is square-root Nyquist with respect to the interval T;
- E_s denotes the average symbol energy;
- $w_{BB}(t)$ is complex-valued zero-mean AWGN with independent real and imaginary parts each having a power spectral density equal to $N_0/2$ over the frequency interval $[-B, B]$. The ratio E_s/N_0 is referred to as the signal-to-noise ratio (SNR).

There are two general strategies to obtain Bayesian (M)CRBs pertaining to the estimation of $\mathbf{\Psi}$ from $\mathbf{r}$, where $\mathbf{r}$ denotes a suitable vector representation of $r_{BB}(t)$. A first strategy consists in deriving (M)CRBs from the 2×2 Bayesian (M)FIM $\mathbf{J}$ related to the likelihood function $L(\nu, \theta; \mathbf{r})$ of (ν, θ). In this case the nuisance parameter vector $\mathbf{v}$ contains the symbols $\{a(k)\}$. Alternatively, one can also derive (M)CRBs from the scalar Bayesian (M)FIMs J_ν and J_θ related to the marginal likelihood functions $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ of ν and θ, respectively. In this case the nuisance parameter vector $\mathbf{v}$ consists of the symbols $\{a(k)\}$ and θ (when considering $L(\nu; \mathbf{r})$) or ν (when considering $L(\theta; \mathbf{r})$).

The computation of the MCRBs related to $L(\nu, \theta; \mathbf{r})$, $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ is straightforward and can be easily performed analytically. The computation of the corresponding true CRBs is considerably more complicated. A complete *analytical* evaluation of the true FIMs related to $L(\nu, \theta; \mathbf{r})$, $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ is often infeasible because of the presence of the nuisance parameters (symbols only, symbols and phase shift, or symbols and frequency offset). On the other hand, a brute-force *numerical* evaluation of the FIMs tends to increase the computational cost dramatically, as it involves calculating various summations with a number of terms that increases exponentially with the length of the symbol sequence $\{a(k)\}$.

In this chapter we will evaluate the MFIMs related to $L(\nu, \theta; \mathbf{r})$, $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$. We will further report on our efforts to compute the FIMs related to $L(\nu, \theta; \mathbf{r})$, $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$. The key idea is to make maximum use of analytical computations (exploiting the specific structure of the observation vector $\mathbf{r}$), so that the number and complexity of the remaining numerical integrations is drastically reduced. Four different estimation modes will be discussed: the *pilot-aided code-aided* (PA CA) mode, the *data-aided* (DA) mode, the *pilot-aided non-code-aided* (PA NCA) mode and the *non-pilot-aided non-code-aided* (NPA NCA) mode. We will also compare the results obtained from the *correct* continuous-time model of the received signal against those resulting from

a (commonly used) *simplified* discrete-time model of the matched filter output samples, which ignores the reduction of the useful signal and the inter-symbol-interference (ISI) caused by a nonzero frequency offset.

Remark

According to chapter 4, the synchronized bit detector uses the signal $r_{BB}(t)$ from (5.3) and an estimate $\hat{\mathbf{\Psi}}$ of $\mathbf{\Psi}$ to compute the quantities $z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)$, $k \in I_d$. The quantities $z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right)$, $k \in I_d$, which are then further processed to recover the value of the transmitted information bits, can be decomposed as follows:

$$z\left(k;\mathbf{r},\hat{\mathbf{\Psi}}\right) = a(k)\, e^{je_\Theta(kT)} + w(k). \tag{5.5}$$

Here, $e_\Theta(kT)$ is the estimation error on the instantaneous carrier phase shift at time instant $t = kT$ (see (5.1)):

$$\begin{aligned} e_\Theta(kT) &= \left(\Theta(kT;\mathbf{\Psi}) - \Theta\left(kT;\hat{\mathbf{\Psi}}\right)\right), \\ &= 2\pi(\nu-\hat{\nu})(k-\kappa_0) + \left(\theta-\hat{\theta}\right),\ k \in I_s \end{aligned} \tag{5.6}$$

and $\mathbf{w} = (w(k) : k \in I_s)$ denotes a sequence of independent and identically distributed (i.i.d.) complex-valued white Gaussian random variables with i.i.d. real and imaginary parts, each having a power spectral density equal to $N_0/2E_s$[1].

Equation (5.5) indicates that the performance of the detector is affected by the estimation errors $e_\Theta(kT)$, $k \in I_d$. For this reason we will not only consider the CRBs related to estimating ν and θ, but also the CRB on $E\left[\left(e_\Theta(kT)\right)^2\right]$ for all $k \in I_d$. As $e_\Theta(kT)$ is a linear combination of the estimation errors $(\nu-\hat{\nu})$ and $\left(\theta-\hat{\theta}\right)$, the CRB on $E\left[\left(e_\Theta(kT)\right)^2\right]$ for $k \in I_d$ is easily obtained from the FIM $\mathbf{J}$ related to $L(\mathbf{\Psi};\mathbf{r}) = p(\mathbf{r}\,|\mathbf{\Psi})$.

For given κ_0, the phase shift $\Theta(t;\mathbf{\Psi})$ is fully specified by the parameter vector $\mathbf{\Psi} = (\nu\ \theta)^T$. Changing κ_0 into $\acute{\kappa}_0$, the same phase shift is now specified by the vector $\acute{\mathbf{\Psi}} = \left(\acute{\nu}\ \acute{\theta}\right)^T$, with $\acute{\nu} = \nu$ and $\acute{\theta} = \theta + 2\pi\nu(\acute{\kappa}_0 - \kappa_0)$, i.e., $\Theta(t;\mathbf{\Psi})$ from (5.1) and

$$\acute{\Theta}\left(t;\acute{\mathbf{\Psi}}\right) = 2\pi\nu(\kappa - \acute{\kappa}_0) + \acute{\theta}$$

are the same. It can be verified (see Appendix C on page 191) that the CRB on $E\left[\left(e_\Theta(kT)\right)^2\right]$ does not depend on the particular selection of the value of κ_0. Essentially, this is because $\mathbf{\Psi}$ and $\acute{\mathbf{\Psi}}$ are equivalent vector representations of the instantaneous phase shift. Hence, when we choose $\kappa_0 = k$, the CRB

[1] The decomposition (5.5) is always valid for the simplified observation model; assuming that the estimation of ν and θ is reasonably accurate, the estimation error $e_\Theta(.)$ is essentially constant over a time interval equal to the duration of the pulse $p(.)$ (which amounts to a few symbol intervals), and (5.5) also holds for the correct observation model.

related to the estimation of θ equals the CRB on $E\left[\left(e_\Theta\left(kT\right)\right)^2\right]$. Therefore, an alternative way to obtain the CRB on $E\left[\left(e_\Theta\left(kT\right)\right)^2\right]$ with $k \in I_d$ is to compute the CRB related to the estimation of θ for $\kappa_0 \in I_d$.

5.2 Results from the Literature

In [49], it has been shown that (in a *Bayesian* framework) the CRBs using $L\left(\nu;\mathbf{r}\right)$ and $L\left(\theta;\mathbf{r}\right)$ are always greater than or equal to the CRBs obtained from $L\left(\nu,\theta;\mathbf{r}\right)$, and thus more tight. However, in the literature on frequency offset and phase shift estimation, much more attention has traditionally been paid to the *non-Bayesian* (or standard) CRBs than to the Bayesian CRBs. The former treat the carrier synchronization parameters ν and θ as deterministic unknowns rather than random variables, but are valid for unbiased estimates only (see chapter 3, section 3.5 for more details). The results from the literature discussed below all pertain to a non-Bayesian framework (unless explicitly mentioned otherwise).

In order to avoid the computational complexity associated with the CRB, a MCRB has been derived in [50, 51]. This MCRB is much simpler to evaluate than the corresponding CRB, but is in general looser than the CRB. In [54], the high signal-to-noise ratio (SNR) limit of the CRB, pertaining to the estimation of a scalar parameter, has been evaluated analytically, and has been shown to coincide with the MCRB when no *continuous-valued* nuisance parameters are present. It is worth noting that, in the presence of coding, synchronization algorithms must operate at a low SNR, so that this high-SNR limit of the CRB might no longer be a relevant benchmark. The NPA NCA CRBs related to $L\left(\nu,\theta;\mathbf{r}\right)$ have been derived for 2-PSK and 4-PSK in [55] and for QAM in [56], assuming the simplified observation model from section 4.5. In [57], the low-SNR limit of the NPA NCA CRBs related to $L\left(\nu,\theta;\mathbf{r}\right)$, again assuming this simplified observation model has been obtained analytically for M-PSK, M-QAM and M-PAM.

The results presented in this chapter are based on our own original work and research, as published in [1–4, 7–14]. In [2, 7, 8, 10], DA and NPA NCA CRBs for frequency offset estimation from a noisy PSK signal have been derived. These papers contrast the CRBs resulting from the correct observation model with those resulting from the simplified observation model used in [55–57], and compare the CRBs related to $L\left(\nu;\mathbf{r}\right)$ with those related to $L\left(\nu,\theta;\mathbf{r}\right)$. In [4, 12, 13], PA NCA CRBs related to $L\left(\nu,\theta;\mathbf{r}\right)$ have been considered. These CRBs can be viewed as a generalization of the DA and NPA NCA CRBs derived in [55, 56] and [2, 7, 8, 10], in the sense that they reduce to the corresponding NPA NCA CRBs when *no* pilot symbols are transmitted, and to the corresponding DA CRBs when pilot symbols *only* are transmitted. Finally, numerical procedures enabling the efficient evaluation of the PA CA CRB in the presence of coding have been presented in [1, 3, 9, 11, 14].

Normally, it makes no sense to compare standard CRBs with Bayesian CRBs,

since the underlying experiments are inherently different. For the problem at hand, however, two specific circumstances make such a comparison possible. First, it was found in [50, 51] and [1–4, 7–14, 54–57] that the standard CRBs related to $L(\nu;\mathbf{r})$ and $L(\nu,\theta;\mathbf{r})$ are independent of ν and θ. Secondly, in the Bayesian framework envisaged here, the random parameters ν and θ are assumed to be uniformly distributed over their domain. The combination of these two facts implies that the Bayesian CRBs computed in this thesis are the same as the standard CRBs presented in [50, 51] and [1–4, 7–14, 54–57].

5.3 Observation Models

The computation of the CRBs pertaining to the estimation of ν, θ and $\Theta(kT;\mathbf{\Psi})$ requires a suitable vector representation of the signal $r_{BB}(t)$ from (5.3). Sampling the signal $r_{BB}(t)$ at a rate $1/T_{sample} \geq 2B$ yields:

$$r_{BB}(kT_{\text{sample}}) = s_{BB}(kT_{\text{sample}};\mathbf{a},\mathbf{\Psi}) + w_{BB}(kT_{\text{sample}}). \tag{5.7}$$

According to the sampling theorem, there is an exact equivalence between the discrete-time signal model (5.7) and the continuous-time signal model (5.3). Scaling (5.7) by a factor $\sqrt{\frac{T_{\text{sample}}}{E_s}}$, finally leads to the following vector representation:

$$\mathbf{r} = \mathbf{s}(\mathbf{a},\mathbf{\Psi}) + \mathbf{w}, \tag{5.8}$$

with

$$(\mathbf{r})_k = \sqrt{\frac{T_{\text{sample}}}{E_s}} r_{BB}(kT_{\text{sample}}), \tag{5.9}$$

$$(\mathbf{s}(\mathbf{a},\mathbf{\Psi}))_k = \sqrt{\frac{T_{\text{sample}}}{E_s}} s_{BB}(kT_{\text{sample}};\mathbf{a},\mathbf{\Psi}), \tag{5.10}$$

$$(\mathbf{w})_k = \sqrt{\frac{T_{\text{sample}}}{E_s}} w_{BB}(kT_{\text{sample}}). \tag{5.11}$$

It follows from $w_{BB}(t)$ being complex-valued zero-mean AWGN with independent real and imaginary parts each having a power spectral density equal to $N_0/2$, that $\mathbf{w}$ is a complex-valued zero-mean AWGN vector with covariance matrix $E\left[\mathbf{w}^H\mathbf{w}\right] = \frac{N_0}{E_s}\mathbf{I}$. Such a pdf will be denoted as $\mathcal{N}_c\left(\mathbf{0}, \frac{N_0}{E_s}\mathbf{I}\right)$.

However, in [55, 56], Cramer-Rao bounds have been derived, assuming an observation vector $\mathbf{r}$ according to (5.8) but with $(\mathbf{s}(\mathbf{a},\mathbf{\Psi}))_k$ given by:

$$(\mathbf{s}(\mathbf{a},\mathbf{\Psi}))_k = a(k)\, e^{j(2\pi\nu(k-\kappa_0)+\theta)}, \tag{5.12}$$

in stead of (5.10). The right-hand side in (5.12) stands for the matched filter output samples taken at $t = kT$ and scaled by a factor $1/\sqrt{E_s}$, when $r(t)$ (5.3) is applied to the matched filter and the normalized frequency offset is assumed to be small (i.e., $|\nu| \ll 1$). It is important to realize that the observation vector

from (5.12) is *not* equivalent to the continuous-time signal model (5.3), as we have already pointed out in section 4.5 of chapter 4.

In this chapter, both the correct model (where the observation consists of the noisy linearly modulated signal at the receiver input) and the simplified observation model adopted in [55, 56] will be considered.

5.4 Estimation Modes

Denoting by $\zeta_{\mathcal{D}}$ the set of legitimate data symbol sequences $\mathbf{a}_d = (a(k) : k \in I_d)$, i.e., $\mathbf{a}_d$ results from the encoding and mapping rule, we obtain for the a priori probability mass function $\Pr[\mathbf{a}]$ of $\mathbf{a}$:

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}}] = \prod_{k \in I_p} I[\tilde{a}(k) = A_k] \cdot 2^{-N_b} I[\tilde{\mathbf{a}}_d \in \zeta_{\mathcal{D}}], \forall \tilde{\mathbf{a}} \in \Omega^{N_s}, \tag{5.13}$$

where $\tilde{\mathbf{a}}$ is a trial value of $\mathbf{a}$, $\tilde{\mathbf{a}}_d$ is a trial value of $\mathbf{a}_d$, A_k is the actual value of the pilot symbol $a(k) \in I_p$, and $I[.]$ is the indicator function, i.e., for a predicate P:

$$I[P] = \begin{cases} 0, \text{ if } P \text{ is false} \\ 1, \text{ if } P \text{ is true} \end{cases}. \tag{5.14}$$

The operation of the various carrier synchronization parameter estimators can subsequently be classified into different estimation modes. Estimators that take into account the full a priori information (5.13) are said to operate in a *pilot-aided code-aided* (PA CA) mode. Their mean square estimation error is lower-bounded by the true PA CA CRB resulting from (5.13). Unfortunately, taking into account all available statistical information on $\mathbf{a}$ in the estimation process, often turns out to be far too complex (essentially because of the complexity of the encoding rule). A solution to simplify the estimator implementation consists in dropping/approximating some parts of the available statistical a priori information. It should be noted that the mean square error resulting from estimation algorithms that rely on such an approximation of $\Pr[\mathbf{a}]$ are lower-bounded by the CRBs that are computed with the same approximate hypothesis about $\Pr[\mathbf{a}]$. Within the framework of carrier synchronization parameter estimation three particular approaches are commonly considered: the *data-aided* (DA), the *pilot-aided non-code-aided* (PA NCA) and the *non-pilot-aided non-code-aided* (NPA NCA) estimation modes.

1. In the DA estimation mode the estimate of $\mathbf{\Psi}$ is computed as if only the pilot symbols $(a(k) : k \in I_p)$ have been transmitted, i.e., by assuming that the data symbols $(a(k) : k \in I_d)$ are zero:

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}}] = \prod_{k \in I_p} I[\tilde{a}(k) = A_k] \prod_{k \in I_d} I[\tilde{a}(k) = 0], \forall \tilde{\mathbf{a}} \in \Omega^{N_s}. \tag{5.15}$$

   This approach neglects the received signal energy associated with the data symbols.

2. Pilot-aided non-code-aided algorithms rely on the assumption that all possible combinations of random data symbols $(a(k) : k \in I_d)$ are equiprobable:

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}}] = M^{-N_d} \prod_{k \in I_p} I[\tilde{a}(k) = A_k], \forall \tilde{\mathbf{a}} \in \Omega^{N_s}. \quad (5.16)$$

   This comes down to neglecting the structure that is enforced upon the data symbols by the channel code and assuming the transmission of a combination of pilot symbols and *uncoded* data symbols.

3. The non-pilot-aided non-code-aided approach boils down to ignoring both the structure of the channel code and the fact that the receiver knows the value of the pilot symbols. This is tantamount to assuming that all possible symbol vectors $\mathbf{a}$ are a priori equiprobable, i.e.,

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}}] = M^{-N_s}, \forall \tilde{\mathbf{a}} \in \Omega^{N_s}. \quad (5.17)$$

   NPA NCA algorithms drop all the available statistical a priori information about the transmitted symbol vector.

We will come back to this algorithm classification in chapter 6. In the current chapter we will derive four types of CRBs corresponding to each of the above estimation modes: PA CA CRBs, DA CRBs, PA NCA CRBs and NPA NCA CRBs.

5.5 MCRBs related to $L(\mathbf{\Psi}; \mathbf{r})$

In this section the MFIM related to the likelihood function $L(\mathbf{\Psi}; \mathbf{r})$ is first computed analytically for the different estimation modes and subsequently the corresponding MCRBs are derived.

5.5.1 Computing the MFIM

We will denote the elements of the Bayesian MFIM related to the likelihood function $L(\mathbf{\Psi}; \mathbf{r}) = p(\mathbf{r}|\mathbf{\Psi})$ of $\mathbf{\Psi} = (\Psi_1 \Psi_2)^T = (\nu\ \theta)^T$ as follows:

$$\mathbf{M} = \begin{pmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{pmatrix} = \begin{pmatrix} M_{\nu,\nu} & M_{\nu,\theta} \\ M_{\nu,\theta} & M_{\theta,\theta} \end{pmatrix}. \quad (5.18)$$

Taking $\mathbf{u} = \mathbf{\Psi}$ and $\mathbf{v} = \mathbf{a}^T$ in (3.37), we obtain:

$$M_{i,j} = E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}} \left[\frac{\partial \ln p(\mathbf{r}|\mathbf{a}, \mathbf{\Psi})}{\partial \Psi_i} \frac{\partial \ln p(\mathbf{r}|\mathbf{a}, \mathbf{\Psi})}{\partial \Psi_j} \right], \quad (5.19)$$

where $E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}[.]$ denotes averaging with respect to $p(\mathbf{r}, \mathbf{a}, \mathbf{\Psi})$. When the observation vector $\mathbf{r}$ satisfies (5.8), $M_{i,j}$ reduces to (see (3.41)):

$$M_{i,j} = \frac{2E_s}{N_0} E_{\mathbf{a},\mathbf{\Psi}} \left[\Re \left\{ \frac{\partial \mathbf{s}(\mathbf{a}, \mathbf{\Psi})}{\partial \Psi_i} \left(\frac{\partial \mathbf{s}(\mathbf{a}, \mathbf{\Psi})}{\partial \Psi_j} \right)^H \right\} \right], \quad (5.20)$$

where $E_{\mathbf{a},\mathbf{\Psi}}[.]$ denotes averaging with respect to $p(\mathbf{a},\mathbf{\Psi})$. The analytical evaluation of (5.20) has been carried out in Appendix D on page 195 for the two observation models and the four estimation modes considered. Below we present the result of this computation. For large values of N_s (PA CA, PA NCA and NPA NCA estimation mode) or large values of N_p (DA estimation mode) the result turns out to be independent of both the observation model and the value and location of the pilot symbols.

For the PA CA, the PA NCA and the NPA NCA modes (and for both observation models), one obtains:

$$M_{\nu,\nu} \approx M_{\theta,\theta} 4\pi^2 \left(\frac{N_s^2}{12} + (\kappa_{m,s} - \kappa_0)^2 \right), \text{ PA CA, PA NCA, NPA NCA,} \quad (5.21)$$

$$M_{\nu,\theta} = M_{\theta,\theta} 2\pi (\kappa_{m,s} - \kappa_0), \text{ PA CA, PA NCA, NPA NCA,} \quad (5.22)$$

$$M_{\theta,\theta} = \frac{2N_s E_s}{N_0}, \text{ PA CA, PA NCA, NPA NCA,} \quad (5.23)$$

where the quantity $\kappa_{m,s}$ is the mean of the symbol index k over the set of indices I_s, i.e.,

$$\kappa_{m,s} = \frac{1}{N_s} \sum_{k \in I_s} k \quad (5.24)$$

and the approximation is valid for large N_s. We note that the normalized time instant $\kappa_{m,s}$ corresponds to the center of the symbol burst. We also note that the MFIM entries (5.21)-(5.23) are independent of the number of pilot symbols and of their location in $\mathbf{a}$.

For the DA mode, we obtain:

$$M_{\nu,\nu} \approx M_{\theta,\theta} \cdot 4\pi^2 \left(\sigma_p^2 + (\kappa_{m,p} - \kappa_0) \right), \text{ DA,} \quad (5.25)$$

$$M_{\nu,\theta} = M_{\theta,\theta} \cdot 2\pi (\kappa_{m,p} - \kappa_0), \text{ DA,} \quad (5.26)$$

$$M_{\theta,\theta} = \frac{2N_p E_s}{N_0}, \text{ DA,} \quad (5.27)$$

where the approximation holds for large N_p. The quantities $\kappa_{m,p}$ and σ_p^2 in (5.25)-(5.26) denote the mean and the variance of the symbol index k over the subset of indices $I_p \subset I_s$, i.e.,

$$\kappa_{m,p} = \frac{1}{N_p} \sum_{k \in I_p} k, \quad (5.28)$$

$$\sigma_p^2 = \frac{1}{N_p} \sum_{k \in I_p} (k - \kappa_{m,p})^2. \quad (5.29)$$

In other words, the normalized time instant $\kappa_{m,p}$ corresponds to the center of the pilot part of the symbol burst, and σ_p^2 serves as a measure for the deviation of the pilot symbols with respect to $\kappa_{m,p}$.

5.5.2 Derivation, Numerical Evaluation and Discussion of the PA CA, the PA NCA and the NPA NCA MCRBs

It follows from the modified Bayesian Cramer-Rao inequality (3.32) that the MSEs regarding the PA CA, the PA NCA or the NPA NCA estimation of ν and θ are lower bounded by the corresponding diagonal elements of the inverse of the MFIM (5.21)-(5.23). Similarly, the modified Bayesian Cramer-Rao inequality (3.32) also yields a lower bound on the MSE regarding the PA CA, the PA NCA or the NPA NCA estimation of the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$. We obtain: $E\left[(\nu-\hat{\nu})^2\right] \geq MCRB_{\nu}(N_s)$, $E\left[\left(\theta-\hat{\theta}\right)^2\right] \geq MCRB_{\theta}(N_s)$ and $E\left[\left(\Theta(kT;\mathbf{\Psi})-\Theta\left(kT;\hat{\mathbf{\Psi}}\right)\right)^2\right] \geq MCRB_{\Theta(kT;\mathbf{\Psi})}(N_s)$, with:

$$\begin{aligned} MCRB_{\nu}(N_s) &= \left(\mathbf{M}^{-1}\right)_{\nu,\nu}, \\ &\approx \frac{1}{M_{\theta,\theta}}\frac{3}{\pi^2 N_s^2}, \text{ PA CA, PA NCA, NPA NCA,} \end{aligned} \tag{5.30}$$

$$\begin{aligned} MCRB_{\theta}(N_s) &= \left(\mathbf{M}^{-1}\right)_{\theta,\theta}, \\ &\approx \frac{1}{M_{\theta,\theta}}\left(1+\frac{12\left(\kappa_{m,s}-\kappa_0\right)^2}{N_s^2}\right), \text{ PA CA, PA NCA, NPA NCA} \end{aligned} \tag{5.31}$$

and

$$MCRB_{\Theta(kT;\mathbf{\Psi})}(N_s) = MCRB_{\theta}(N_s)|_{\kappa_0=k}, \tag{5.32}$$

where $\mathbf{M}^{-1}$ denotes the inverse of the MFIM with entries $\left(\mathbf{M}^{-1}\right)_{1,1} = \left(\mathbf{M}^{-1}\right)_{\nu,\nu}$, $\left(\mathbf{M}^{-1}\right)_{1,2} = \left(\mathbf{M}^{-1}\right)_{\nu,\theta}$ and $\left(\mathbf{M}^{-1}\right)_{2,2} = \left(\mathbf{M}^{-1}\right)_{\theta,\theta}$, $M_{\theta,\theta} = 2N_s(E_s/N_0)$, and the approximations hold for large N_s.

The bounds from (5.30)-(5.32) are inversely proportional to E_s/N_0. The bound on $E\left[\left(\theta-\hat{\theta}\right)^2\right]$ from (5.31) depends on the normalized time instant κ_0 at which the phase shift is estimated. Its *minimum* value is achieved at $\kappa_0 = \kappa_{m,s}$, i.e., when the phase shift is estimated at the center of the symbol burst. This minimum value is equal to $1/M_{\theta,\theta}$, which is inversely proportional to the number of transmitted symbols N_s. The bounds (5.30) and (5.32) regarding the estimation of the normalized frequency offset ν and the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$ do not depend on κ_0. The $MCRB_{\nu}$ is inversely proportional to the third power of N_s.

Selecting $\kappa_0 = \kappa_{m,s}$, Fig. 5.1 shows the MCRBs pertaining to the PA CA, PA NCA or NPA NCA estimation of ν and θ as a function of E_s/N_0, and for N_s = 50, 500. We make the following observations. The MCRBs are inversely proportional to E_s/N_0. When N_s increases from 50 to 500, the MCRBs pertaining to frequency offset and phase shift estimation respectively decrease by a factor

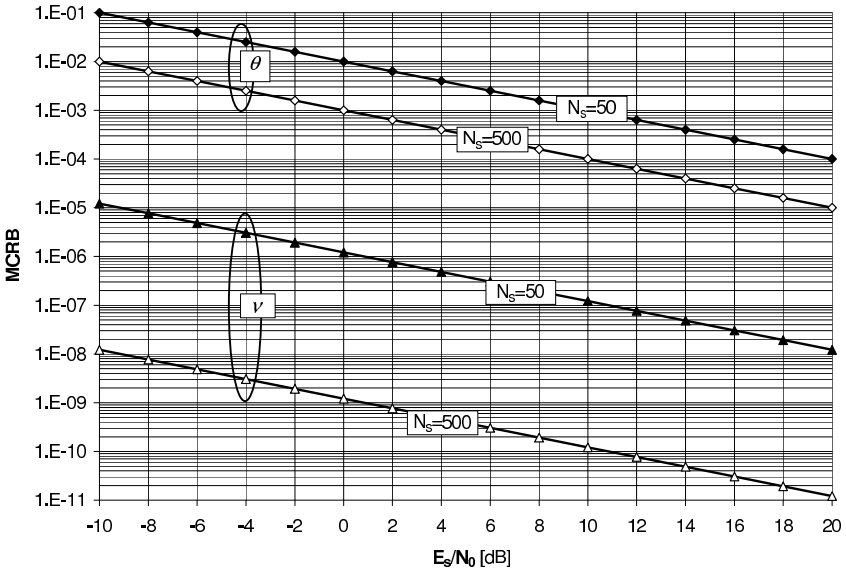

Figure 5.1: MCRBs related to $L(\mathbf{\Psi};\mathbf{r})$ and pertaining to the PA CA, the PA NCA or the NPA NCA estimation of the normalized frequency offset ν and the phase shift θ (in radian) at the center of a burst of N_s symbols.

of 1000 and 10. This is consistent with (5.30) and (5.31). The $MCRB_{\Theta(kT;\mathbf{\Psi})}$ is *quadratic* in k. Fig. 5.2 shows the MCRB pertaining to the PA CA, PA NCA or NPA NCA estimation of $\Theta(kT;\mathbf{\Psi})$ as a function of $(k-\kappa_{m,s})$, for $E_s/N_0 = 2$ dB and $N_s = 50$, 100. The smallest values of $MCRB_{\Theta(kT;\mathbf{\Psi})}$ (close to $1/M_{\theta,\theta}$) are achieved when the value of k lies close to $\kappa_{m,s}$, i.e., the normalized time instant that corresponds to the center of the symbol burst. The $MCRB_{\Theta(kT;\mathbf{\Psi})}$ achieves its maximum value at the edges of the interval I_s. The difference between the minimum and the maximum value of $MCRB_{\Theta(kT;\mathbf{\Psi})}$ over I_s amounts to $(N_s)^2\pi^2 MCRB_\nu$, for $N_s \gg 1$, which is inversely proportional to N_s (see (5.31) and (5.32)).

5.5.3 Derivation, Numerical Evaluation and Discussion of the DA MCRBs

It follows from the modified Bayesian Cramer-Rao inequality (3.32) that the MSEs regarding the DA estimation of ν and θ are lower-bounded by the corresponding diagonal elements of the inverse of the MFIM (5.25)-(5.27). Similarly, the modified Bayesian Cramer-Rao inequality (3.32) also yields a lower bound on the MSE regarding the DA estimation of the instantaneous phase

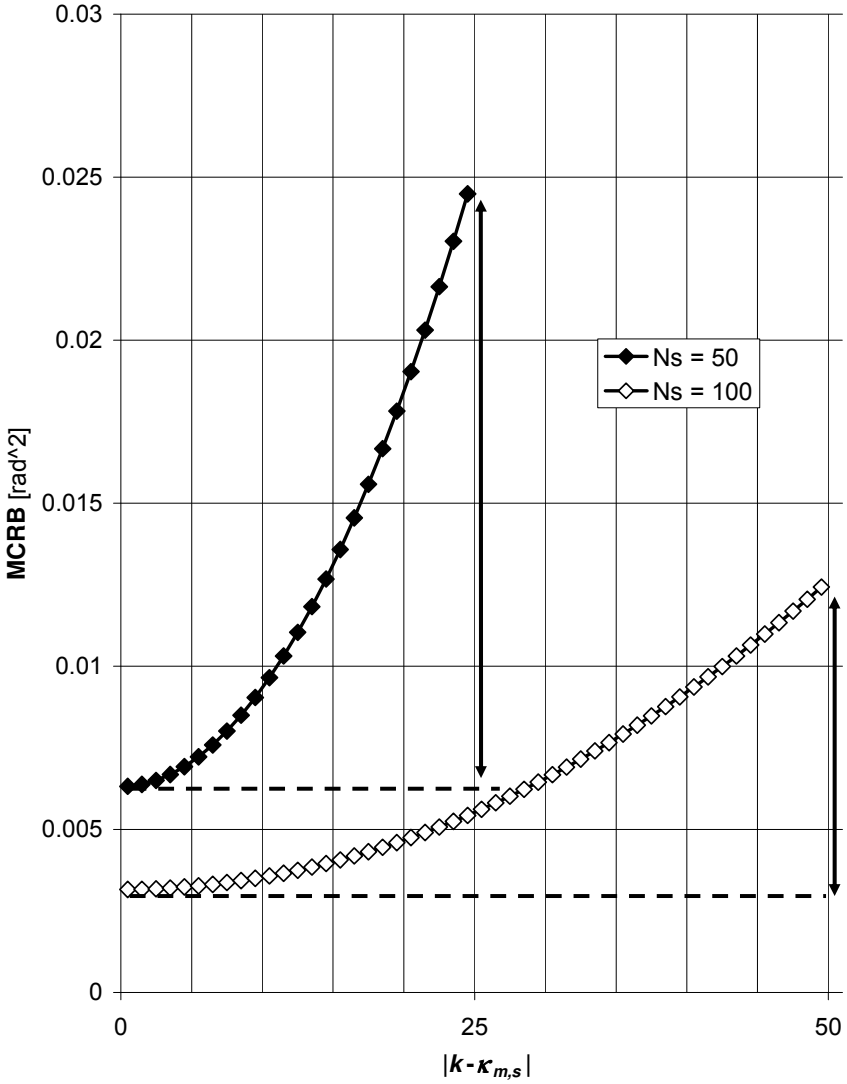

Figure 5.2: MCRB related to $L(\mathbf{\Psi};\mathbf{r})$ and pertaining to the PA CA, the PA NCA or the NPA NCA estimation of the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$, $k \in I_s$ with I_s a set of N_s consecutive integers.

shift $\Theta(kT;\mathbf{\Psi})$. We obtain: $E\left[(\nu-\hat{\nu})^2\right] \geq MCRB_\nu(N_p)$, $E\left[\left(\theta-\hat{\theta}\right)^2\right] \geq MCRB_\theta(N_p)$ and $E\left[\left(\Theta(kT;\mathbf{\Psi})-\Theta\left(kT;\hat{\mathbf{\Psi}}\right)\right)^2\right] \geq MCRB_{\Theta(kT;\mathbf{\Psi})}(N_p)$, with:

$$\begin{aligned} MCRB_\nu(N_p) &= \left(\mathbf{M}^{-1}\right)_{\nu,\nu}, \\ &\approx \frac{1}{4\pi^2\sigma_p^2 M_{\theta,\theta}}, \text{ DA}, \end{aligned} \tag{5.33}$$

$$\begin{aligned} MCRB_\theta(N_p) &= \left(\mathbf{M}^{-1}\right)_{\theta,\theta}, \\ &\approx \frac{1}{M_{\theta,\theta}}\left(1+\frac{(\kappa_{m,p}-\kappa_0)^2}{\sigma_p^2}\right), \text{ DA} \end{aligned} \tag{5.34}$$

and

$$MCRB_{\Theta(kT;\mathbf{\Psi})}(N_p) = MCRB_\theta(N_p)|_{\kappa_0=k}, \tag{5.35}$$

where $\mathbf{M}^{-1}$ denotes the inverse of the MFIM with entries $\left(\mathbf{M}^{-1}\right)_{1,1}=\left(\mathbf{M}^{-1}\right)_{\nu,\nu}$, $\left(\mathbf{M}^{-1}\right)_{1,2}=\left(\mathbf{M}^{-1}\right)_{\nu,\theta}$and $\left(\mathbf{M}^{-1}\right)_{2,2}=\left(\mathbf{M}^{-1}\right)_{\theta,\theta}$, $M_{\theta,\theta}=2N_p(E_s/N_0)$, and the approximations hold for large N_p.

The bounds from (5.33)-(5.35) are inversely proportional to E_s/N_0. The bound on $E\left[\left(\theta-\hat{\theta}\right)^2\right]$ from (5.34) depends on the normalized time instant κ_0 at which the phase shift is estimated. Its *minimum* value is achieved at $\kappa_0=\kappa_{m,p}$, i.e., when the phase shift is estimated at the center of the pilot symbols. This minimum value is equal to $1/M_{\theta,\theta}$, which is inversely proportional to the number of pilot symbols N_p. The bounds (5.33) and (5.35) regarding the estimation of the normalized frequency offset ν and the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$ do not depend on κ_0. Assuming a symbol vector $\mathbf{a}$ with two blocks of $N_p/2$ consecutive pilot symbols spaced with $\epsilon_{Sp}N_p$ data symbols, and selecting $\kappa_0=\kappa_{m,p}$, one obtains:

$$MCRB_\nu(N_p) \approx \frac{3}{2\pi^2 N_p^3\left(1+3\left(\epsilon_{Sp}+\epsilon_{Sp}^2\right)\right)E_s/N_0}, \tag{5.36}$$

$$MCRB_\theta(N_p) \approx \frac{1}{2N_pE_s/N_0}, \tag{5.37}$$

$$MCRB_{\Theta(kT;\mathbf{\Psi})}(N_p) \approx \frac{1}{2N_pE_s/N_0}\left(1+\frac{4N_p^2(\kappa_{m,p}-k)^2}{1+3\left(\epsilon_{Sp}+\epsilon_{Sp}^2\right)}\right), \tag{5.38}$$

where the approximation holds for large N_p. The quantity ϵ_{Sp} is referred to as the *spacing ratio*. Fig. 5.3 shows the $MCRB_\nu$ (5.36) and the $MCRB_\theta$ (5.37) as a function of the E_s/N_0, for $N_p=32$, 64 and $\epsilon_{Sp}=0.5$, 1.5. The DA MCRB

related to phase shift estimation is inversely proportional to N_p and E_s/N_0. The DA MCRB related to frequency offset estimation depends on the spacing ratio ϵ_{Sp}. For a given value of ϵ_{Sp}, $MCRB_\nu$ is inversely proportional to $(N_p)^3$ and E_s/N_0. For a given value of N_p and E_s/N_0, $MCRB_\nu$ decreases as the spacing ratio ϵ_{Sp} increases. This is consistent with (5.36) and (5.37). Fig. 5.4 shows the $MCRB_{\Theta(kT;\mathbf{\Psi})}$ (5.38) as a function of k, for $N_p = 64$, $\epsilon_{Sp} = 0.5$, $N_s = 320$ and $E_s/N_0 = 2$ dB. Two different burst structures are considered, both of which are shown in Fig. 5.5. The shaded areas indicate the location of the pilot symbols within the symbol vector $\mathbf{a}$. In structure #1 the burst starts with a block of pilot symbols. In this case, the center of the pilot symbols $\kappa_{m,p}$ is located at 48 symbol intervals from the start of the observation interval. In burst structure #2 the two pilot symbol blocks are symmetrically positioned about the center of the symbol vector. In this case, $\kappa_{m,p}$ corresponds to the center of the observation interval. We observe that the $MCRB_{\Theta(kT;\mathbf{\Psi})}$ is *quadratic* in k. The smallest values of $MCRB_{\Theta(kT;\mathbf{\Psi})}$ are achieved when the value of k is close to $\kappa_{m,p}$, i.e. the normalized time instant that corresponds to the center of the pilot part of the symbol burst. The $MCRB_{\Theta(kT;\mathbf{\Psi})}$ achieves its maximum value at the edge of the interval I_s which is at maximum distance from $\kappa_{m,p}$ (or simultaneously at both edges if $\kappa_{m,p} = \kappa_{m,s}$). This maximum is larger when $\kappa_{m,p}$ is further away from the center of the observation interval. This finding is consistent with (5.38).

5.6 MCRBs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$

In this section we will first compute, for each one of the different estimation modes, the MFIMs related to the likelihood functions $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$ and subsequently derive the corresponding MCRBs.

5.6.1 Computing the MFIMs

We denote the scalar Bayesian MFIMs related to the likelihood functions $L(\nu;\mathbf{r}) = p(\mathbf{r}|\nu)$ of ν and $L(\theta;\mathbf{r}) = p(\mathbf{r}|\theta)$ of θ respectively as M_ν and M_θ. Taking $\mathbf{u} = (\nu)$ and $\mathbf{v} = (\theta\ \mathbf{a})^T$ in (3.37) one obtains:

$$M_\nu = 1/MCRB_\nu(N_s), \tag{5.39}$$

$$= E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}\left[\left(\frac{\partial \ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})}{\partial\nu}\right)^2\right] \tag{5.40}$$

and taking $\mathbf{u} = (\theta)$ and $\mathbf{v} = (\nu\ \mathbf{a})^T$ in (3.37) yields:

$$M_\theta = 1/MCRB_\theta(N_s), \tag{5.41}$$

$$= E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}\left[\left(\frac{\partial \ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})}{\partial\theta}\right)^2\right], \tag{5.42}$$

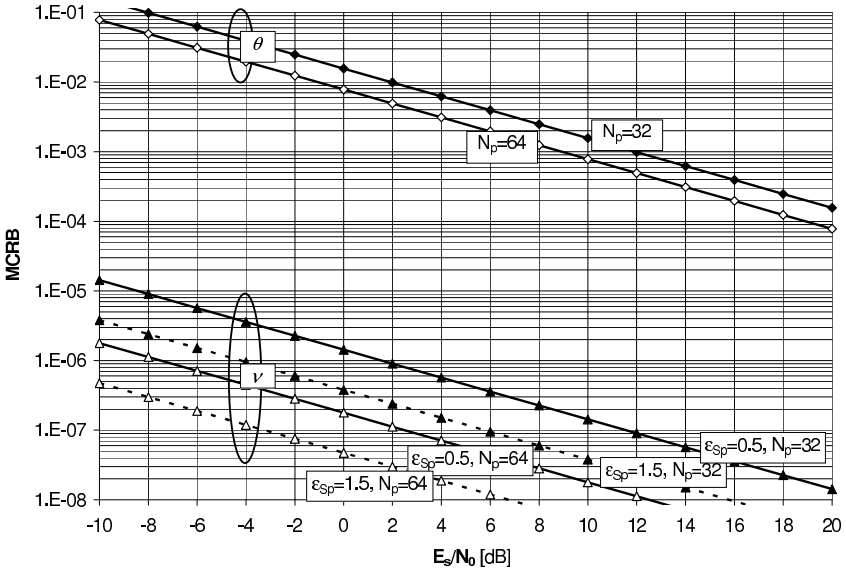

Figure 5.3: MCRBs related to $L(\mathbf{\Psi};\mathbf{r})$ and pertaining to the DA estimation of the normalized frequency offset ν and the phase shift θ (in radian) at the center of the pilot part of a burst with two blocks of $N_p/2$ pilot symbols spaced with $\epsilon_{Sp}N_p$ data symbols.

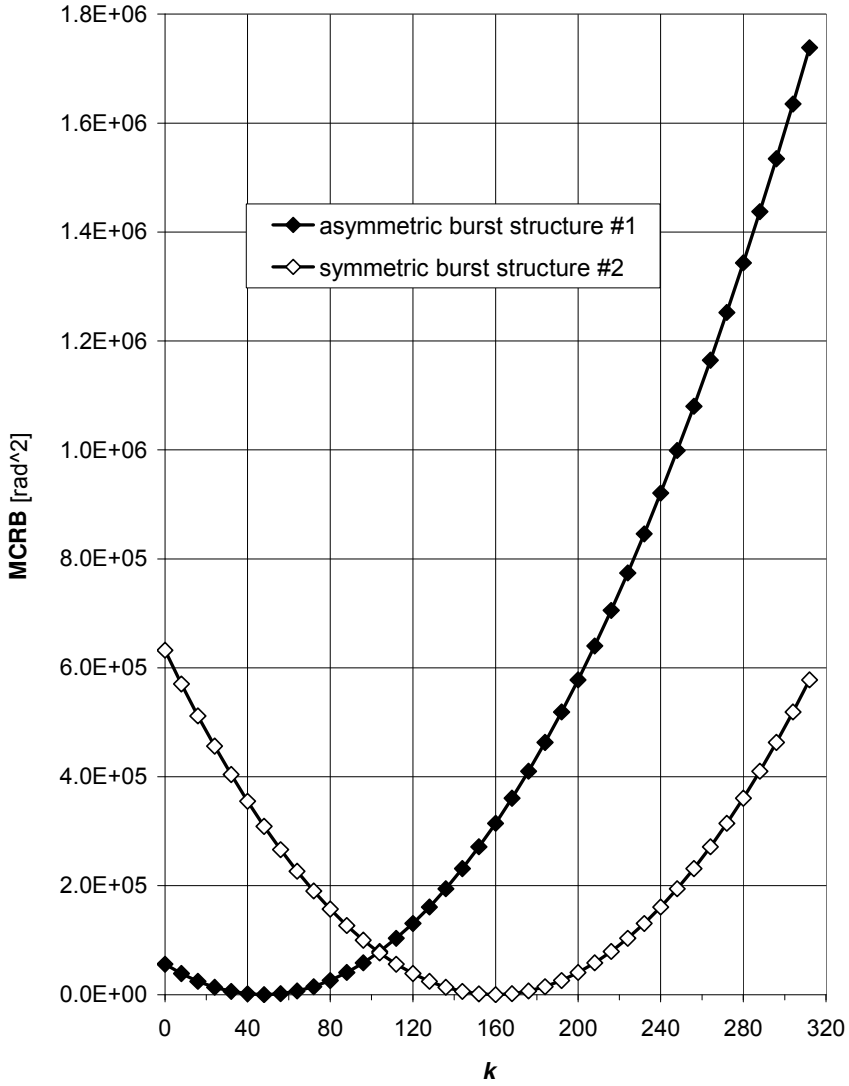

Figure 5.4: MCRBs related to $L(\mathbf{\Psi};\mathbf{r})$ and pertaining to the DA estimation of the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$, $k \in I_s$ with $I_s = [0, 320]$ and for $N_p = 64$, $\epsilon_{Sp} = 0.5$ and $E_s/N_0 = 2$ dB.

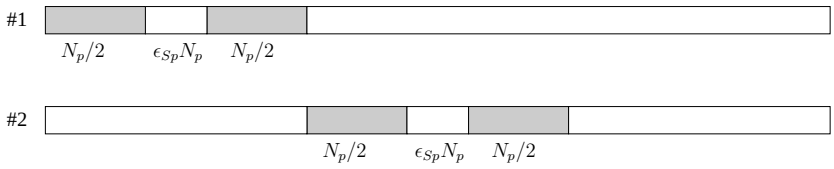

Figure 5.5: Burst Structures. The shaded areas indicate the location of the pilot symbols.

where $E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}[.]$ denotes averaging with respect to $p(\mathbf{r},\mathbf{a},\mathbf{\Psi})$. This reveals that (compare (5.19) and (5.40)-(5.42)):

$$M_\nu = M_{\nu,\nu},$$
$$M_\theta = M_{\theta,\theta},$$

where $M_{\nu,\nu}$ and $M_{\theta,\theta}$ are the diagonal elements of the MFIM related to $L(\mathbf{\Psi};\mathbf{r})$.

5.6.2 Derivation, Numerical Evaluation and Discussion of the PA CA, the PA NCA and the NPA NCA MCRBs

If we use (5.21) and (5.23) we obtain for the MCRBs corresponding to PA CA, PA NCA or NPA NCA estimation:

$$MCRB_\nu(N_s) = \frac{1}{M_{\nu,\nu}}, \tag{5.43}$$
$$\approx \frac{1}{M_{\theta,\theta}4\pi^2\left(\frac{N_s^2}{12}+(\kappa_{m,s}-\kappa_0)^2\right)}, \text{PA CA, PA NCA, NPA NCA,}$$

$$MCRB_\theta(N_s) = \frac{1}{M_{\theta,\theta}}, \tag{5.44}$$
$$= \frac{N_0}{2N_sE_s}, \text{PA CA, PA NCA, NPA NCA,}$$

where the approximation in the second line of (5.43) is valid for large N_s.

The MCRBs from (5.43) and (5.44) are inversely proportional to E_s/N_0, and coincide with the MCRBs from (5.30) and (5.31) when $\kappa_0 = \kappa_{m,s}$ (which are shown in Fig. 5.1). This does not come as a surprise, since the off-diagonal elements of the PA CA, the PA NCA and the NPA NCA MFIMs related to $L(\mathbf{\Psi};\mathbf{r})$ vanish when $\kappa_0 = \kappa_{m,s}$. The bound on $E\left[(\nu-\hat{\nu})^2\right]$ from (5.43) depends on κ_0, and achieves a maximum, given by (5.30), when $\kappa_0 = \kappa_{m,s}$. The lower bound on $E\left[\left(\theta-\hat{\theta}\right)^2\right]$ from (5.44) is independent of κ_0. It is identical to the minimum value of the lower bound on $E\left[\left(\theta-\hat{\theta}\right)^2\right]$ from (5.31), which is achieved for $\kappa_0 = \kappa_{m,s}$.

5.6.3 Derivation, Numerical Evaluation and Discussion of the DA MCRBs

For the MCRBs corresponding to DA estimation one obtains:

$$MCRB_{\nu}\left(N_p\right) = \frac{1}{M_{\nu,\nu}} = \frac{1}{M_{\theta,\theta} \cdot 4\pi^2 \left(\sigma_p^2 + \left(\kappa_{m,p} - \kappa_0\right)^2\right)}, \text{ DA}, \tag{5.45}$$

$$MCRB_{\theta}\left(N_p\right) = \frac{1}{M_{\theta,\theta}} = \frac{N_0}{2N_p E_s}, \text{ DA}. \tag{5.46}$$

The MCRBs from (5.45) and (5.46) are inversely proportional to E_s/N_0, and coincide with the DA MCRBs from (5.33) and (5.34) when $\kappa_0 = \kappa_{m,p}$ (which are shown in Fig. 5.3). This does not come as a surprise either, since the off-diagonal elements of the DA MFIM related to $L\left(\mathbf{\Psi};\mathbf{r}\right)$ vanish when $\kappa_0 = \kappa_{m,p}$. The bound on $E\left[\left(\nu - \hat{\nu}\right)^2\right]$ from (5.45) depends on κ_0, and achieves a maximum, given by (5.33), when $\kappa_0 = \kappa_{m,p}$. The lower bound on $E\left[\left(\theta - \hat{\theta}\right)^2\right]$ from (5.46) is independent of κ_0 and is identical to the minimum value of the DA $MCRB_{\theta}$ from (5.34), which is achieved for $\kappa_0 = \kappa_{m,p}$.

5.7 CRBs related to $L\left(\mathbf{\Psi};\mathbf{r}\right)$

In this section we compute the FIM (and the associated CRBs) related to $L\left(\mathbf{\Psi};\mathbf{r}\right)$. In contrast to the MFIM, the analytical evaluation of the FIM is not feasible. A brute-force evaluation of the FIM by means of simulation is conceptually simple, but the associated computational complexity increases exponentially with the number of symbols transmitted. Taking into account the linear modulation we derive a numerical procedure for evaluating the FIM with a complexity that is linear in the number of symbols transmitted. This procedure, which holds for all estimation modes considered, can be further simplified in case of the DA, PA NCA and NPA NCA estimation modes.

5.7.1 The FIM

Taking $\mathbf{u} = \mathbf{\Psi}$ and $\mathbf{v} = \mathbf{a}^T$ in (3.26), yields the elements of the true Bayesian FIM $\mathbf{J}$ related to the likelihood function $L\left(\mathbf{\Psi};\mathbf{r}\right)$ of ν and θ:

$$\mathbf{J} = \begin{pmatrix} J_{1,1} & J_{1,2} \\ J_{2,1} & J_{2,2} \end{pmatrix} = \begin{pmatrix} J_{\nu,\nu} & J_{\nu,\theta} \\ J_{\nu,\theta} & J_{\theta,\theta} \end{pmatrix}, \tag{5.47}$$

with

$$J_{i,j} \;=\; E_{\mathbf{r},\mathbf{\Psi}}\left[\ell_i\left(\mathbf{\Psi};\mathbf{r}\right)\ell_j\left(\mathbf{\Psi};\mathbf{r}\right)\right], \tag{5.48}$$

where $E_{\mathbf{r},\mathbf{\Psi}}[.]$ denotes averaging with respect to $p\left(\mathbf{r},\mathbf{\Psi}\right)$, and

$$\ell_i\left(\mathbf{\Psi};\mathbf{r}\right) = \frac{\partial \ln L\left(\mathbf{\Psi};\mathbf{r}\right)}{\partial \Psi_i} \tag{5.49}$$

is a short-hand notation for the derivative of the logarithm of the likelihood function $L(\mathbf{\Psi};\mathbf{r})$ with respect to Ψ_i.

According to (3.30), the derivatives $\ell_i(\mathbf{\Psi};\mathbf{r})$, $i=1,2$ (5.49) can be put into the following form:

$$\ell_i(\mathbf{\Psi};\mathbf{r}) = \sum_{\tilde{\mathbf{a}}} \frac{\partial \ln p(\mathbf{r}|\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi})}{\partial \Psi_i} \Pr[\mathbf{a}=\tilde{\mathbf{a}}|\mathbf{r},\mathbf{\Psi}]. \tag{5.50}$$

As the conditional pdf $p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})$ is Gaussian (see (5.8)), the logarithm $\ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})$ is readily available in closed-form:

$$\ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi}) = -\frac{E_s}{N_0}|\mathbf{r}-\mathbf{s}(\mathbf{a},\mathbf{\Psi})|^2 \tag{5.51}$$

and the joint symbol a posteriori probabilities (APP) $\Pr[\mathbf{a}=\tilde{\mathbf{a}}|\mathbf{r},\mathbf{\Psi}]$ can be computed from $p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})$ and the a priori distribution $\Pr[\mathbf{a}]$ of the symbol vector $\mathbf{a}$, according to:

$$\Pr[\mathbf{a}=\tilde{\mathbf{a}}|\mathbf{r},\mathbf{\Psi}] = \frac{\Pr[\mathbf{a}=\tilde{\mathbf{a}}]\,p(\mathbf{r}|\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi})}{\sum_{\check{\mathbf{a}}\in\Omega^{N_s}} p(\mathbf{r}|\mathbf{a}=\check{\mathbf{a}},\mathbf{\Psi})\Pr[\mathbf{a}=\check{\mathbf{a}}]}. \tag{5.52}$$

5.7.2 High-SNR Behavior of the FIM

Assuming that $\mathbf{r}$ is given by (5.8), we compute the high-SNR limit of the FIM elements $J_{i,j}$ (5.48) in Appendix E on page 203. At high E_s/N_0, one obtains:

$$\begin{aligned}
&E_{\mathbf{r},\mathbf{\Psi}}[\ell_i(\mathbf{\Psi};\mathbf{r})\,\ell_j(\mathbf{\Psi};\mathbf{r})],\\
&\approx E_{\tilde{\mathbf{a}},\mathbf{\Psi}}\left[E_{\mathbf{r}|\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi}}\left[\frac{\partial \ln p(\mathbf{r}|\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi})}{\partial \Psi_i}\frac{\partial \ln p(\mathbf{r}|\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi})}{\partial \Psi_j}\right]\right],\\
&= E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}\left[\frac{\partial \ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})}{\partial \Psi_i}\frac{\partial \ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})}{\partial \Psi_j}\right].
\end{aligned} \tag{5.53}$$

For both observation models and all estimation modes, the high-SNR limit (5.53) of the FIM entries $J_{i,j}$ (5.48) equals the corresponding MFIM entries $M_{i,j}$ (5.20).

A similar reasoning has been used in [54] to derive the high-SNR asymptote of the non-Bayesian CRB, pertaining to the the estimation of a scalar parameter in the presence of nuisance parameters.

5.7.3 Computing the FIM by "Brute Force" Simulation

A brute force numerical evaluation of the FIM involves replacing the statistical average $E[.]$ in (5.48) by an arithmetical average over a large number of realizations of $(\mathbf{r},\mathbf{\Psi})$, which are computer-generated according to the joint distribution of $\mathbf{r}$ and $\mathbf{\Psi}$. This leads to the computation algorithm that is described in Algorithm 1, where the generation of a realization $\mathbf{a}^{(n)}$ according to $\Pr[\mathbf{a}]$ involves

Algorithm 1 FIM related to $L(\mathbf{\Psi};\mathbf{r})$ by brute force simulation

1. For $n = 1, 2, ..., N$:
   - generate $\mathbf{a}^{(n)}$ and $\mathbf{\Psi}^{(n)} = \left(\nu^{(n)}\theta^{(n)}\right)^T$ according to the distributions $\Pr[\mathbf{a}]$, $p(\nu)$ and $p(\theta)$;
   - generate $\mathbf{w}^{(n)}$ according to $\mathcal{N}_c\left(\mathbf{0}, \frac{N_0}{E_s}\mathbf{I}\right)$;
   - compute $\mathbf{r}^{(n)} = \mathbf{s}\left(\mathbf{a}^{(n)}, \mathbf{\Psi}^{(n)}\right) + \mathbf{w}^{(n)}$;
   - compute $\ell_i\left(\mathbf{\Psi}^{(n)}; \mathbf{r}^{(n)}\right)$ for $i \in \{1, 2\}$ according to (5.50).
2. Compute $J_{i,j}$ for $(i, j) \in \{1, 2\}^2$ as:

$$J_{i,j} = \frac{1}{N}\sum_{n=1}^{N} \ell_i\left(\mathbf{\Psi}^{(n)}; \mathbf{r}^{(n)}\right) \ell_j\left(\mathbf{\Psi}^{(n)}; \mathbf{r}^{(n)}\right).$$

the generation of a pilot symbol sequence $\mathbf{p}^{(n)}$, the generation of an information bit sequence $\mathbf{b}^{(n)}$, and the computation of $\mathbf{a}^{(n)}$ from $\mathbf{p}^{(n)}$and $\mathbf{b}^{(n)}$according to the encoding, the mapping and the pilot symbol insertion rules at hand. It should be noted that, due to the summations in (5.50) and (5.52), the computation of the derivatives $\ell_i(\mathbf{\Psi};\mathbf{r})$, $i = 1, 2$ in Step 1 gives rise to a computational complexity that is exponential in the burst size N_s.

5.7.4 Computing the FIM for Linear Modulation

We will now show that the complexity associated with the computation of the derivatives $\ell_i\left(\mathbf{\Psi}^{(n)}; \mathbf{r}^{(n)}\right)$, $i = 1, 2$ corresponding to the realizations $\left(\mathbf{r}^{(n)}, \mathbf{\Psi}^{(n)}\right)$, $n = 1, 2, ..., N$ of $(\mathbf{r}, \mathbf{\Psi})$ can be drastically reduced for linear modulation. In that case, the signal vector $\mathbf{s}(\mathbf{a}, \mathbf{\Psi})$ is decomposable into the matrix product of the symbol vector $\mathbf{a}$ and a matrix $\mathbf{S}(\mathbf{\Psi})$ that only depends on $\mathbf{\Psi}$:

$$\mathbf{s}(\mathbf{a}, \mathbf{\Psi}) = \mathbf{a}\mathbf{S}(\mathbf{\Psi}). \tag{5.54}$$

For the correct observation model using (5.10), the (k, l)th element $S((k, l); \mathbf{\Psi})$ of $\mathbf{S}(\mathbf{\Psi})$ is given by:

$$S((k, l); \mathbf{\Psi}) = \sqrt{T_{\text{sample}}}\, p(lT_{\text{sample}} - kT)\, e^{j\left(\theta + 2\pi\nu\left(l\frac{T_{\text{sample}}}{T} - \kappa_0\right)\right)}, \tag{5.55}$$

whereas, for the simplified observation model using (5.12), we obtain:

$$S((k, l); \mathbf{\Psi}) = e^{j(\theta + 2\pi\nu(k - \kappa_0))}\delta_{k-l}. \tag{5.56}$$

For both observation models, the following holds:

$$\mathbf{S}(\mathbf{\Psi})\,\mathbf{S}^H(\mathbf{\Psi}) = \mathbf{I}. \tag{5.57}$$

It follows from (5.57) that (within a factor not depending on $(\mathbf{a},\mathbf{\Psi})$) the joint log-likelihood function $\ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})$ (5.51) of $\mathbf{a}$ and $\mathbf{\Psi}$ may be rewritten as follows:

$$\ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi}) \propto -\frac{E_s}{N_0}\left(|\mathbf{a}|^2 - 2\Re\left\{\mathbf{z}(\mathbf{r},\mathbf{\Psi})\,\mathbf{a}^H\right\}\right), \tag{5.58}$$

with

$$\mathbf{z}(\mathbf{r},\mathbf{\Psi}) = \mathbf{r}\mathbf{S}^H(\mathbf{\Psi}). \tag{5.59}$$

Differentiating with respect to Ψ_i yields:

$$\frac{\partial \ln p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})}{\partial \Psi_i} = \frac{2E_s}{N_0}\Re\left\{\mathbf{z}_i(\mathbf{r},\mathbf{\Psi})\,\mathbf{a}^H\right\}, \tag{5.60}$$

where $\mathbf{z}_i(\mathbf{r},\mathbf{\Psi})$ is a short-hand notation for the derivative of $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$ with respect to Ψ_i , i.e.,

$$\mathbf{z}_i(\mathbf{r},\mathbf{\Psi}) = \frac{\partial \mathbf{z}(\mathbf{r},\mathbf{\Psi})}{\partial \Psi_i}. \tag{5.61}$$

Substituting (5.60) into (5.50) one obtains:

$$\ell_i(\mathbf{\Psi};\mathbf{r}) = \frac{2E_s}{N_0}\Re\left\{\mathbf{z}_i(\mathbf{r},\mathbf{\Psi})\,\boldsymbol{\mu}^H(\mathbf{z}(\mathbf{r},\mathbf{\Psi}))\right\}, \tag{5.62}$$

where $\boldsymbol{\mu}(\mathbf{z}(\mathbf{r},\mathbf{\Psi}))$ is a short-hand notation for the a posteriori average of the symbol vector $\mathbf{a}$. The kth component of this average is given by:

$$\begin{aligned}
\mu(k;\mathbf{z}(\mathbf{r},\mathbf{\Psi})) &= E_{\mathbf{a}}[a(k)|\mathbf{r},\mathbf{\Psi}], && (5.63)\\
&= \sum_{\omega\in\Omega}\Pr[a(k)=\omega|\mathbf{r},\mathbf{\Psi}]\,\omega, && (5.64)\\
&= \sum_{\omega\in\Omega}\Pr[a(k)=\omega|\mathbf{z}(\mathbf{r},\mathbf{\Psi})]\,\omega, && (5.65)
\end{aligned}$$

where the averaging $E_{\mathbf{a}}[.|\mathbf{r},\mathbf{\Psi}]$ is with respect to $\Pr[\mathbf{a}|\mathbf{r},\mathbf{\Psi}]$ and Ω denotes the set of constellation points. We emphasize that no approximation is involved in obtaining (5.64); this line simply expresses the a posteriori average of the kth symbol $a(k)$ in terms of the *marginal* APP $\Pr[a(k)|\mathbf{r},\mathbf{\Psi}]$ of $a(k)$, rather than the *joint* APP $\Pr[\mathbf{a}|\mathbf{r},\mathbf{\Psi}]$ of all components of $\mathbf{a}$. In principle the marginal APPs $\Pr[a(k)=\omega|\mathbf{r},\mathbf{\Psi}]$ can always be computed from $p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})$ and the a priori distribution $\Pr[\mathbf{a}]$ of the symbol vector $\mathbf{a}$, according to:

$$\Pr[a(k)=\omega|\mathbf{r},\mathbf{\Psi}] = \frac{\sum_{\tilde{\mathbf{a}}\in\Omega^{N_s};\,\tilde{a}(k)=\omega}\Pr[\mathbf{a}=\tilde{\mathbf{a}}]\,p(\mathbf{r}|\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi})}{\sum_{\check{\mathbf{a}}\in\Omega^{N_s}}p(\mathbf{r}|\mathbf{a}=\check{\mathbf{a}},\mathbf{\Psi})\Pr[\mathbf{a}=\check{\mathbf{a}}]}, \tag{5.66}$$

where the summation in the numerator is with respect to all $\tilde{\mathbf{a}} = (\tilde{a}(k) : k\in I_s)$ in Ω^{N_s} for which $\tilde{a}(k)$ is equal to ω. It follows from (5.58) and (5.66) that $\Pr[a(k)|\mathbf{r},\mathbf{\Psi}]$ depends on $(\mathbf{r},\mathbf{\Psi})$ only through $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$, and consequently this is also the case for the a posteriori average of the symbol $a(k)$. This explains (5.65) and the left hand side of (5.63).

Algorithm 2 FIM related to $L(\mathbf{\Psi};\mathbf{r})$ for linear modulation

1. For $n = 1, 2, ..., N$:

   - generate $\mathbf{a}^{(n)}$ and $\mathbf{\Psi}^{(n)} = \left(\nu^{(n)}\theta^{(n)}\right)^T$ according to the distributions $\Pr[\mathbf{a}]$, $p(\nu)$ and $p(\theta)$;
   - generate $\mathbf{w}^{(n)}$ according to $\mathcal{N}_c\left(\mathbf{0}, \frac{N_0}{E_s}\mathbf{I}\right)$;
   - compute $\mathbf{r}^{(n)} = \mathbf{a}^{(n)}\mathbf{S}\left(\mathbf{\Psi}^{(n)}\right) + \mathbf{w}^{(n)}$;
   - compute $\mathbf{z}\left(\mathbf{r}^{(n)}, \mathbf{\Psi}^{(n)}\right) = \mathbf{r}^{(n)}\mathbf{S}^H\left(\mathbf{\Psi}^{(n)}\right)$;
   - compute $\Pr\left[a(k) = \omega \left| \mathbf{r}^{(n)}, \mathbf{\Psi}^{(n)} \right.\right]$ for $\omega \in \Omega$ and $k \in I_s$ by applying the sum-product algorithm to a factor graph representing a suitable factorization of $\Pr\left[\mathbf{a} \left| \mathbf{r}^{(n)}, \mathbf{\Psi}^{(n)} \right.\right] \propto \Pr[\mathbf{a}] \exp\left(-\frac{E_s}{N_0}\left(|\mathbf{a}|^2 - 2\Re\left\{\mathbf{z}\left(\mathbf{r}^{(n)}, \mathbf{\Psi}^{(n)}\right)\mathbf{a}^H\right\}\right)\right)$;
   - compute $\mathbf{z}_i\left(\mathbf{r}^{(n)}, \mathbf{\Psi}^{(n)}\right) = \mathbf{r}^{(n)} \left.\frac{\partial \mathbf{S}^H(\mathbf{\Psi})}{\partial \Psi_i}\right|_{\mathbf{\Psi}=\mathbf{\Psi}^{(n)}}$;
   - compute $\mu\left(k; \mathbf{z}\left(\mathbf{r}^{(n)}, \mathbf{\Psi}^{(n)}\right)\right)$, for $k \in I_s$ according to (5.64);
   - compute $\ell_i\left(\mathbf{\Psi}^{(n)}; \mathbf{r}^{(n)}\right)$ for $i \in \{1, 2\}$ according to (5.62).

2. Compute $J_{i,j}$ for $(i,j) \in \{1,2\}^2$ as:

$$J_{i,j} = \frac{1}{N}\sum_{n=1}^{N} \ell_i\left(\mathbf{\Psi}^{(n)}; \mathbf{r}^{(n)}\right) \ell_j\left(\mathbf{\Psi}^{(n)}; \mathbf{r}^{(n)}\right).$$

Computing the marginal APPs according to (5.66) still yields a complexity that increases exponentially with N_s. In most practical scenarios, however, the required marginal symbol APPs can directly be computed in an efficient way by applying the sum-product algorithm to a factor graph representing a suitable factorization of the joint APP $\Pr[\mathbf{a} | \mathbf{r}, \mathbf{\Psi}] \propto \Pr[\mathbf{a}]\, p(\mathbf{r} | \mathbf{a}, \mathbf{\Psi})$. For more details we refer to chapter 4.

Based on the above considerations, the FIM elements $J_{i,j}$, $(i,j) \in \{1,2\}^2$ can be computed as in Algorithm 2. The complexity of this procedure is linear (and not exponential) in the number of symbols N_s.

5.7.5 Computing the FIM for Linear Modulation: Direct Generation of $\mathbf{z}(\mathbf{r}, \mathbf{\Psi})$

In the case of the correct observation model, the computation of $\mathbf{s}\left(\mathbf{a}^{(n)}, \mathbf{\Psi}^{(n)}\right)$ involves the convolution of the symbol vector $\mathbf{a}^{(n)}$ with samples of the transmit filter impulse response $p(t)$, while the computation of $\mathbf{z}(\mathbf{r}, \mathbf{\Psi})$ and $\mathbf{z}_i(\mathbf{r}, \mathbf{\Psi})$ in-

volves the convolution of the observation vector $\mathbf{r}^{(n)}$ with samples of the matched filter impulse response (again $p(t)$, as we restrict attention to real-valued symmetrical $p(t)$). We will now demonstrate that this complexity can be reduced by directly generating realizations of $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$. Generating these realizations is easy as, for given $(\mathbf{a},\mathbf{\Psi})$, $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$ can be decomposed as:

$$\begin{aligned}\mathbf{z}(\mathbf{r},\mathbf{\Psi}) &= (\mathbf{a}\mathbf{S}(\mathbf{\Psi})+\mathbf{w})\,\mathbf{S}^H(\mathbf{\Psi}),\\ &= \mathbf{a}+\mathbf{n},\end{aligned}$$

where $\mathbf{n}=\mathbf{w}\mathbf{S}^H(\mathbf{\Psi})$ is a complex-valued AWGN vector with zero mean and covariance matrix $E\left[\mathbf{n}^H\mathbf{n}\right]$ given by:

$$\begin{aligned}E\left[\mathbf{n}^H\mathbf{n}\right] &= \mathbf{S}(\mathbf{\Psi})\,E\left[\mathbf{w}^H\mathbf{w}\right]\mathbf{S}^H(\mathbf{\Psi}),\\ &= \frac{N_0}{E_s}\mathbf{I}.\end{aligned}$$

Note that, for given $(\mathbf{a},\mathbf{\Psi})$, the statistics of $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$ do not depend on $\mathbf{\Psi}$.

It follows from (5.62) that $\ell_i(\mathbf{\Psi};\mathbf{r})$ only depends on $(\mathbf{r},\mathbf{\Psi})$ through $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$ and $\mathbf{z}_i(\mathbf{r},\mathbf{\Psi})$. The elements $J_{i,j}$, $(i,j)\in\{1,2\}^2$of the FIM (5.48) can therefore be rewritten as follows:

$$J_{i,j}=\left(\frac{2E_s}{N_0}\right)^2\sum_{k,l\in I_s}E_{\mathbf{z}}\left[H_{i,j}(k,l;\mathbf{z})\right], \tag{5.67}$$

with $\mathbf{z}$ a short-hand notation of $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$, and

$$H_{i,j}(k,l;\mathbf{z})=E_{z_i(k),z_j(l)|\mathbf{z}}\left[\Re\left\{z_i(k)\,\mu^*(k;\mathbf{z})\right\}\Re\left\{z_j(l)\,\mu^*(l;\mathbf{z})\right\}\right], \tag{5.68}$$

with $z_i(k)$, $(i,k)\in\{1,2\}\times I_s$ a short-hand notation of $z_i(k;\mathbf{r},\mathbf{\Psi})$. Taking into account the statistical properties of $\mathbf{z}$, $\frac{\partial\mathbf{z}}{\partial\nu}$ and $\frac{\partial\mathbf{z}}{\partial\theta}$, we were able to perform analytically the average in (5.68) over $z_i(k)$ and $z_j(l)$, conditioned on $\mathbf{z}$, based on $E\left[z_i(k)\,z_j(l)\,|\mathbf{z}\right]$ and $E\left[z_i^*(k)\,z_j(l)\,|\mathbf{z}\right]$. The details of this procedure are outlined in Appendix F on page 207. Further evaluation of (5.67) subsequently only requires numerically averaging over $\mathbf{z}$. As a consequence, the numerical complexity is considerably reduced, leading to the algorithm described in Algorithm 3.

Remarks

From the analytical expression for $H_{i,j}\left(k,l;\mathbf{z}^{(n)}\right)$ resulting from Appendix F on page 207 the following conclusions can be drawn. Firstly, the FIM resulting from the simplified observation model does not depend on the shape of the transmit pulse. Secondly, the FIM entries $J_{\nu,\theta}=J_{\theta,\nu}$ and $J_{\theta,\theta}$ do not depend on the observation model (correct or simplified). Thirdly, the FIM entry $J_{\nu,\nu}$ resulting from the correct observation model depends on the shape of the transmit pulse and equals the sum of the FIM entry $J_{\nu,\nu}$ resulting from the simplified observation model and an additional (pulse shape dependent) term.

Algorithm 3 FIM related to $L(\mathbf{\Psi};\mathbf{r})$ for linear modulation (bis)

1. For $n = 1, 2, ..., N$:
   - generate $\mathbf{a}^{(n)}$ according to the distribution $\Pr[\mathbf{a}]$;
   - generate $\mathbf{n}^{(n)}$ according to $\mathcal{N}_c\left(\mathbf{0}, \frac{N_0}{E_s}\mathbf{I}\right)$;
   - compute $\mathbf{z}^{(n)} = \mathbf{a}^{(n)} + \mathbf{n}^{(n)}$;
   - compute $\Pr\left[a(k) = \omega \left| \mathbf{z}^{(n)} \right.\right]$ for $\omega \in \Omega$ and $k \in I_s$ by applying the sum-product algorithm to a factor graph representing a suitable factorization of $\Pr\left[\mathbf{a} \left| \mathbf{z}^{(n)} \right.\right] \propto \Pr[\mathbf{a}]\, p\left(\mathbf{z}^{(n)} \left| \mathbf{a} \right.\right)$;
   - compute $\mu^*\left(k; \mathbf{z}^{(n)}\right)$ for $k \in I_s$;
   - compute $H_{i,j}\left(k, l; \mathbf{z}^{(n)}\right)$ (5.68) for $(i,j) \in \{1,2\}^2$ and $(k,l) \in (I_s)^2$ as outlined in Appendix F on page 207.
2. Compute $J_{i,j}$ for $(i,j) \in \{1,2\}^2$ as:

$$J_{i,j} = \frac{1}{N}\sum_{n=1}^{N}\left(\left(\frac{2E_s}{N_0}\right)^2 \sum_{k,l\in I_s} H_{i,j}\left(k, l; \mathbf{z}^{(n)}\right)\right).$$

The results obtained in subsections 5.7.4-5.7.5 hold for both observation models and for different estimation modes. For the PA CA mode, no further simplifications are possible. For the DA, PA NCA and NPA NCA modes, the a priori distribution $\Pr[\mathbf{a}]$ takes a form that does allow additional simplifications. We will elaborate on this in subsections 5.7.6-5.7.8.

5.7.6 Computing the DA FIM for Linear Modulation

Appendix G on page 213 shows that, when $\Pr[\mathbf{a}]$ satisfies (5.15) (DA approximation), one obtains:

$$\begin{aligned} &E_{\mathbf{r},\mathbf{\Psi}}\left[\ell_i(\mathbf{\Psi};\mathbf{r})\,\ell_j(\mathbf{\Psi};\mathbf{r})\right], \\ &= E_{\mathbf{r},\mathbf{\Psi}}\left[\frac{\partial \ln p(\mathbf{r}\,|\mathbf{a}=\boldsymbol{\alpha},\mathbf{\Psi})}{\partial \Psi_i}\frac{\partial \ln p(\mathbf{r}\,|\mathbf{a}=\boldsymbol{\alpha},\mathbf{\Psi})}{\partial \Psi_j}\right], \\ &= E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}\left[\frac{\partial \ln p(\mathbf{r}\,|\mathbf{a},\mathbf{\Psi})}{\partial \Psi_i}\frac{\partial \ln p(\mathbf{r}\,|\mathbf{a},\mathbf{\Psi})}{\partial \Psi_j}\right], \end{aligned} \tag{5.69}$$

where $\boldsymbol{\alpha} = (\alpha(k) : k \in I_s)$ with $\alpha(k) = A_k$ for $k \in I_p$ and $\alpha(k) = 0$ for $k \in I_d$, and averaging over $\mathbf{a}$ denotes averaging with respect to the DA approximation (5.15) of $\Pr[\mathbf{a}]$. It follows from (5.69) that for the DA estimation mode (and for

both observation models):

$$J_{i,j} = M_{i,j}^{(DA)}, \text{ DA}, \tag{5.70}$$

where $M_{1,1}^{(DA)} = M_{\nu,\nu}^{(DA)}$, $M_{1,2}^{(DA)} = M_{2,1}^{(DA)} = M_{\nu,\theta}^{(DA)} = M_{\theta,\nu}^{(DA)}$ and $M_{\theta,\theta}^{(DA)}$ denote the elements of the DA MFIM (5.25)-(5.27).

Remarks

The significance of the DA FIM extends beyond the DA estimation mode as is demonstrated by the following properties of the PA CA FIM (5.67):

1. Appendix H on page 215 shows that, when $\Pr[\mathbf{a}]$ is given by the correct PA CA expression (5.13), one obtains:

$$E_{\mathbf{z}}[H_{i,j}(k,l;\mathbf{z})] = 0, \quad (k,l) \in (I_p \times I_d) \bigcup (I_d \times I_p). \tag{5.71}$$

   Substituting (5.71) into (5.67) reveals that the PA CA FIM $\mathbf{J}$ can be decomposed as the sum of a pilot symbol contribution $\mathbf{J}_p$ and a data symbol contribution $\mathbf{J}_d$, i.e.,

$$\mathbf{J} = \mathbf{J}_p + \mathbf{J}_d.$$

   The elements of $\mathbf{J}_d$ are given by (5.67), provided that we replace the double summation over I_s by a double summation over I_d only. Similarly, the elements of $\mathbf{J}_p$ are given by (5.67), provided we replace the double summation over I_s by a double summation over I_p only. The latter fact implies neglecting the received signal energy associated with the data symbols and computing the FIM as if only the pilot symbols had been transmitted (DA assumption). It follows that $\mathbf{J}_p$ is nothing but the DA FIM (5.70).

2. For small E_s/N_0, the (dominant part of the) FIM may be obtained by replacing in (5.48) the true log-likelihood function by a low-SNR approximation. For linear modulation this in turn is obtained by:

   - writing the joint likelihood function of $\mathbf{a}$ and $\mathbf{\Psi}$ (see also (5.58)) as follows:

$$\begin{aligned} &p(\mathbf{r}|\mathbf{a},\mathbf{\Psi}) \\ &\propto \prod_{k\in I_s} \exp\left(-\frac{E_s}{N_0}\left(|a(k)|^2 - 2\Re\{z(k;\mathbf{r},\mathbf{\Psi})\,a^*(k)\}\right)\right); \end{aligned} \tag{5.72}$$

   - expanding the exponential function in (5.72) into a Taylor series;
   - working out the multiplication;
   - averaging each resulting term with respect to the data symbols;
   - keeping only the relevant terms that correspond to the smallest powers of E_s/N_0.

Assuming a non-zero number of pilot symbols, Appendix I on page 219 shows that the low-SNR limit of the FIM is precisely the DA FIM (5.70), such that:

$$\lim_{E_s/N_0 \to 0} \mathbf{J} = \mathbf{M}^{(DA)},\ I_p \neq \emptyset,$$

with $\mathbf{M}^{(DA)}$ denoting the DA MFIM with entries (5.25)-(5.27).

3. It is easily verified from (5.13)-(5.17) that, when all transmitted symbols are pilot symbols,

$$\text{PA CA FIM} = \text{PA NCA FIM} = \text{NPA NCA FIM} = \text{DA FIM} =$$
$$\text{PA CA MFIM} = \text{PA NCA MFIM} = \text{NPA NCA MFIM} = \text{DA MFIM}.$$

This implies that for any scenario in which not all transmitted symbols are pilot symbols the ratio $CRB/MCRB\left(N_s\right)$ with $MCRB\left(N_s\right)$ denoting the PA CA, the PA NCA or the NPA NCA MCRB given by (5.30)-(5.31), is a measure of the penalty caused by not a priori knowing the value of the data symbols $a\left(k\right)$, for $k \in I_d$.

5.7.7 Computing the PA NCA FIM and the Associated CRBs for Linear Modulation

It follows from (5.15) and (5.16) that the PA NCA FIM is a special case of the FIM, namely for uncoded transmission. Taking into account (5.16) it is easily verified from (5.58) and (5.66) that the marginal symbol APP $\Pr\left[a\left(k\right)|\mathbf{r},\mathbf{\Psi}\right]$ reduces to:

$$\Pr\left[a\left(k\right)=\omega\,|\mathbf{r},\mathbf{\Psi}\right] = \begin{cases} I\left[\omega = A_k\right] & ,\, k \in I_p \\ G\left(\omega, z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right) & ,\, k \in I_d \end{cases}, \tag{5.73}$$

where A_k is the actual value of the pilot symbol $a\left(k\right)$, $k \in I_p$,

$$G\left(\omega, z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right) \;=\; \frac{F\left(\omega, z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)}{\sum_{m=0}^{M-1} F\left(\omega_m, z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)} \tag{5.74}$$

and

$$F\left(\omega, z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right) = \exp\left(\frac{E_s}{N_0}\left(2\Re\left\{z\left(k;\mathbf{r},\mathbf{\Psi}\right)\omega^*\right\} - \left|\omega\right|^2\right)\right). \tag{5.75}$$

Since $\Pr\left[a\left(k\right)|\mathbf{r},\mathbf{\Psi}\right]$ depends on $\left(\mathbf{r},\mathbf{\Psi}\right)$ only through $z\left(k;\mathbf{r},\mathbf{\Psi}\right)$, this also holds for the a posteriori average of the symbol $a\left(k\right)$ in (5.62), which we will therefore denote as $\mu\left(k;z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)$ (instead of $\mu\left(k;\mathbf{z}\left(\mathbf{r},\mathbf{\Psi}\right)\right)$). We obtain:

$$\ell_i\left(\mathbf{\Psi};\mathbf{r}\right) \;\propto\; \frac{2E_s}{N_0}\sum_{k\in I_s}\Re\left\{z_i\left(k;\mathbf{r},\mathbf{\Psi}\right)\mu^*\left(k;z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)\right\}, \tag{5.76}$$

with

$$\mu\left(k;z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)=\left\{\begin{array}{ll} A_k & ,\, k\in I_p \\ \sum_{m=0}^{M-1} G\left(\omega_m, z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)\omega_m & ,\, k\in I_d \end{array}\right. . \tag{5.77}$$

The PA NCA FIM can therefore be computed as in Algorithm 3, with in Step 1 the computation of the APPs performed according to (5.73), and the computation of the a posteriori symbol averages $\mu\left(k;z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)$ as in (5.77). Appendix J on page 221 shows that the entries $J_{1,1}=J_{\nu,\nu}$, $J_{1,2}=J_{\nu,\theta}$, $J_{2,1}=J_{\theta,\nu}$ and $J_{2,2}=J_{\theta,\theta}$ of the PA NCA FIM can be written into the following form:

$$J_{\nu,\nu} = J_{\theta,\theta}\cdot 4\pi^2\left(\sigma_W^2+\left(\kappa_{m,W}-\kappa_0\right)^2\right)+\delta J_{\nu,\nu},\ \text{PA NCA}, \tag{5.78}$$

$$J_{\nu,\theta} = J_{\theta,\theta}\cdot 2\pi\left(\kappa_{m,W}-\kappa_0\right),\ \text{PA NCA}, \tag{5.79}$$

$$J_{\theta,\theta} = \frac{2N_W E_s}{N_0},\ \text{PA NCA}, \tag{5.80}$$

with

$$N_W=\sum_{k\in I_s} W\left(k\right), \tag{5.81}$$

$$\kappa_{m,W}=\frac{\sum_{k\in I_s} kW\left(k\right)}{N_W}, \tag{5.82}$$

$$\sigma_W^2=\frac{\sum_{k\in I_s}\left(k-\kappa_{m,W}\right)^2 W\left(k\right)}{N_W} \tag{5.83}$$

and

$$W\left(k\right)=\left\{\begin{array}{ll} 1 & ,\, k\in I_p \\ R_\Omega\left(E_s/N_0\right) & ,\, k\in I_d \end{array}\right. , \tag{5.84}$$

with

$$R_\Omega\left(\frac{E_s}{N_0}\right)=\frac{2E_s}{N_0}E\left[\Im\left\{\mu^*\left(k;z\left(k\right)\right)z\left(k\right)\right\}^2\right] \tag{5.85}$$

and $z\left(k\right)$ a short-hand notation of $z\left(k;\mathbf{r},\mathbf{\Psi}\right)$. The quantity $W\left(k\right)$ (5.84) serves as a *weight function* defined on the discrete set of indices I_s that is used to compute the *weighted cardinality* N_W (5.81) of I_s, the *weighted average* $\kappa_{m,W}$ (5.82) of the symbol indices $k\in I_s$, and the *weighted variance* (5.83) σ_W^2 of the symbol indices $k\in I_s$. When the term "weight" is interpreted in its physical sense, the time instant $\kappa_{m,W}T$ corresponds to the "*center of gravity*" of the symbol vector $\mathbf{a}$. The quantity $\delta J_{\nu,\nu}$ in (5.78) only occurs for the correct observation model. For the simplified (si) observation model we have $\delta J_{\nu,\nu}=0$, while for the correct (co) observation model:

$$\begin{aligned}\delta J_{\nu,\nu} &= 8\pi^2\frac{E_s}{N_0}\sum_{k,l\in I_p}\mathrm{f}\left(\left(k-l\right)T\right)\Re\left\{a^*\left(k\right)a\left(l\right)\right\} \\ &\quad +8\pi^2\frac{E_s}{N_0}N_d\mathrm{f}\left(0\right)Q_\Omega\left(E_s/N_0\right),\ \text{(co)},\ \text{PA NCA},\end{aligned} \tag{5.86}$$

with

$$\mathrm{f}(t) = \int \left(\frac{\tilde{t}}{T}\right)^2 p\left(\tilde{t} + \frac{t}{2}\right) p\left(\tilde{t} - \frac{t}{2}\right) \mathrm{d}\tilde{t} \tag{5.87}$$

and

$$Q_\Omega\left(\frac{E_s}{N_0}\right) = E\left[\left|\mu\left(k; z\left(k\right)\right)\right|^2\right]. \tag{5.88}$$

Note that $R_\Omega\left(E_s/N_0\right)$ from (5.85) and $Q_\Omega\left(E_s/N_0\right)$ from (5.88) depend on E_s/N_0 and on the type and size of the constellation, but not on the shape of the transmit pulse. The shape of the transmit pulse $p\left(t\right)$ affects only the quantity $J_{\nu,\nu}$ involving $\mathrm{f}\left(t\right)$. Depending on the constellation type some further simplifications are possible:

- It is easily verified from (5.75) that, for a *real-valued* M-PAM constellation, the quantities $R_\Omega\left(E_s/N_0\right)$ from (5.85) and $Q_\Omega\left(E_s/N_0\right)$ from (5.88) further reduce to:

$$R_{M-PAM}\left(\frac{E_s}{N_0}\right) = Q_{M-PAM}\left(\frac{E_s}{N_0}\right), \tag{5.89}$$

$$= E\left[\left(\sum_{m=0}^{M-1} G\left(\omega_m, \Re\left\{z\left(k\right)\right\}\right)\omega_m\right)^2\right], \tag{5.90}$$

  where the expectation operator $E\left[.\right]$ denotes averaging with respect to the real part of $z\left(k\right)$ only rather than $z\left(k\right)$. We recall that $\Re\left\{z\left(k\right)\right\} = a\left(k\right) + n_r\left(k\right)$, where $a\left(k\right)$, $k \in I_d$ is a real-valued random variable that takes any value from the M-PAM symbol alphabet Ω with equal probability M^{-1}, and $n_r\left(k\right)$ is real-valued zero-mean Gaussian noise with a variance equal to $N_0/2E_s$.

- An M-QAM constellation can be viewed as the orthogonal superposition of two $\sqrt{M}$-PAM constellations that are scaled down by a factor $\sqrt{2}$. Taking this into account, it is easily shown that the quantities $R_{M-QAM}\left(.\right)$ and $Q_{M-QAM}\left(.\right)$ for M-QAM are related to the quantity $R_{\sqrt{M}-PAM}\left(.\right)$ for $\sqrt{M}$-PAM by:

$$R_{M-QAM}\left(\frac{E_s}{N_0}\right) = \left(1 + \frac{E_s}{N_0}\right) R_{\sqrt{M}-PAM}\left(\frac{E_s}{2N_0}\right) - \frac{E_s}{N_0}, \tag{5.91}$$

$$Q_{M-QAM}\left(\frac{E_s}{N_0}\right) = Q_{\sqrt{M}-PAM}\left(\frac{E_s}{2N_0}\right). \tag{5.92}$$

- For an M-PSK constellation with $M > 4$, no simplification of (5.85) and (5.88) seems possible.

The evaluation of the PA NCA FIM entries (5.78)-(5.80) involves replacing in (5.85),(5.88) (for M-PSK) or (5.90) (for M-PAM and M-QAM) the statistical average $E\left[.\right]$ by an arithmetical average over a large number of realizations of $z\left(k\right)$ or $\Re\left\{z\left(k\right)\right\}$, respectively. This yields the computation algorithms that are described in Algorithm 4 (for M-PSK) and Algorithm 5 (for M-PAM and M-QAM).

Algorithm 4 PA NCA FIM related to $L(\mathbf{\Psi};\mathbf{r})$ for linear modulation

1. For $n = 1, 2, ..., N$:
   - generate $a^{(n)}$ according to the distribution $\Pr[a = \omega]$, for $\omega \in \Omega$;
   - generate $n^{(n)}$ according to $\mathcal{N}_c\left(0, \frac{N_0}{E_s}\right)$;
   - compute $z^{(n)} = a^{(n)} + n^{(n)}$;
   - compute $G\left(\omega, z^{(n)}\right)$ for $\omega \in \Omega$ according to (5.74).
2. Compute $R_\Omega\left(\frac{E_s}{N_0}\right)$ and $Q_\Omega\left(\frac{E_s}{N_0}\right)$ as:

$$R_\Omega\left(\frac{E_s}{N_0}\right) = \frac{1}{N}\sum_{n=1}^{N}\frac{2E_s}{N_0}\Im\left(\sum_{m=0}^{M-1} G\left(\omega_m, z^{(n)}\right)\omega_m^* z^{(n)}\right)^2,$$

$$Q_\Omega\left(\frac{E_s}{N_0}\right) = \frac{1}{N}\sum_{n=1}^{N}\left|\sum_{m=0}^{M-1} G\left(\omega_m, z^{(n)}\right)\omega_m\right|^2.$$

3. Compute N_W, $\kappa_{m,W}$, σ_W^2 and $\delta J_{\nu,\nu}$ according to (5.81)-(5.84) and (5.86)-(5.87).
4. Compute $J_{\nu,\nu}$, $J_{\nu,\theta}$ and $J_{\theta,\theta}$ according to (5.78)-(5.80).

Algorithm 5 PA NCA CRBs related to $L(\mathbf{\Psi};\mathbf{r})$ for linear modulation (M-PAM, M^2-QAM)

1. For $n = 1, 2, ..., N$:
   - generate $a_r^{(n)}$ according to the distribution $\Pr[a_r = \omega]$, for $\omega \in M$-PAM;
   - generate $n_r^{(n)}$ according to $\mathcal{N}\left(0, \frac{N_0}{2E_s}\right)$;
   - compute $z_r^{(n)} = a_r^{(n)} + n_r^{(n)}$;
   - compute $G\left(\omega, z_r^{(n)}\right)$ for $\omega \in M$-PAM according to (5.74).
2. Compute:
$$X\left(\frac{E_s}{N_0}\right) = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{\omega\in\Omega_{M-PAM}} G\left(\omega, z_r^{(n)}\right)\omega\right)^2.$$
3. Compute $R_{M-\mathrm{PAM}}\left(\frac{E_s}{N_0}\right)$ and $Q_{M-\mathrm{PAM}}\left(\frac{E_s}{N_0}\right)$ as:
$$R_{M-\mathrm{PAM}}\left(\frac{E_s}{N_0}\right) = Q_{M-\mathrm{PAM}}\left(\frac{E_s}{N_0}\right) = X\left(\frac{E_s}{N_0}\right)$$
   or $R_{M^2-\mathrm{QAM}}\left(\frac{E_s}{N_0}\right)$ and $Q_{M^2-\mathrm{QAM}}\left(\frac{E_s}{N_0}\right)$ as:
$$R_{M^2-QAM}\left(\frac{E_s}{N_0}\right) = \left(1+\frac{E_s}{N_0}\right)X\left(\frac{E_s}{2N_0}\right) - \frac{E_s}{N_0},$$
$$Q_{M^2-QAM}\left(\frac{E_s}{N_0}\right) = X\left(\frac{E_s}{2N_0}\right).$$
4. Compute N_W, $\kappa_{m,W}$, σ_W^2 and $\delta J_{\nu,\nu}$ according to (5.81)-(5.84) and (5.86)-(5.87).
5. Compute $J_{\nu,\nu}$, $J_{\nu,\theta}$ and $J_{\theta,\theta}$ according to (5.78)-(5.80).

Remarks

The remarks formulated in subsection 5.7.6 pertaining to the pilot-symbol/data-symbol decomposition of the FIM, also hold for the PA NCA FIM from (5.78)-(5.80). In particular, when $I_p \neq \emptyset$, the PA NCA low-SNR asymptotic FIM equals the DA (M)FIM with entries (5.25)-(5.27).

5.7.8 Computing the NPA NCA FIM and the Associated CRBs for Linear Modulation

It follows from (5.16) and (5.17) that the NPA NCA FIM is a special case of the PA NCA FIM, namely for pilot symbol free transmission. The procedure for computing the NPA NCA FIM is therefore the same one as for the PA NCA FIM (Algorithms 4 and 5), with $(I_p, I_d) = (\emptyset, I_p \bigcup I_d)$.

5.7.9 Derivation of the CRBs

It follows from the Bayesian Cramer-Rao inequality (3.21) that the MSEs regarding to the estimation of ν and θ are lower-bounded by the corresponding diagonal elements of the inverse of the FIM (5.47). Similarly, the Bayesian Cramer-Rao inequality (3.21) also yields a lower bound on the MSE regarding the estimation of the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$. We obtain: $E\left[(\nu-\hat{\nu})^2\right] \geq CRB_\nu$, $E\left[\left(\theta-\hat{\theta}\right)^2\right] \geq CRB_\theta$ and

$$E\left[\left(\Theta(kT;\mathbf{\Psi}) - \Theta\left(kT;\hat{\mathbf{\Psi}}\right)\right)^2\right] \geq CRB_{\Theta(kT;\mathbf{\Psi})},$$

with

$$CRB_\nu = \frac{J_{\theta,\theta}}{J_{\theta,\theta}J_{\nu,\nu} - (J_{\theta,\nu})^2}, \tag{5.93}$$

$$CRB_\theta = \frac{J_{\nu,\nu}}{J_{\theta,\theta}J_{\nu,\nu} - (J_{\theta,\nu})^2} \tag{5.94}$$

and

$$CRB_{\Theta(kT;\mathbf{\Psi})} = CRB_\theta|_{\kappa_0=k}\,. \tag{5.95}$$

where κ_0 is the normalized time instant at which the phase shift is estimated, and $J_{1,1} = J_{\nu,\nu}$, $J_{1,2} = J_{\nu,\theta}$ and $J_{2,2} = J_{\theta,\theta}$ are the FIM entries (5.48). The above results hold for both observation models and for the different estimation modes. For the PA CA mode, no further simplifications are possible. For the DA, PA NCA and NPA NCA modes, the expressions of the FIM entries obtained in subsections 5.7.6-5.7.8 take a form that allows further analytical computation of the CRBs (5.93)-(5.95). We will further discuss this in subsections 5.7.10-5.7.12.

5.7.10 Derivation of the DA CRBs

It follows directly from (5.70) that the DA CRBs regarding the estimation of ν, θ and $\Theta\left(kT;\mathbf{\Psi}\right)$ equal the corresponding DA MCRBs, which, for large values of N_p, are given by (5.33)-(5.35). Furthermore, when $I_p \neq \emptyset$, these DA (M)CRBs are the low-SNR asymptotes (ACRBs) of the CRBs from (5.93)-(5.95) (see remark at the end of section 5.7.6). In other words:

$$ACRB_{\nu} = DACRB_{\nu} = MCRB_{\nu}\left(N_p\right),\ I_p \neq \emptyset, \tag{5.96}$$

$$ACRB_{\theta} = DACRB_{\theta} = MCRB_{\theta}\left(N_p\right),\ I_p \neq \emptyset, \tag{5.97}$$

with $MCRB_{\nu}\left(N_p\right)$ from (5.33) and $MCRB_{\theta}\left(N_p\right)$ from (5.34). Equations (5.96) and (5.97) hold for both observation models and for the different estimation modes.

5.7.11 Derivation of the PA NCA CRBs

Substituting (5.78)-(5.80) into (5.93) and (5.94) yields:

$$CRB_{\nu} = \frac{N_0}{8\pi^2 N_W E_s\left(\sigma_W^2 + \delta\tilde{J}_{\nu,\nu}\right)}, \tag{5.98}$$

$$CRB_{\theta} = \frac{N_0}{2N_W E_s}\left(\frac{\sigma_W^2 + \left(\kappa_{m,W} - \kappa_0\right)^2 + \delta\tilde{J}_{\nu,\nu}}{\sigma_W^2 + \delta\tilde{J}_{\nu,\nu}}\right), \tag{5.99}$$

with N_W from (5.81), $\kappa_{m,W}$ from (5.82), σ_W^2 from (5.83) and $\delta\tilde{J}_{\nu,\nu}$ a short-hand notation for

$$\delta\tilde{J}_{\nu,\nu} = \delta J_{\nu,\nu}\left(\frac{N_0}{8\pi^2 N_W E_s}\right),\ \text{PA NCA}, \tag{5.100}$$

with $\delta J_{\nu,\nu}$ equal to zero for the simplified observation model and given by (5.86) for the correct observation model.

5.7.12 Derivation of the NPA NCA CRBs

The closed-form expression for the PA NCA CRBs (5.98)-(5.99) also holds for the NPA NCA estimation mode, provided that the quantities N_W (5.81), $\kappa_{m,W}$ (5.82), σ_W^2 (5.83) and $\delta J_{\nu,\nu}$ (5.86) are evaluated for $(I_p, I_d) = (\emptyset, I_p \bigcup I_d)$. We obtain:

$$N_W = N_s \cdot R_{\Omega}\left(E_s/N_0\right),\ \text{NPA NCA}, \tag{5.101}$$

$$\kappa_{m,W} = \kappa_{m,s},\ \text{NPA NCA}, \tag{5.102}$$

$$\sigma_W^2 = \frac{\left(N_s^2 - 1\right)}{12} \approx \frac{N_s^2}{12},\ \text{NPA NCA}, \tag{5.103}$$

$$\delta J_{\nu,\nu} = 8\pi^2 \frac{E_s}{N_0} N_s \mathrm{f}\left(0\right) Q_{\Omega}\left(E_s/N_0\right),\ (\mathrm{co}),\ \text{NPA NCA}, \tag{5.104}$$

with $R_\Omega(E_s/N_0)$ from (5.85), $\kappa_{m,s}$ from (5.24), $Q_\Omega(E_s/N_0)$ from (5.88), and $\mathrm{f}(t)$ defined as in (5.87). Substituting (5.101)-(5.104) into (5.98)-(5.99) yields:

$$CRB_\nu \approx \frac{1}{R_\Omega(E_s/N_0)} \cdot \frac{3N_0}{2\pi^2 E_s N_s \left(N_s^2 + 3\delta\tilde{J}_{\nu,\nu}\right)} \tag{5.105}$$

and

$$CRB_\theta \approx \frac{1}{R_\Omega(E_s/N_0)} \cdot \frac{N_0}{2N_s E_s} \left(\frac{N_s^2 + 12\left(\kappa_{m,s} - \kappa_0\right)^2 + 12\delta\tilde{J}_{\nu,\nu}}{N_s^2 + 12\delta\tilde{J}_{\nu,\nu}}\right), \tag{5.106}$$

where the approximations hold for large N_s, and $\delta\tilde{J}_{\nu,\nu}$ is given by:

$$\delta\tilde{J}_{\nu,\nu} = \delta J_{\nu,\nu}\left(\frac{1}{R_\Omega(E_s/N_0)} \cdot \frac{N_0}{8\pi^2 N_s E_s}\right), \text{ NPA NCA}, \tag{5.107}$$

with $\delta J_{\nu,\nu}$ equal to zero for the simplified observation model, and given by (5.104) for the correct observation model.

5.7.13 Numerical Evaluation and Discussion of the CRBs

Assuming a square-root cosine roll-off transmit pulse with a 30% excess bandwidth, we will compute the CRBs and their low-SNR asymptotes (ACRBs) for the PA CA, the PA NCA and the NPA NCA estimation modes. The ACRBs are obtained analytically using closed-form expressions that are available in the literature. The computation of CRBs involves the numerical evaluation of the FIM entries in (5.48). Algorithm 3 is used to evaluate the PA CA FIM, Algorithm 4 is used to evaluate the PA NCA and the NPA NCA FIMs for PSK, and Algorithm 5 is used to evaluate the PA NCA and the NPA NCA FIMs for PAM and QAM. Step 2 of all three algorithms involves the evaluation of one or more numerical averages. Note that these averages are independent of the transmit pulse and with respect to either a complex-valued vector of size N_s (for the PA CA mode), a complex-valued scalar (for the PA NCA and the NPA NCA modes, and for PSK) or a real-valued scalar (for the PA NCA and the NPA NCA modes, and for PAM and QAM).

The results are presented and discussed in sections 5.7.14-5.7.16. We will use

$$(A)CRB_u^{(co)} \tag{5.108}$$

and

$$(A)CRB_u^{(si)} \tag{5.109}$$

to denote the (A)CRB that pertains to the estimation of u with u equal to ν, θ or $\Theta(kT;\mathbf{\Psi})$, and results from the correct (referred to as (co)) or the simplified (referred to as (si)) observation model, respectively. Instead of analyzing the (A)CRBs (5.108)-(5.109) directly, it is often more convenient to examine the ratios

$$\frac{(A)CRB_u^{(co)}}{MCRB_u(N_s)}, \tag{5.110}$$

$$\frac{(A)CRB_u^{(si)}}{MCRB_u\left(N_s\right)}, \tag{5.111}$$

where $MCRB_u\left(N_s\right)$ denotes the MCRB related to the PA CA, PA NCA or NPA NCA estimation of u with u equal to ν, θ or $\Theta\left(kT;\mathbf{\Psi}\right)$. We recall from section 5.5.2 that, for $N_s \gg 1$, the MCRBs regarding the estimation of ν, θ and $\Theta\left(kT;\mathbf{\Psi}\right)$ are equal for the PA CA, the PA NCA and the NPA NCA estimation modes and do not depend on the observation model being considered. We further recall from the remark at the end of section 5.7.6 that the ratios (5.110)-(5.111) can be interpreted as a measure of the penalty that is caused by not a priori knowing the value of the data symbols at the receiver.

5.7.14 Numerical Evaluation and Discussion of the NPA NCA CRBs

Results pertaining to the NPA NCA CRBs are presented for $\kappa_0 = \kappa_{m,s}$ (i.e., the phase shift is estimated at the center of the symbol burst). Imposing this condition, one obtains from (5.104)-(5.107) and (5.30)-(5.31):

$$\begin{aligned}\frac{CRB_\nu^{(si)}}{MCRB_\nu\left(N_s\right)} &= \frac{CRB_\theta^{(si)}\Big|_{\kappa_0=\kappa_{m,s}}}{MCRB_\theta\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}} = \frac{CRB_\theta^{(co)}\Big|_{\kappa_0=\kappa_{m,s}}}{MCRB_\theta\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}}, \\ &= \frac{1}{R_\Omega\left(\frac{E_s}{N_0}\right)},\end{aligned} \tag{5.112}$$

$$\frac{CRB_\nu^{(co)}}{MCRB_\nu\left(N_s\right)} = \frac{N_s^2}{N_s^2 R_\Omega\left(\frac{E_s}{N_0}\right) + 12\mathrm{f}\left(0\right) Q_\Omega\left(\frac{E_s}{N_0}\right)}, \tag{5.113}$$

with $R_\Omega\left(E_s/N_0\right)$ from (5.85), $Q_\Omega\left(E_s/N_0\right)$ from (5.88), and $\mathrm{f}\left(t\right)$ defined as in (5.87). The low-SNR asymptotes of the considered NPA NCA CRBs have been derived in [2, 57]. We obtain:

$$\begin{aligned}\frac{ACRB_\nu^{(si)}}{MCRB_\nu\left(N_s\right)} &= \frac{ACRB_\theta^{(si)}\Big|_{\kappa_0=\kappa_{m,s}}}{MCRB_\theta\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}} = \frac{ACRB_\theta^{(co)}\Big|_{\kappa_0=\kappa_{m,s}}}{MCRB_\theta\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}}, \\ &= \frac{K!}{K^2\left|S_K\right|^2}\left(\frac{N_0}{E_s}\right)^{K-1},\end{aligned} \tag{5.114}$$

$$\frac{ACRB_\nu^{(co)}}{MCRB_\nu\left(N_s\right)} = \begin{cases} \frac{N_0}{2E_s} & , K=2 \\ \frac{1}{12\mathrm{f}(0)}\left(N_s\right)^2\frac{N_0}{E_s} & , K>2 \end{cases}, \tag{5.115}$$

where K is related to the symmetry angle $2\pi/K$ of the constellation ($K=2$ for M-PAM, $K=4$ for M-QAM, $K=M$ for M-PSK) and $S_K = E\left[\left(a\left(k\right)\right)^K\right]$.

In Fig. 5.6 we have plotted (5.112) (curves indicated with (*)) and (5.113)

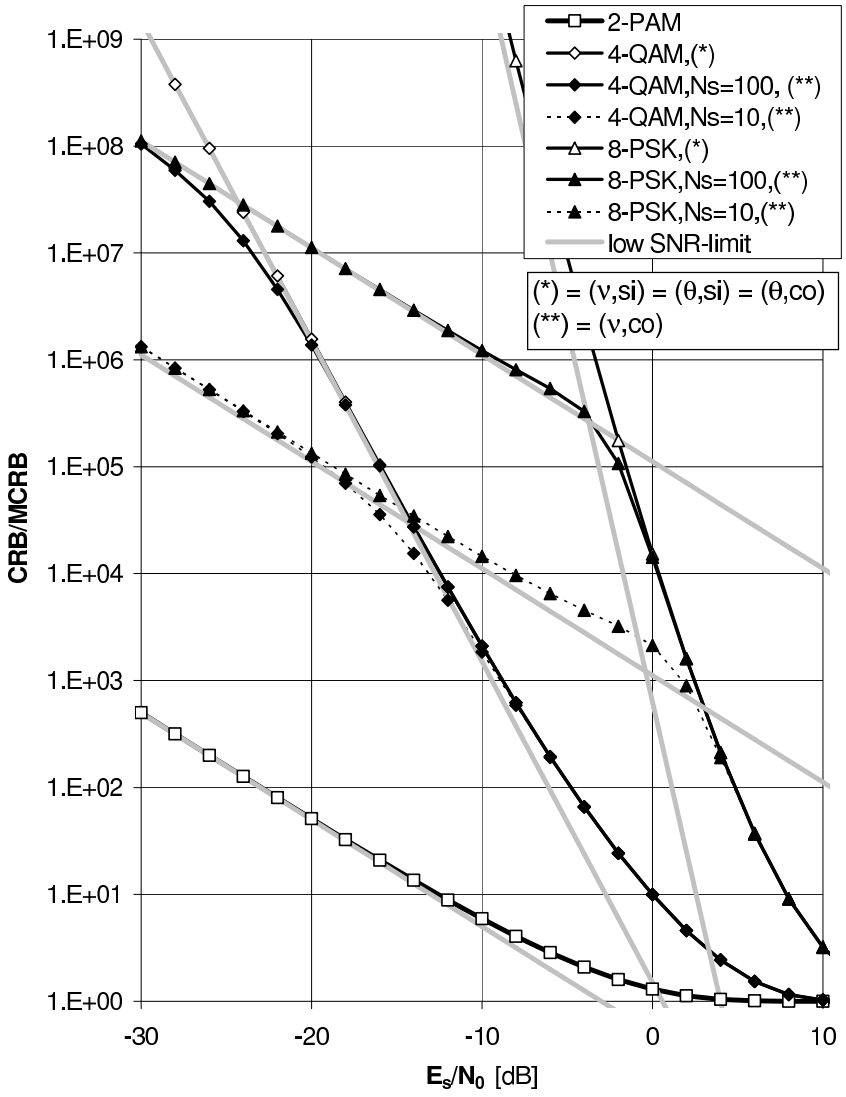

Figure 5.6: Ratio CRB/MCRB related to $L(\mathbf{\Psi};\mathbf{r})$, resulting from the correct (co) or the simplified (si) observation model and pertaining to the NPA NCA estimation of the normalized frequency offset ν and the phase shift θ at the center of the burst, as a function of the SNR E_s/N_0 for random 2-PAM, 4-QAM and 8-PSK.

(curves indicated with (**)) as a function of E_s/N_0, for 2-PAM, 4-QAM and 8-PSK, and for observation intervals of $N_s = 10, 100$ symbol periods (see (5.112)-(5.115)). The considered range of E_s/N_0 contains extremely low values because it was our intention to illustrate that the numerically computed ratios CRB/MCRB are consistent with their analytically obtained low-SNR limits ACRB/MCRB. The following observations can be made about these results:

- The ratios ACRB/MCRB (5.114) resulting from the simplified observation model are inversely proportional to $(N_0/E_s)^{K-1}$.
- The ratio CRB/MCRB (5.112) regarding the estimation of the phase shift θ at $t/T = \kappa_{m,s}$ does not depend on the observation model.
- For the ratio CRB/MCRB (5.112)-(5.113) regarding the estimation of the normalized frequency offset ν and resulting from the correct observation model, we have to make a distinction between the cases $K > 2$ and $K = 2$. For $K > 2$ (complex-valued PSK and QAM constellations), the correct and the simplified observation models result in a considerably different CRB_ν for very small E_s/N_0 (see (5.114)-(5.115)). For the correct observation model, the ratio $ACRB_\nu/MCRB_\nu\,(N_s)$ is inversely proportional to N_0/E_s, whereas, for the simplified observation model, the ratio $ACRB_\nu/MCRB_\nu\,(N_s)$ is inversely proportional to $(N_0/E_s)^{K-1}$ with $K > 2$. This indicates that normalized frequency offset estimation from the simplified observation model is highly inaccurate at very low E_s/N_0, as far as complex-valued PSK and QAM constellations are concerned. Conversely, for $K = 2$ (real-valued PSK and PAM constellations), we find that for large N_s both observation models yield essentially the same ratio $ACRB_\nu/MCRB_\nu\,(N_s)$. The results for 2-PAM presented in Fig. 5.6 show that both observation models also yield essentially the same ratio $CRB_\nu/MCRB_\nu\,(N_s)$, irrespective of the number of observed samples. We have verified (results not reported here) that the same behavior also applies to higher order M-PAM constellations.
- It follows from:

$$\mathrm{f}\,(0) = \int \left(\frac{\tilde{t}}{T}\right)^2 p^2\left(\tilde{t}\right) \mathrm{d}\tilde{t} \geq 0$$

and

$$Q_\Omega\,(E_s/N_0) = E\left[\left|\mu\,(k; z\,(k))\right|^2\right] \geq 0,$$

that the CRB pertaining to normalized frequency offset estimation is smaller for the correct observation model than for the simplified observation model (see (5.112)-(5.113)). This is consistent with the fact that, in the presence of a non-zero normalized frequency offset ν, carrier recovery from the simplified observation vector using (5.12) is sub-optimum. However, $CRB_\nu^{(co)}$ converges to $CRB_\nu^{(si)}$ when, for a given value of E_s/N_0, the number N_s of transmitted symbols goes to infinity (also see (5.112)-(5.113)). The results for 4-QAM and 8-PSK presented in Fig. 5.6 indicate

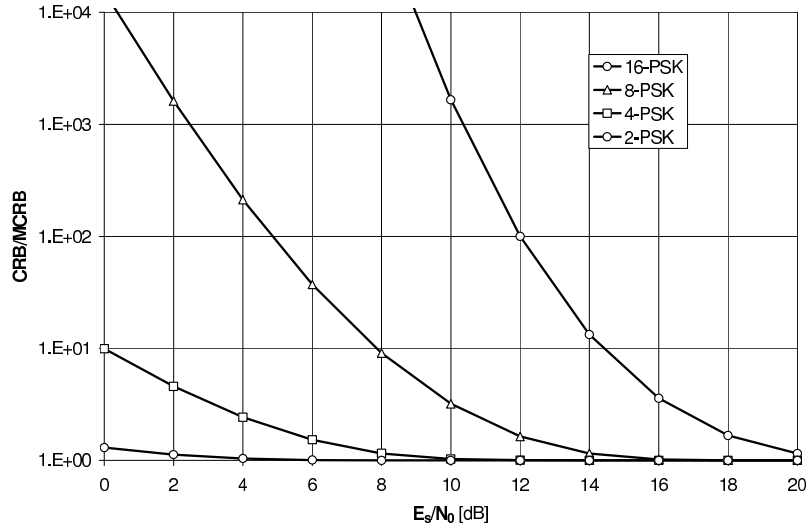

Figure 5.7: Ratio CRB/MCRB related to $L(\mathbf{\Psi};\mathbf{r})$ for the NPA NCA estimate of the normalized frequency offset or the phase shift at $t/T = \kappa_{m,s}$, for M-PSK.

that, for all SNR values of practical interest (let us assume with a BER of less than about 10^{-3}(see Fig. 4.7)) both observation models yield the same ratio CRB/MCRB for NPA NCA frequency estimation, which is essentially independent of the number of observed symbol intervals and of the shape of the transmit pulse. We have verified (results not shown here) that the same behavior also applies to higher order M-QAM and M-PSK constellations.

Assuming the simplified observation model, Fig. 5.7, Fig. 5.8 and Fig. 5.9 show more results for M-PSK, M-QAM and M-PAM, respectively. We have verified that, for the ranges of E_s/N_0 and of CRB/MCRB considered, the correct observation model yields the same curves. The following observations can be made:

- For all three constellation types (PAM,QAM,PSK), we observe that for a given value of E_s/N_0 the ratio CRB/MCRB increases with M, which indicates that for the larger constellations carrier recovery is inherently harder to accomplish. This effect is clearly evident for M-PSK, less visible for M-QAM, in which case the curves corresponding to large M exhibit an almost horizontal portion, and almost unnoticeable for M-PAM. Fig. 5.8 for QAM and Fig. 5.9 for PAM also show the limiting curve for M approaching infinity; this situation corresponds to data symbols that are continuous random variables, that are uniformly distributed in the interval $[-\sqrt{3},\sqrt{3}]$ for PAM and in a square with side $\sqrt{6}$ for QAM. For PSK

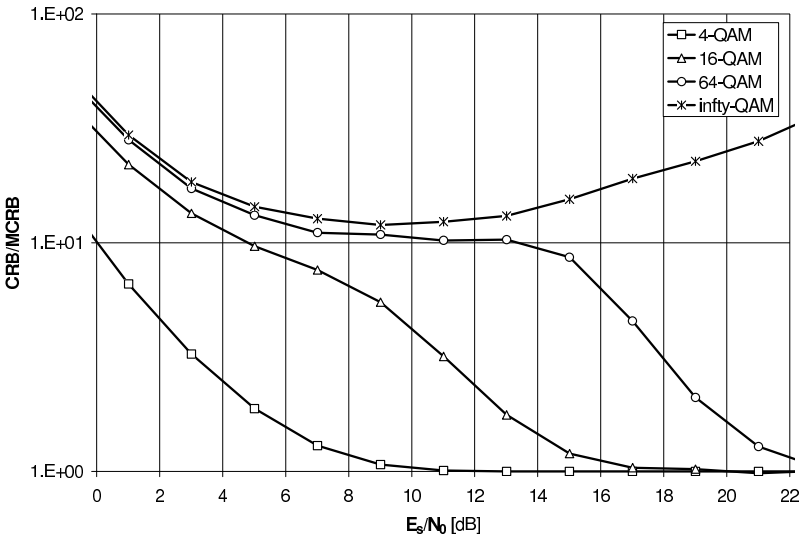

Figure 5.8: Ratio CRB/MCRB related to $L(\mathbf{\Psi};\mathbf{r})$ for the NPA NCA estimate of the normalized frequency offset or the phase shift at $t/T = \kappa_{m,s}$, for M-QAM.

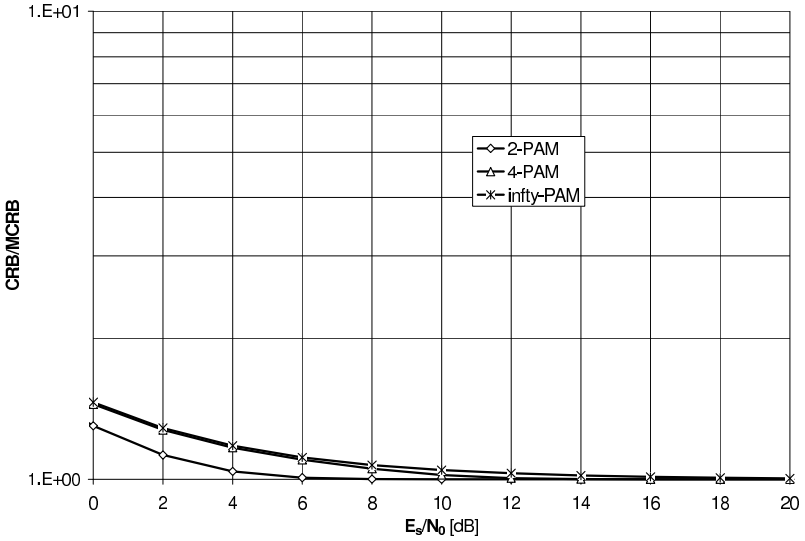

Figure 5.9: Ratio CRB/MCRB related to $L(\mathbf{\Psi};\mathbf{r})$ for the NPA NCA estimate of the normalized frequency offset or the phase shift at $t/T = \kappa_{m,s}$, for M-PAM.

the curve for M approaching infinity (data symbols uniformly distributed over the unit circle) falls outside the range displayed in Fig. 5.7. In the case of infinite-size constellations the CRBs do not necessarily converge to the corresponding MCRBs for large SNR; according to [54]. This is due to the non-diagonal nature of the FIM, related to the joint likelihood function $p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})$ of $\mathbf{a}$ and $\mathbf{\Psi}$.

- For finite M, the CRB does converge to the MCRB when E_s/N_0 is sufficiently large. The value of E_s/N_0, at which CRB is close to MCRB, increases by about 6 dB when M doubles (PAM, PSK) or quadruples (QAM). This is consistent with the observation in Appendix K on page 225, namely that for uncoded pilot-symbol-free transmission, the convergence of the CRB to the MCRB is mainly determined by the value of $\frac{E_s}{N_0}(d_M)^2$, with d_M denoting the minimum Euclidean distance between the constellation points. It is easily verified from Fig. 4.7, in which the SER curves corresponding to uncoded transmission with perfect synchronization are shown, that the convergence of the CRBs to the MCRBs occurs at a value of $(E_s/N_0)_{thr}$ which corresponds to a SER in the order of 10^{-3}, irrespective of the constellation. This too has been motivated in Appendix K. For these reasons, we can claim that as far as the CRB for carrier phase shift or frequency offset estimation is concerned, transmission at a SER of less than 10^{-3} is nearly equivalent to transmitting a sequence of known pilot symbols. At the normal operating SNR of uncoded digital communication systems, these CRBs are therefore very well approximated by the corresponding MCRBs.

5.7.15 Numerical Evaluation and Discussion of the PA NCA CRBs

The results pertaining to the PA NCA CRBs will be presented for the simplified observation model and for a symbol vector $\mathbf{a}$ that contains two blocks of $N_p/2$ consecutive pilot symbols spaced with $S = \epsilon_{Sp}N_p$ data symbols. Note that $0 \leq S \leq N_d$ with $N_d = N_s - N_p$ the total number of data symbols in the burst, or, equivalently, $0 \leq \epsilon_{Sp} \leq \left(\lambda_p^{-1} - 1\right)$ where $\lambda_p = N_p/N_s$ is the pilot symbol ratio. The quantity ϵ_{Sp} is referred to as the *spacing ratio*. Again an asymmetric and a symmetric burst structure are considered, both of which are shown in Fig. 5.5 where the shaded areas indicate the location of the pilot symbols within the symbol vector $\mathbf{a}$. In the case of the asymmetric burst structure #1, the center of gravity $\kappa_{m,W}$ from (5.82) is a function of the SNR. It follows directly from (5.82), (5.112) and Fig. 5.6 that, for very low SNR, $\kappa_{m,W}$ equals $\kappa_{m,p}$ (i.e., the center of gravity of the pilot symbols from (5.28)), whereas, for very high SNR, $\kappa_{m,W}$ equals $\kappa_{m,s}$ (i.e., the center of the transmitted symbol vector $\mathbf{a}$). In the case of the symmetric burst structure #2, $\kappa_{m,W} = \kappa_{m,p} = \kappa_{m,s}$ (independent of the SNR).

5.7.15.1 The CRB Regarding the Estimation of the Phase Shift at $\kappa_0 = \kappa_{m,W}$.

Fig. 5.10 corresponds to the minimum phase shift estimation error, showing the ratio of the minimum value of the PA NCA CRB from (5.99) to the minimum value of the PA NCA MCRB from (5.31). Note that the minimum value of (5.99) is obtained when $\kappa_0 = \kappa_{m,W}$, whereas the minimum value of (5.31) is obtained when $\kappa_0 = \kappa_{m,s}$. Using (5.99), (5.31) and (5.81), we can write:

$$\frac{CRB_{\theta}^{(si)}\Big|_{\kappa_0=\kappa_{m,W}}}{MCRB_{\theta}\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}} = \frac{N_s}{N_W} = \left(\lambda_p + R_{\Omega}\left(\frac{E_s}{N_0}\right)(1-\lambda_p)\right)^{-1}. \tag{5.116}$$

For a given constellation Ω and a given value of E_s/N_0, (5.116) only depends on the pilot symbol ratio $\lambda_p = N_p/N_s$, in such a way that the curves for all burst structures with the same λ_p coincide. Results are presented for 4-PSK, and λ_p equal to 0.1 or 0.2. For the sake of comparison, the ratio CRB/MCRB (5.112) related to NPA NCA estimation (or, equivalently, $\lambda_p = 0$) is also displayed. We come to the following findings: At very high SNR, the ratio CRB/MCRB (5.116) converges to 1. At low and intermediate SNR, the ratio CRB/MCRB (5.116) decreases as λ_p increases. At very low SNR, the ratio CRB/MCRB (5.116) converges to its low-SNR asymptote that, for $\lambda_p > 0$, is given by:

$$\lim_{E_s/N_0 \to 0} \frac{CRB_{\theta}^{(si)}\Big|_{\kappa_0=\kappa_{m,W}}}{MCRB_{\theta}\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}} = \left(\lambda_p\right)^{-1}. \tag{5.117}$$

To arrive at (5.117) we made use of (5.116) and

$$\lim_{E_s/N_0 \to 0} R_{\Omega}\left(E_s/N_0\right) = 0,$$

which follows directly from (5.112) and Fig. 5.6. Alternatively, the low-SNR limit (5.117) can also be obtained by taking into account the remark at the end of subsection 5.7.7 concerning the PA NCA low-SNR asymptotic CRBs. When $I_p \neq \emptyset$, the low-SNR asymptote of $CRB_{\theta}^{(si)}$ equals the DA $MCRB_{\theta}\left(N_p\right)$ from (5.34). This $MCRB_{\theta}\left(N_p\right)$ achieves its minimum value when $\kappa_0 = \kappa_{m,p}$, which in turn is the low-SNR limit of $\kappa_{m,W}$. As a result:

$$\frac{CRB_{\theta}^{(si)}\Big|_{\kappa_0=\kappa_{m,W}}}{MCRB_{\theta}\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}} \rightarrow \frac{MCRB_{\theta}\left(N_p\right)|_{\kappa_0=\kappa_{m,p}}}{MCRB_{\theta}\left(N_s\right)|_{\kappa_0=\kappa_{m,s}}} \tag{5.118}$$

for low values of E_s/N_0. Taking into account (5.34) and (5.31) again yields (5.117).

5.7.15.2 The CRB Regarding the Estimation of the Normalized Frequency Offset

Figs. 5.11 and 5.12 correspond to the normalized frequency offset estimation

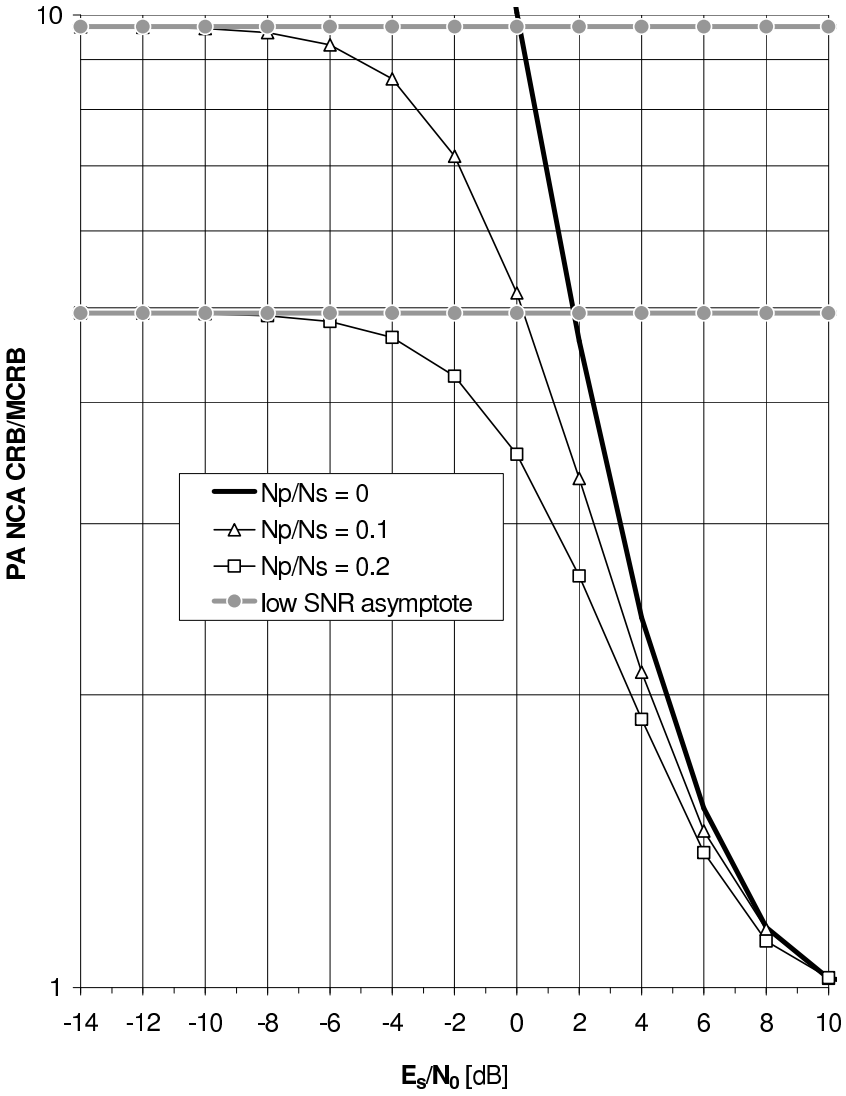

Figure 5.10: Ratio of the CRB related to $L(\mathbf{\Psi};\mathbf{r})$, resulting from the simplified observation model and pertaining to the PA NCA estimate of the phase shift at $\kappa_0 = \kappa_{m,W}$ to the MCRB related to the PA NCA estimate of the phase shift at $\kappa_0 = \kappa_{m,s}$, assuming a 4-PSK signaling constellation.

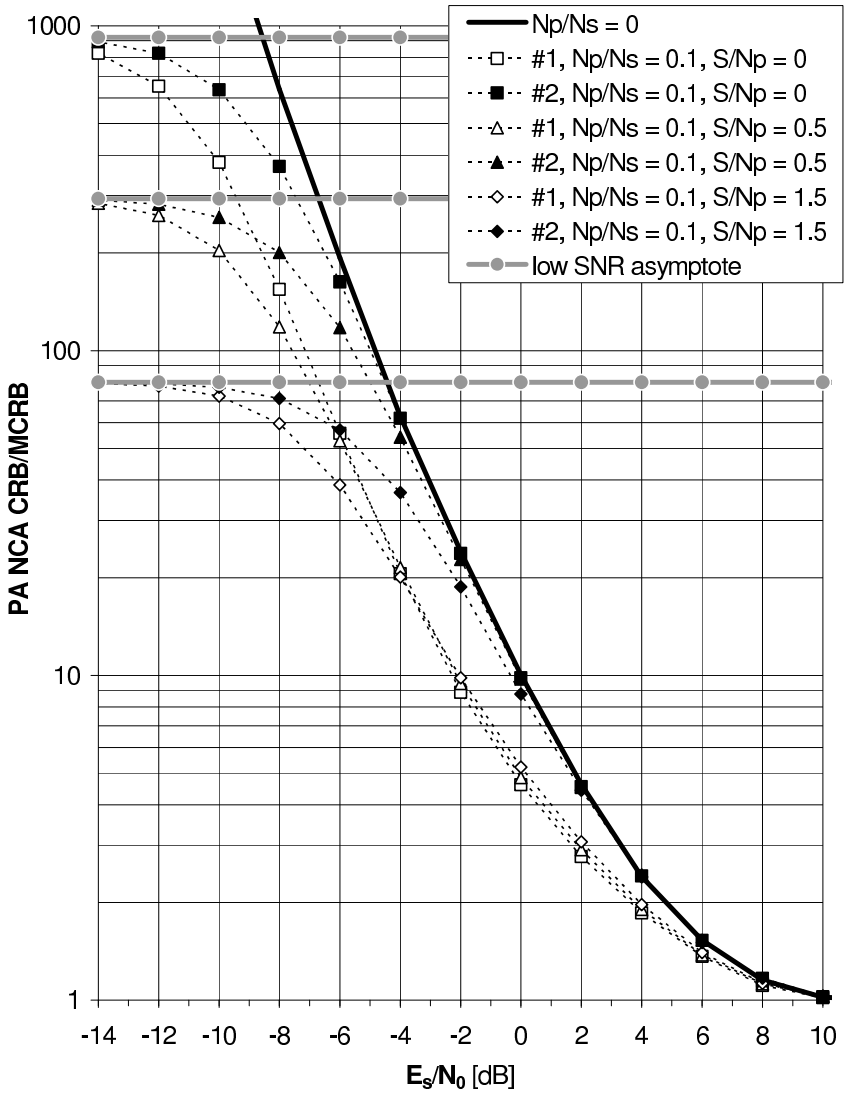

Figure 5.11: Ratio CRB/MCRB related to $L(\mathbf{\Psi}; \mathbf{r})$, resulting from the simplified observation model and pertaining to the PA NCA estimate of the normalized frequency offset, assuming a 4-PSK signaling constellation and $\lambda_p = 0.1$. An asymmetric (#1) and a symmetric (#2) burst structure are considered, both of which are shown in Fig. 5.5.

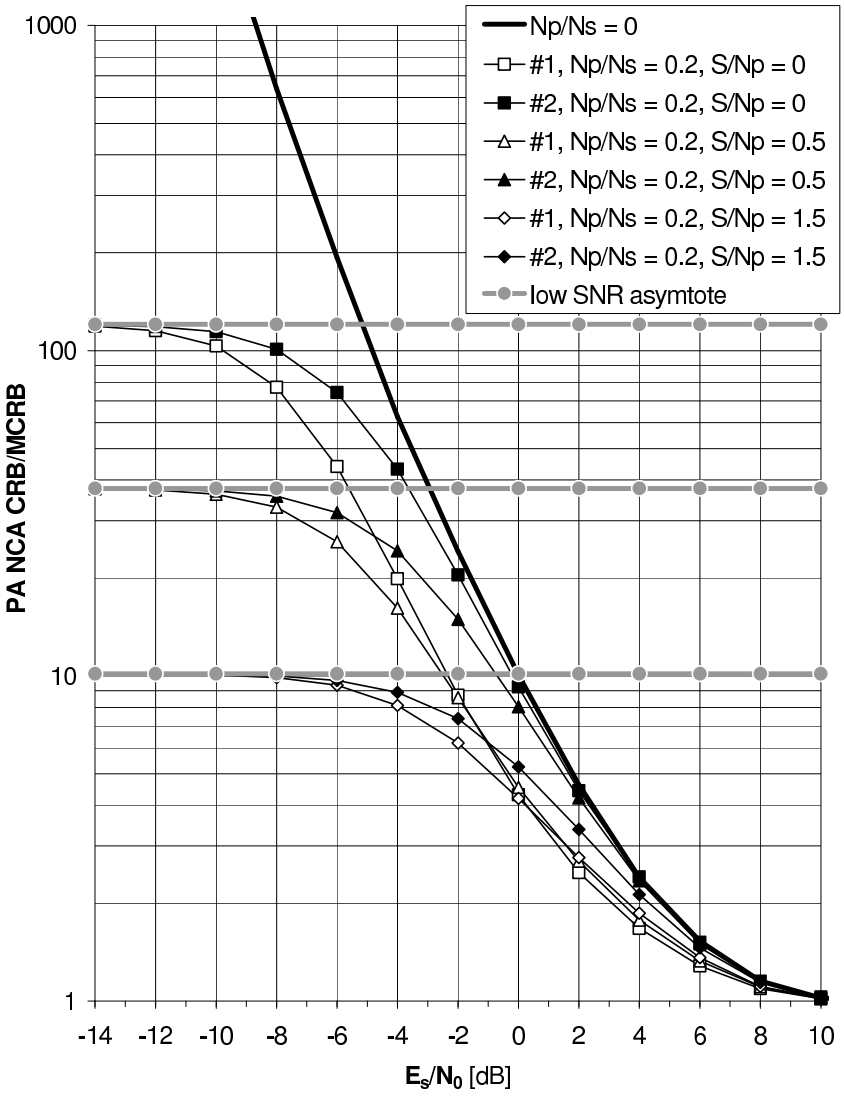

Figure 5.12: Ratio CRB/MCRB related to $L(\mathbf{\Psi};\mathbf{r})$, resulting from the simplified observation model and pertaining to the PA NCA estimate of the normalized frequency offset, assuming a 4-PSK signaling constellation and $\lambda_p = 0.2$. An asymmetric (#1) and a symmetric (#2) burst structure are considered, both of which are shown in Fig. 5.5.

error, showing the ratio of the PA NCA CRB_ν from (5.98) to the PA NCA $MCRB_\nu(N_s)$ from (5.30). As we restrict our attention to the simplified observation model, the quantity $\delta\tilde{J}_{\nu,\nu}$ in (5.98) reduces to zero. Using (5.98), (5.30) and (5.83)-(5.84), we can therefore write:

$$\frac{CRB_\nu^{(si)}}{MCRB_\nu(N_s)} = \frac{N_s^3}{12N_W\sigma_W^2}, \tag{5.119}$$

$$= \frac{N_s^3}{12\left(\sum_{k\in I_p}(k-\kappa_{m,W})^2 + R_\Omega\left(\frac{E_s}{N_0}\right)\sum_{k\in I_d}(k-\kappa_{m,W})^2\right)}. \tag{5.120}$$

It follows directly from the definition of σ_W^2 (see (5.83)) and N_W (see (5.81)) that (5.119) does not depend on the choice of the normalized time instant κ_0 at which the phase shift is estimated. In contrast with (5.116), the ratio CRB/MCRB (5.119)-(5.120) depends on the specific position of the pilot symbols in $\mathbf{a}$. Results are presented for 4-PSK, $N_s = 320$ transmitted symbols, a pilot symbol ratio of $\lambda_p = N_p/N_s = 0.1, 0.2$, and a spacing ratio of $\epsilon_{Sp} = S/N_p = 0.5, 1.5$. For reasons of comparison, the ratio CRB/MCRB (5.112) related to NPA NCA estimation (or, equivalently, $\lambda_p = 0$) is also displayed. We come to the following findings.

At very high SNR the CRB converges to its high SNR asymptote, such that (see section 5.7.2):

$$\frac{CRB_\nu^{(si)}}{MCRB_\nu(N_s)} \quad \rightarrow \quad 1.$$

At very low SNR, the CRB comes close to its low-SNR asymptote. It follows from the remark at the end of subsection 5.7.7 that, when $I_p \neq \emptyset$, this asymptote equals the MCRB pertaining to the DA estimation of ν. Consequently, for low values of E_s/N_0:

$$\frac{CRB_\nu^{(si)}}{MCRB_\nu(N_s)} \quad \rightarrow \quad \frac{ACRB_\nu^{(si)}}{MCRB_\nu(N_s)}, \tag{5.121}$$

$$= \quad \frac{MCRB_\nu(N_p)}{MCRB_\nu(N_s)}, \tag{5.122}$$

where $MCRB_\nu(N_p)$ denotes the MCRB pertaining to the DA estimation of ν from (5.33). As (5.33) nor (5.30) are affected by a time-shift of the pilot sequence within the symbol vector $\mathbf{a}$, both the asymmetric and symmetric burst structures yield the same low-SNR asymptote, for fixed λ_p and fixed ϵ_{Sp}. Evaluating (5.122) for the specific pilot symbol arrangement considered, we obtain:

$$\frac{ACRB_\nu^{(si)}}{MCRB_\nu(N_s)} \approx \left((\lambda_p)^3\left(1+3\epsilon_{Sp}\left(1+\epsilon_{Sp}\right)\right)\right)^{-1}, \tag{5.123}$$

where the approximation holds for large N_s. Note that (5.123) depends only on $(\lambda_p, \epsilon_{Sp})$. Consequently, for a fixed pilot symbol ratio λ_p, the ratio $CRB/MCRB$ from (5.119)-(5.120) is, at low SNR, mainly determined by the spacing ratio ϵ_{Sp}. At low and intermediate SNR, and for a given burst structure, increasing λ_p at fixed ϵ_{Sp}, or increasing ϵ_{Sp} at fixed λ_p decreases the ratio CRB/MCRB (5.119)-(5.120) (see Fig. 5.11 versus Fig. 5.12). For given values of λ_p and ϵ_{Sp}, the symmetric burst structure #2 yields a ratio CRB/MCRB that is larger than or equal to that of the asymmetric burst structure #1. This is consistent with (5.120). For a given constellation Ω and a given value of E_s/N_0, the denominator of (5.120) achieves its maximum value when $\kappa_{m,W} = \kappa_{m,p} = \kappa_{m,s}$, i.e., for a symmetric burst structure.

5.7.15.3 The Relation of the PA NCA CRB with the DA CRB and the NPA NCA CRB

It follows from (5.118), (5.122) and (5.70) that the low-SNR asymptote curves shown in Fig. 5.10, Fig. 5.11 and Fig. 5.12 can also be interpreted as pertaining to DA estimation. We also recall that the curves with $N_p/N_s = 0$ correspond to NPA NCA estimation. As a consequence, it may be concluded from Fig. 5.10, Fig. 5.11 and Fig. 5.12 that the PA NCA CRB is lower than both the DA and the NPA NCA CRB. This indicates that it is potentially more accurate to estimate ν and θ using a PA NCA algorithm that exploits both pilot symbols and data symbols (in an intelligent way) rather than using a DA algorithm that only exploits the pilot symbols (and ignores the data symbols) or using a NPA NCA algorithm that takes into account all received symbols (pilot symbols and data symbols) but ignores the a priori knowledge about the pilot symbols. The ratio (DA CRB)/(PA NCA CRB) ((NPA NCA CRB)/(PA NCA CRB)) depends on the operating SNR and on the transmission scheme, and indicates to what extent synchronizer performance can be improved by making clever use of the presence of the data symbols (of the knowledge about the pilot symbols) in the estimation process.

5.7.15.4 The Effect of the Burst Structure

For a fixed λ_p and a fixed ϵ_{Sp}, the burst structures #1 and #2 yield the same CRB for the phase shift estimation error at $t/T = \kappa_{m,W}$ (see (5.116) and Fig. 5.10), while the asymmetric burst structure #1 yields the smallest CRB for the normalized frequency offset estimation error (at any SNR) (see (5.119)-(5.120) and Figs. 5.11-5.12). However, as the following example illustrates, we must be very careful when interpreting these results. It was argued in section 5.1 that the CRB regarding the estimation of the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$ is given by:

$$CRB_{\Theta(kT;\mathbf{\Psi})} = CRB_{\theta}|_{\kappa_0=k}\,. \tag{5.124}$$

For the simplified observation model and the PA NCA estimation mode, we obtain from (5.99) and (5.124):

$$CRB_{\Theta(kT;\mathbf{\Psi})} = \frac{1}{2N_W\left(E_s/N_0\right)}\left(1+\frac{\left(k-\kappa_{m,W}\right)^2}{\sigma_W^2}\right). \tag{5.125}$$

The CRB from (5.125) depends on the relative location of k in the burst (i.e., the distance with respect to the center of gravity $\kappa_{m,W}$) rather than on the absolute value of k. Its *minimum* value is achieved at $k = \kappa_{m,W}$ and is equal to $\left(2N_W\left(E_s/N_0\right)\right)^{-1}$, which depends on the number of pilot symbols N_p and the number of data symbols N_d but not on the specific position of the pilot symbols in the symbol vector $\mathbf{a}$. The smallest values of $CRB_{\Theta(kT;\mathbf{\Psi})}$, $k \in I_d$ are achieved when the value of k is close to $\kappa_{m,W}$. The CRB (5.125) achieves its maximum value for $k = k_{max}$, i.e., at the edge of the interval I_d which is at maximum distance from $\kappa_{m,W}$ (or simultaneously at both edges if $\kappa_{m,W} = \kappa_{m,d}$, with $\kappa_{m,d}$ being the center of the data part of the symbol burst). Consequently, the detection of data symbols located near the edge k_{max} suffers from a larger instantaneous phase error variance when $\kappa_{m,W}$ is closer to one of the edges of the interval I_d, or equivalently, when $\kappa_{m,W}$ is further away from the center of the observation interval.

Fig. 5.13 depicts the CRB for the instantaneous phase shift error as a function of k. Both the symmetric and the asymmetric burst structure are considered, $N_s = 320$, $E_s/N_0 = 2\,\text{dB}$, $\lambda_p = 0.2$ and $\epsilon_{Sp} = 0.5$. We come to the following conclusions:

1. Although burst structure #1 yields the smallest CRB for the normalized frequency offset estimation error, the CRB on the instantaneous phase shift estimation error at $k = k_{max}$ is larger than for burst structure #2. This can be explained by noting that, at a value of E_s/N_0 as low as 2 dB, the distance $|k_{max} - \kappa_{m,W}|$ between the positions of the minimum and maximum value of the CRB on the phase shift estimation error is significantly larger for burst structure #1 than for burst structure #2.

2. Although burst structure #2 results in the smallest maximum for the CRB on the phase shift estimation error over the observation interval, other than for burst structure #1, this maximum value is reached at both edges of the observation interval. This implies that in the case of burst structure #2 more symbols are affected by a large mean square phase shift estimation error than in the case of burst structure #1.

Summarizing, it can be stated that the optimal burst structure strongly depends on the operating SNR and on the maximum allowable phase shift estimation error for proper symbol detection.

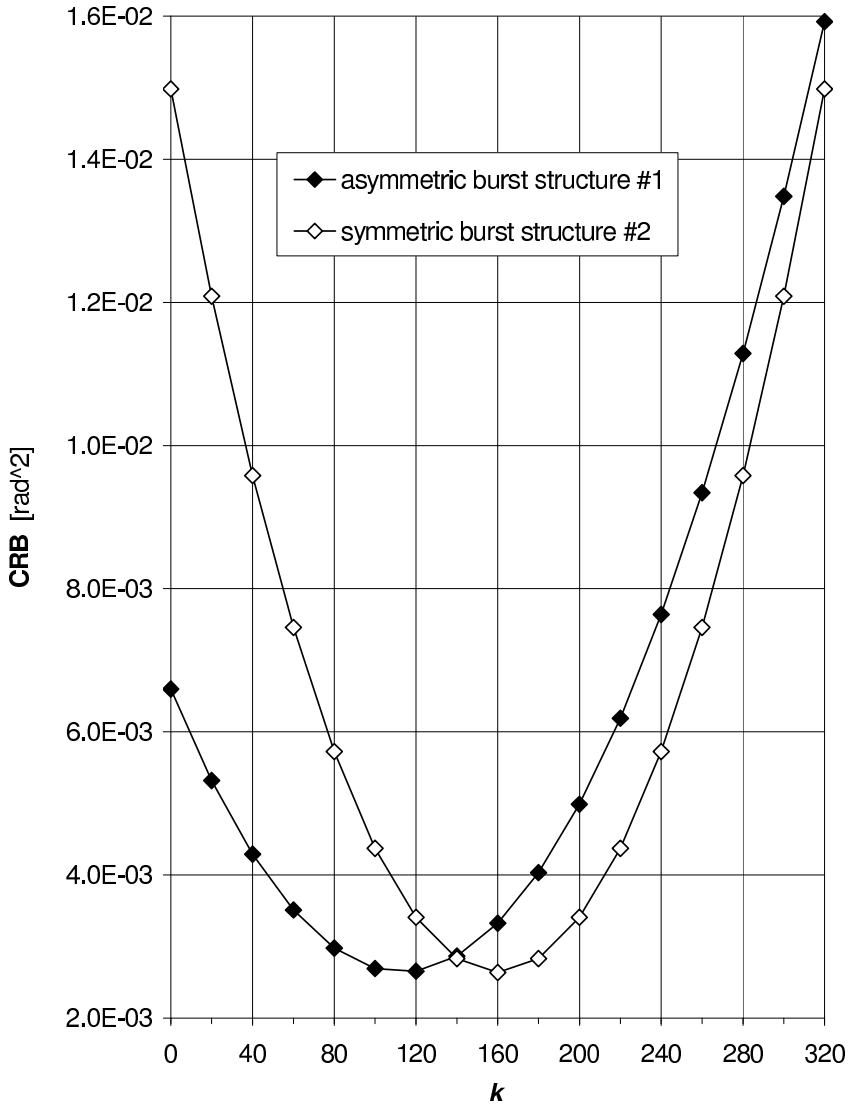

Figure 5.13: CRBs related to $L(\mathbf{\Psi};\mathbf{r})$ and pertaining to the PA NCA estimation of the instantaneous phase shift $\Theta(kT;\mathbf{\Psi})$, $k \in I_s$ with $I_s = [0, 320]$ and for $N_p = 64$, $\epsilon_{Sp} = 0.5$ and $E_s/N_0 = 2$ dB.

5.7.16 Numerical Evaluation and Discussion of the PA CA CRBs

Results pertaining to the PA CA CRBs are obtained for $I_p = \emptyset$, i.e., all symbols in $\mathbf{a}$ are data symbols, and $\kappa_0 = \kappa_{m,s}$ (see (5.24)), i.e., the phase shift is estimated at the center of the symbol burst. Using these conditions, we compute the (PA CA) FIM (5.47) for the channel codes (turbo (T), LDPC) and the symbol mappings (2-PSK, 4-PSK) that were considered in Fig. 4.9 of chapter 4. As far as this simulation set-up is concerned, our numerical results indicate that $(J_{\nu,\theta})^2 \ll J_{\nu,\nu} J_{\theta,\theta}$. This implies that the (PA CA) CRB (5.93) on the normalized frequency offset estimation error and the (PA CA) CRB (5.94) on the phase shift estimation error at $t/T = \kappa_{m,s}$ are essentially decoupled and well-approximated by $1/J_{\nu,\nu}$ and $1/J_{\theta,\theta}$, respectively. (A similar observation regarding the MFIM and the MCRBs was reported in section 5.6.) Fig. 5.14 shows the ratio CRB/MCRB related to PA CA phase shift estimation at $t/T = \kappa_{m,s}$ as a function of E_s/N_0 (curves with circular and triangular markers). For the values of E_s/N_0 and CRB/MCRB displayed in Fig. 5.14, the ratio CRB/MCRB related to PA CA normalized frequency offset estimation yields essentially the same curves, irrespective of the observation model. Fig. 5.14 also shows the result for uncoded transmission (curves with diamond markers): in this case, the ratios CRB/MCRB for PA CA estimation equal the ratios CRB/MCRB for PA NCA estimation (compare (5.13) with (5.16)), which in turn equal the ratios CRB/MCRB for NPA NCA estimation from (5.112) because we assume pilot symbol free transmission ($I_p = \emptyset$).

We make the following observations:

- For large E_s/N_0, the ratio CRB/MCRB converges to 1; this is consistent with [54].

- When E_s/N_0 decreases, a point is reached where the CRB starts to diverge from the MCRB. A comparison of Fig. 5.14 and Fig. 4.9 shows that this happens at a value of $E_s/N_0 = (E_s/N_0)_{thr}$ that corresponds to a SER of about 10^{-3}. In other words, the SNR threshold $(E_s/N_0)_{thr}$ for uncoded transmission exceeds the SNR threshold $(E_s/N_0)_{thr}$ for coded transmission by an amount equal to the coding gain at 10^{-3} SER. This indicates that, even at the (very low) operating SNR of the coded system, the CRB is very well approximated by the MCRB (which is much simpler to evaluate). It might be useful to note that, for small constellations like 2-PSK (2-PAM) and 4-PSK (4-QAM), SER≈BER (see also Fig. 4.9, with BER in dashed line). The BER is a more important code performance parameter than the SER and is therefore more often available in technical literature.

- The PA CA CRB for a powerful code, evaluated at its normal operating SNR (this excludes very low SNR at which the code becomes unreliable, as well as very high SNR at which uncoded transmission becomes reliable), is considerably smaller than the (N)PA NCA CRB. It follows that PA CA

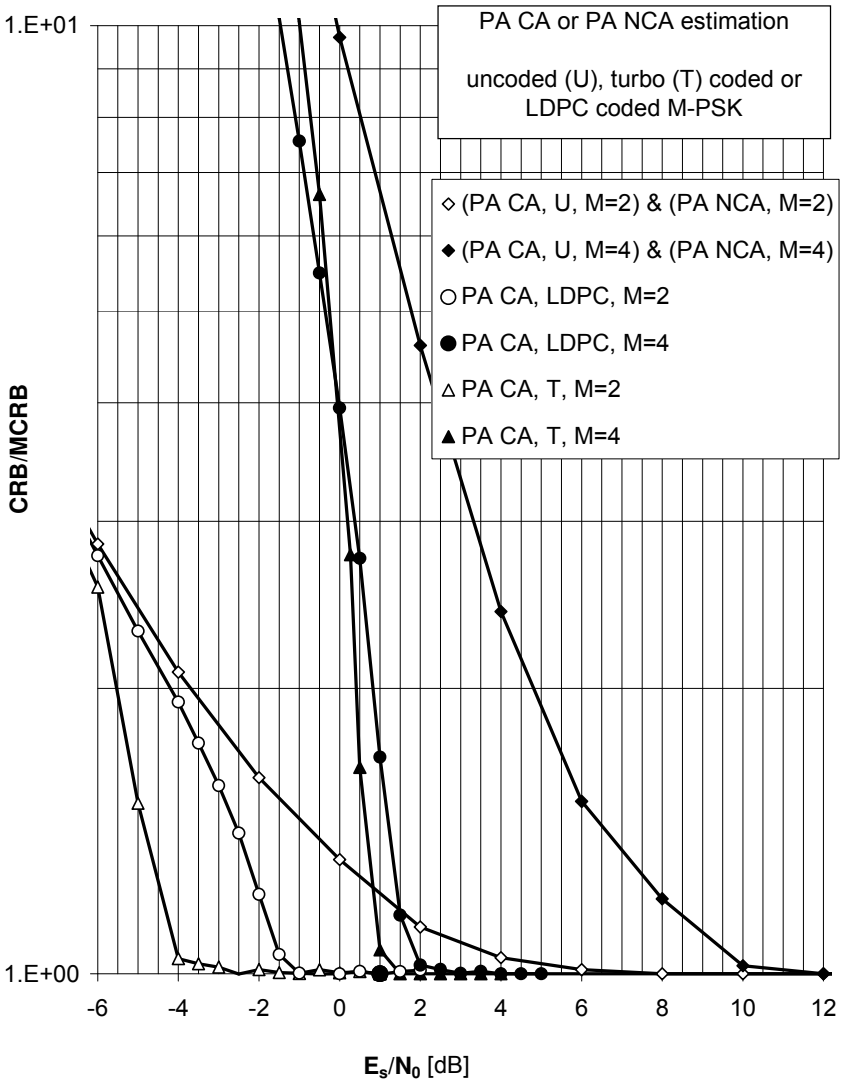

Figure 5.14: Ratio CRB/MCRB related to $L(\mathbf{\Psi};\mathbf{r})$ for the PA CA or PA NCA estimate of the normalized frequency offset or the phase shift at $t/T = \kappa_{m,s}$, for uncoded (U), turbo (T) coded or LDPC coded M-PSK ($M \in \{2,4\}$).

synchronizers are potentially more accurate than (N)PA NCA synchronizers when operating on coded signals. The ratio (PA NCA CRB)/(PA CA CRB) provides a quantitative indication of the extent to which synchronizer performance can be improved by making clever use of the code structure.

5.8 CRBs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$

This section computes the FIMs (and the associated CRBs) related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$. As for the FIM related to $L(\mathbf{\Psi};\mathbf{r})$, the analytical evaluation of these FIMs is not feasible and a brute-force evaluation of the FIMs by means of simulation is conceptually simple but the associated computational complexity increases exponentially with the number of symbols transmitted. Taking into account the linear modulation and assuming statistically independent symbols, we derive numerical procedures for evaluating the FIMs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$, respectively, with a complexity that is linear in the number of symbols transmitted. These procedures hold for the DA, the PA NCA and the NPA NCA estimation modes, but not for the PA CA estimation mode. The computation of the PA CA FIMs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$ seems practically unfeasible. The procedure for evaluating the DA, PA NCA or NPA NCA FIM related to $L(\nu;\mathbf{r})$ can be further simplified in the case of the correct observation model; whereas a similar reduction of the computational complexity is not possible for the FIM related to $L(\theta;\mathbf{r})$.

5.8.1 The FIMs

Taking $\mathbf{u} = \Psi_i$ and $\mathbf{v} = (\Psi_{3-i}\,\mathbf{a})^T$ in (3.26) yields the scalar Bayesian FIM J_i related to the marginal likelihood function $L(\Psi_i;\mathbf{r})$ of Ψ_i, for $i \in \{1,2\}$, with $\Psi_1 = \nu$ and $\Psi_2 = \theta$:

$$J_i \;=\; E_{\mathbf{r},\Psi_i}\left[\left(\ell'(\Psi_i;\mathbf{r})\right)^2\right], \tag{5.126}$$

where $E_{\mathbf{r},\Psi_i}[.]$ denotes averaging with respect to the joint pdf $p(\mathbf{r},\Psi_i)$ of $\mathbf{r}$ and Ψ_i, and $\ell'(\Psi_i;\mathbf{r})$ is a short-hand notation for the derivative of the log-likelihood function $\ln L(\Psi_i;\mathbf{r})$ with respect to Ψ_i, i.e.,

$$\ell'(\Psi_i;\mathbf{r}) = \frac{\partial \ln L(\Psi_i;\mathbf{r})}{\partial \Psi_i}. \tag{5.127}$$

According to (3.30), the derivative $\ell'(\Psi_i;\mathbf{r})$ (5.127) can be put into the following form :

$$\begin{aligned}&\ell'(\Psi_i;\mathbf{r}) \qquad\qquad (5.128)\\ &= \int \sum_{\tilde{\mathbf{a}}} \frac{\partial \ln p(\mathbf{r}\,|\mathbf{a}=\tilde{\mathbf{a}},\Psi_{3-i}=\psi,\Psi_i)}{\partial \Psi_i} p(\mathbf{a}=\tilde{\mathbf{a}},\Psi_{3-i}=\psi\,|\mathbf{r},\Psi_i)\,\mathrm{d}\psi.\end{aligned}$$

As the conditional pdf $p(\mathbf{r}\,|\mathbf{a},\Psi_{3-i},\Psi_i)$ is Gaussian (see (5.8)), the logarithm $\ln p(\mathbf{r}\,|\mathbf{a},\Psi_{3-i},\Psi_i)$ is readily available in closed-form:

$$\ln p(\mathbf{r}\,|\mathbf{a},\Psi_{3-i},\Psi_i) = -\frac{E_s}{N_0}\left|\mathbf{r}-\mathbf{s}(\mathbf{a},\mathbf{\Psi})\right|^2 \tag{5.129}$$

and the joint symbol a posteriori probabilities (APP) $p(\mathbf{a},\Psi_{3-i}\,|\mathbf{r},\Psi_i)$ can be computed from $p(\mathbf{r}\,|\mathbf{a},\Psi_{3-i},\Psi_i)$, $\Pr[\mathbf{a}]$ and $p(\Psi_{3-i})$, according to:

$$\begin{aligned} &p(\mathbf{a}=\tilde{\mathbf{a}},\Psi_{3-i}=\psi\,|\mathbf{r},\Psi_i) \qquad (5.130)\\ &= \frac{p(\mathbf{r}\,|\mathbf{a}=\tilde{\mathbf{a}},\Psi_{3-i}=\psi,\Psi_i)\Pr[\mathbf{a}=\tilde{\mathbf{a}}]\,p(\Psi_{3-i}=\psi)}{\int\sum_{\check{\mathbf{a}}} p\left(\mathbf{r}\,\middle|\mathbf{a}=\check{\mathbf{a}},\Psi_{3-i}=\tilde{\psi},\Psi_i\right)\Pr[\mathbf{a}=\check{\mathbf{a}}]\,p\left(\Psi_{3-i}=\tilde{\psi}\right)\mathrm{d}\tilde{\psi}}. \end{aligned}$$

5.8.2 High SNR Behavior of the FIMs

Assuming that $\mathbf{r}$ is given by (5.8), we now compute the high-SNR limit of the FIMs J_i (5.126), $i\in\{1,2\}$. Appendix L on page 229 shows that, at high E_s/N_0, one obtains:

$$\begin{aligned} &E_{\mathbf{r},\Psi_i}\left[\left(\ell'(\Psi_i;\mathbf{r})\right)^2\right]\\ &\approx E_{\mathbf{a},\Psi_i,\Psi_{3-i}}\left[E_{\mathbf{r}|\mathbf{a},\Psi_i,\Psi_{3-i}}\left[\left(\frac{\partial\ln p(\mathbf{r}\,|\mathbf{a},\Psi_i,\Psi_{3-i})}{\partial\Psi_i}\right)^2\right]\right],\\ &= E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}\left[\left(\frac{\partial\ln p(\mathbf{r}\,|\mathbf{a},\mathbf{\Psi})}{\partial\Psi_i}\right)^2\right]. \qquad (5.131) \end{aligned}$$

This reveals that (for both observation models and all estimation modes) the high-SNR limit of the FIMs $J_1=J_\nu$ and $J_2=J_\theta$ (5.126) are given by the corresponding MFIM entries M_ν (5.40) and M_θ (5.42). A similar reasoning has been used in section 5.7.2 regarding the high-SNR limit of the FIM related to $L(\mathbf{\Psi};\mathbf{r})$.

5.8.3 Computing the FIMs by 'Brute Force' Simulation

A brute force numerical evaluation of the scalar PA CA FIMs involves numerical integrations with respect to ψ and $\tilde{\psi}$ in (5.128) and (5.130), and the substitution in (5.126) of the statistical average $E[.]$ by an arithmetical average over a large number of realizations of $(\mathbf{r},\Psi_i)$. The latter are computer-generated according to the joint distribution of $\mathbf{r}$ and Ψ_i. This leads to the computation algorithm that is described in Algorithm 6. It should be noted that, due to the summations in (5.128) and (5.130), the computation of the derivatives $\ell_i(\mathbf{\Psi};\mathbf{r})$, $i=1,2$ in Step 1 gives rise to a computational complexity that is exponential in the burst size N_s.

Algorithm 6 FIMs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$ by brute force simulation

1. For $n = 1, 2, ..., N$:
   - generate $\mathbf{a}^{(n)}$ and $\mathbf{\Psi}^{(n)} = \left(\nu^{(n)}\theta^{(n)}\right)^T$ according to the distributions $\Pr[\mathbf{a}]$, $p(\nu)$ and $p(\theta)$;
   - generate $\mathbf{w}^{(n)}$ according to $\mathcal{N}_c\left(\mathbf{0}, \frac{N_0}{E_s}\mathbf{I}\right)$;
   - compute $\mathbf{r}^{(n)}$ as $\mathbf{r}^{(n)} = \mathbf{s}\left(\mathbf{a}^{(n)}, \mathbf{\Psi}^{(n)}\right) + \mathbf{w}^{(n)}$;
   - compute $\ell'\left(\Psi_i^{(n)}; \mathbf{r}^{(n)}\right)$ for $i \in \{1, 2\}$ according to (5.128).
2. Compute J_i for $i \in \{1, 2\}$ as:

$$J_i = \frac{1}{N}\sum_{n=1}^{N}\left(\ell'\left(\Psi_i^{(n)}; \mathbf{r}^{(n)}\right)\right)^2.$$

5.8.4 Computing the FIMs for Linear Modulation

In the case of linear modulation, the signal vector $\mathbf{s}(\mathbf{a}, \mathbf{\Psi})$ can be decomposed into the matrix product of the symbol vector $\mathbf{a}$, a scalar $e^{j\theta}$ and a matrix $\check{\mathbf{S}}(\nu)$ that depends only on ν:

$$\mathbf{s}(\mathbf{a}, \mathbf{\Psi}) = \mathbf{a}e^{j\theta}\check{\mathbf{S}}(\nu). \tag{5.132}$$

For the correct observation model using (5.10), the (k,l)th element $\check{S}((k,l);\nu)$ of $\check{\mathbf{S}}(\nu)$ is given by

$$\check{S}((k,l);\nu) = \sqrt{T_{\text{sample}}}p(lT_{\text{sample}} - kT)\, e^{j2\pi\nu\left(l\frac{T_{\text{sample}}}{T} - \kappa_0\right)}, \tag{5.133}$$

whereas, for the simplified observation model using (5.12), we obtain:

$$\check{S}((k,l);\nu) = e^{j2\pi\nu(k-\kappa_0)}\delta_{k-l}. \tag{5.134}$$

For both observation models, the following holds:

$$\check{\mathbf{S}}(\nu)\,\check{\mathbf{S}}^H(\nu) = \mathbf{I}. \tag{5.135}$$

It follows from (5.135) that (within a factor not depending on $(\mathbf{a}, \mathbf{\Psi})$) the logarithm $\ln p(\mathbf{r}|\mathbf{a}, \mathbf{\Psi})$ (5.129) can be rewritten as follows:

$$\ln p(\mathbf{r}|\mathbf{a}, \mathbf{\Psi}) \propto -\frac{E_s}{N_0}\left(|\mathbf{a}|^2 - 2\Re\left\{e^{-j\theta}\check{\mathbf{z}}(\mathbf{r}, \nu)\,\mathbf{a}^H\right\}\right), \tag{5.136}$$

with

$$\check{\mathbf{z}}(\mathbf{r}, \nu) = \mathbf{r}\check{\mathbf{S}}^H(\nu). \tag{5.137}$$

Differentiating with respect to Ψ_i yields:

$$\frac{\partial \ln p(\mathbf{r}\,|\mathbf{a},\boldsymbol{\Psi})}{\partial \nu} = \frac{2E_s}{N_0}\Re\left\{e^{-j\theta}\check{\mathbf{z}}_\nu(\mathbf{r},\nu)\,\mathbf{a}^H\right\}, \tag{5.138}$$

$$\frac{\partial \ln p(\mathbf{r}\,|\mathbf{a},\boldsymbol{\Psi})}{\partial \theta} = \frac{2E_s}{N_0}\Im\left\{e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\nu)\,\mathbf{a}^H\right\}, \tag{5.139}$$

where $\check{\mathbf{z}}_\nu(\mathbf{r},\nu)$ is a short-hand notation for the derivative of $\check{\mathbf{z}}(\mathbf{r},\nu)$ with respect to ν , i.e.,

$$\check{\mathbf{z}}_\nu(\mathbf{r},\nu) = \frac{\partial \check{\mathbf{z}}(\mathbf{r},\nu)}{\partial \nu}. \tag{5.140}$$

Using (5.138)-(5.139) one obtains from (5.128):

$$\ell'(\nu;\mathbf{r}) \tag{5.141}$$

$$= \int \sum_{\tilde{\mathbf{a}}} \frac{2E_s}{N_0}\Re\left\{e^{-j\tilde{\theta}}\check{\mathbf{z}}_\nu(\mathbf{r},\nu)\,\tilde{\mathbf{a}}^H\right\} p\left(\mathbf{a}=\tilde{\mathbf{a}},\theta=\tilde{\theta}\,|\mathbf{r},\nu\right)\mathrm{d}\tilde{\theta}, \tag{5.142}$$

$$= \frac{2E_s}{N_0}\sum_{k\in I_s} X_1(k;\mathbf{r},\nu), \tag{5.143}$$

$$\ell'(\theta;\mathbf{r}) \tag{5.144}$$

$$= \int \sum_{\tilde{\mathbf{a}}} \frac{2E_s}{N_0}\Im\left\{e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\tilde{\nu})\,\tilde{\mathbf{a}}^H\right\} p(\mathbf{a}=\tilde{\mathbf{a}},\nu=\tilde{\nu}\,|\mathbf{r},\theta)\,\mathrm{d}\tilde{\nu}, \tag{5.145}$$

$$= \frac{2E_s}{N_0}\sum_{k\in I_s} X_2(k;\mathbf{r},\theta), \tag{5.146}$$

with

$$X_1(k;\mathbf{r},\nu) = \Re\left\{\check{z}_\nu(k;\mathbf{r},\nu)\int\sum_{m=0}^{M-1} e^{-j\tilde{\theta}}\omega_m^* p\left(a(k)=\omega_m,\theta=\tilde{\theta}\,|\mathbf{r},\nu\right)\mathrm{d}\tilde{\theta}\right\}, \tag{5.147}$$

$$X_2(k;\mathbf{r},\theta) = \Im\left\{e^{-j\theta}\int\sum_{m=0}^{M-1}\check{z}(k;\mathbf{r},\tilde{\nu})\,\omega_m^* p(a(k)=\omega_m,\nu=\tilde{\nu}\,|\mathbf{r},\theta)\,\mathrm{d}\tilde{\nu}\right\}, \tag{5.148}$$

It is worth noting that no approximation is involved in obtaining (5.143) and (5.146).

No further analytical simplification of $\ell'(\nu;\mathbf{r})$ and $\ell'(\theta;\mathbf{r})$ seems possible. Since ν and θ are defined over a *continuous* domain, the framework of the sum-product algorithm and factor graphs is much less attractive for the computation of the APPs $p(a(k)=\omega,\theta\,|\mathbf{r},\nu)$ and $p(a(k)=\omega,\nu\,|\mathbf{r},\theta)$ involved in $\ell'(\nu;\mathbf{r})$ (see (5.147)) and $\ell'(\theta;\mathbf{r})$ (see (5.148)) than for the APPs $\Pr[a(k)=\omega\,|\mathbf{r},\boldsymbol{\Psi}]$ involved

in $\ell_i(\mathbf{\Psi};\mathbf{r})$, $i \in \{1,2\}$ (see section 5.7.4). This is because the "summations" in the sum-product message-passing algorithm have become integrals and because the messages might also be probability density functions now, rather than probability mass functions only. Of course, the APPs $p(a(k), \Psi_{3-i}|\mathbf{r}, \Psi_i)$, $i = 1, 2$ can always be computed as marginals of the joint APPs $p(\mathbf{a}, \Psi_{3-i}|\mathbf{r}, \Psi_i)$ (5.130), but then the computational complexity will still increase exponentially with N_s. This will generally imply that the complexity associated with the computation of the FIMs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$ cannot be drastically reduced when $\mathbf{s}(\mathbf{a},\mathbf{\Psi})$ satisfies (5.132). This is in contrast to the FIM related to $L(\mathbf{\Psi};\mathbf{r})$ (see section 5.7.4).

5.8.5 Computing the PA NCA FIMs for Linear Modulation

We will now compute the PA NCA FIMs. It follows from (5.13) and (5.16) that the PA NCA FIMs are a special case of the PA CA FIMs, namely for the case of uncoded transmission. Equations (5.126)-(5.148) therefore remain valid for the PA NCA scenario. However, we show in the following that the computational complexity of the APPs $p(a(k) = \omega, \theta = \psi|\mathbf{r}, \nu)$ and $p(a(k) = \omega, \nu = \psi|\mathbf{r}, \theta)$ can be significantly reduced when the a priori probability mass function $\Pr[\mathbf{a}]$ satisfies the PA NCA assumption (5.16).

Substituting (5.16) and (5.136) into (5.130) and summing the result over all symbols $a(l)$, $l \neq k$, $l \in I_s$ yields:

$$p(a(k) = \omega, \theta|\mathbf{r}, \nu) = \begin{cases} Y_{1,s}(\mathbf{r}, \theta, \nu)\, I[\omega = A_k] & , k \in I_p \\ Y_{1,s}(\mathbf{r}, \theta, \nu)\, G\left(\omega, e^{-j\theta}\check{z}(k;\mathbf{r}, \nu)\right) & , k \in I_d \end{cases} \quad (5.149)$$

and

$$p(a(k) = \omega, \nu|\mathbf{r}, \theta) = \begin{cases} Y_{2,s}(\mathbf{r}, \theta, \nu)\, I[\omega = A_k] & , k \in I_p \\ Y_{2,s}(\mathbf{r}, \theta, \nu)\, G\left(\omega, e^{-j\theta}\check{z}(k;\mathbf{r}, \nu)\right) & , k \in I_d \end{cases}, \quad (5.150)$$

where A_k denotes the actual value of the pilot symbol $a(k)$, $k \in I_p$, and $I[.]$ is the indicator function defined in (5.14), i.e., $I[P] = 0$ when P is false and $I[P] = 1$ when P is true. The quantities $G(.,.)$, $Y_{1,s}(.,.)$ and $Y_{2,s}(.,.)$ are given by:

$$G\left(\omega, e^{-j\theta}\check{z}(k;\mathbf{r}, \nu)\right) = \frac{F\left(\omega, e^{-j\theta}\check{z}(k;\mathbf{r}, \nu)\right)}{\sum_{m=0}^{M-1} F\left(\omega_m, e^{-j\theta}\check{z}(k;\mathbf{r}, \nu)\right)}, \quad (5.151)$$

$$Y_{1,s}(\mathbf{r}, \theta, \nu) = \frac{U_s\left(e^{-j\theta}\check{\mathbf{z}}(\mathbf{r}, \nu)\right)}{\int_{-\pi}^{\pi} U_s\left(e^{-j\tilde{\theta}}\check{\mathbf{z}}(\mathbf{r}, \nu)\right) \mathrm{d}\tilde{\theta}}, \quad (5.152)$$

$$Y_{2,s}(\mathbf{r}, \theta, \nu) = \frac{U_s\left(e^{-j\theta}\check{\mathbf{z}}(\mathbf{r}, \nu)\right)}{\int_{-\nu_{max}}^{\nu_{max}} U_s\left(e^{-j\theta}\check{\mathbf{z}}(\mathbf{r}, \tilde{\nu})\right) \mathrm{d}\tilde{\nu}}, \quad (5.153)$$

Algorithm 7 PA NCA FIMs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$ for linear modulation

1. For $n = 1, 2, ..., N$:
   - generate $\mathbf{a}^{(n)}$, $\nu^{(n)}$ and $\theta^{(n)}$ according to the distributions $\Pr[\mathbf{a}]$, $p(\nu)$ and $p(\theta)$;
   - generate $\mathbf{w}^{(n)}$ according to $\mathcal{N}_c\left(\mathbf{0}, \frac{N_0}{E_s}\mathbf{I}\right)$;
   - compute $\mathbf{r}^{(n)} = \mathbf{a}^{(n)} e^{j\theta^{(n)}} \check{\mathbf{S}}\left(\nu^{(n)}\right) + \mathbf{w}^{(n)}$;
   - if $i = 1$ (normalized frequency offset estimation):
     - compute $\check{\mathbf{z}}\left(\mathbf{r}^{(n)}, \nu^{(n)}\right) = \mathbf{r}^{(n)} \check{\mathbf{S}}^H\left(\nu^{(n)}\right)$;
     - compute $p\left(a(k) = \omega, \theta \left| \mathbf{r}^{(n)}, \nu^{(n)}\right.\right)$ for $\omega \in \Omega$, $\theta \in [-\pi, \pi]$ and $k \in I_s$, according to (5.149);
     - compute $\check{\mathbf{z}}_\nu\left(\mathbf{r}^{(n)}, \nu^{(n)}\right) = \mathbf{r}^{(n)} \left.\frac{\partial \check{\mathbf{S}}^H(\nu)}{\partial \nu}\right|_{\nu=\nu^{(n)}}$;
     - compute $\ell'\left(\nu^{(n)}; \mathbf{r}^{(n)}\right)$ according to (5.143),(5.147).
   - if $i = 2$ (phase shift estimation):
     - compute $\check{\mathbf{z}}\left(\mathbf{r}^{(n)}, \nu\right) = \mathbf{r}^{(n)} \check{\mathbf{S}}^H(\nu)$ for $\nu \in [-\nu_{max}, \nu_{max}]$;
     - compute $p\left(a(k) = \omega, \nu \left| \mathbf{r}^{(n)}, \theta^{(n)}\right.\right)$ for $\omega \in \Omega$, $\nu \in [-\nu_{max}, \nu_{max}]$ and $k \in I_s$, according to (5.150);
     - compute $\ell'\left(\theta^{(n)}; \mathbf{r}^{(n)}\right)$ according to (5.146),(5.148).
2. Compute J_i, $i \in \{1, 2\}$ as:

$$J_i = \frac{1}{N} \sum_{n=1}^{N} \left(\ell'\left(\Psi_i^{(n)}; \mathbf{r}^{(n)}\right)\right)^2.$$

with

$$U_s\left(e^{-j\theta}\check{\mathbf{z}}(\mathbf{r}, \nu)\right) = \prod_{k \in I_p} F\left(A_k, e^{-j\theta}\check{z}(k; \mathbf{r}, \nu)\right) \prod_{k \in I_d} \sum_{m=0}^{M-1} F\left(\omega_m, e^{-j\theta}\check{z}(k; \mathbf{r}, \nu)\right) \tag{5.154}$$

and

$$F\left(\omega, e^{-j\theta}\check{z}(k; \mathbf{r}, \nu)\right) = e^{\frac{E_s}{N_0}\left(2\Re\left\{e^{-j\theta}\check{z}(k;\mathbf{r},\nu)\omega^*\right\} - |\omega|^2\right)}.$$

Using (5.149)-(5.154) and (5.141)-(5.148), the PA NCA FIMs $J_i = 1/CRB_i$ for $i \in \{1, 2\}$ can be computed as in Algorithm 7. The complexity of this procedure is linear (and not exponential) in the number of symbols N_s.

5.8.6 Computing the DA FIMs for Linear Modulation

The DA FIMs are also a special case of the PA CA FIMs, namely for all data symbols equal to zero (compare (5.13) and (5.15)). In the following, the computation of the DA FIM is found to be similar to that of the PA NCA FIMs, although there are some subtle differences.

Let us consider the computation of the APPs $p\left(a\left(k\right)=\omega,\theta=\psi\left|\mathbf{r},\nu\right.\right)$ and $p\left(a\left(k\right)=\omega,\nu=\psi\left|\mathbf{r},\theta\right.\right)$ when the a priori probability mass function $\Pr\left[\mathbf{a}\right]$ satisfies (5.15). Substituting (5.15) and (5.136) into (5.130) and summing the result over all symbols $a\left(l\right)$, $l\neq k$, $l\in I_s$ yields:

$$p\left(a\left(k\right)=\omega,\theta\left|\mathbf{r},\nu\right.\right)=\left\{\begin{array}{ll} Y_{1,p}\left(\mathbf{r},\nu,\theta\right)I\left[\omega=A_k\right] & ,\,k\in I_p \\ Y_{1,p}\left(\mathbf{r},\nu,\theta\right)I\left[\omega=0\right] & ,\,k\in I_d \end{array}\right. \tag{5.155}$$

and

$$p\left(a\left(k\right)=\omega,\nu\left|\mathbf{r},\theta\right.\right)=\left\{\begin{array}{ll} Y_{2,p}\left(\mathbf{r},\nu,\theta\right)I\left[\omega=A_k\right] & ,\,k\in I_p \\ Y_{2,p}\left(\mathbf{r},\nu,\theta\right)I\left[\omega=0\right] & ,\,k\in I_d \end{array}\right., \tag{5.156}$$

where A_k denotes the actual value of the pilot symbol $a\left(k\right)$, $k\in I_p$, and $I\left[.\right]$ is the indicator function defined in (5.14), i.e., $I\left[P\right]=0$ when P is false and $I\left[P\right]=1$ when P is true. The quantities $Y_{1,p}\left(.,.\right)$ and $Y_{2,p}\left(.,.\right)$ are given by:

$$Y_{1,p}\left(\mathbf{r},\nu,\theta\right)=\frac{U_p\left(e^{-j\theta}\check{Z}_a\left(\mathbf{z}\left(\mathbf{r},\nu\right)\right)\right)}{\int_{-\pi}^{\pi}U_p\left(e^{-j\tilde{\theta}}\check{Z}_a\left(\mathbf{z}\left(\mathbf{r},\nu\right)\right)\right)\mathrm{d}\tilde{\theta}}, \tag{5.157}$$

$$Y_{2,p}\left(\mathbf{r},\nu,\theta\right)=\frac{U_p\left(e^{-j\theta}\check{Z}_a\left(\mathbf{z}\left(\mathbf{r},\nu\right)\right)\right)}{\int_{-\nu_{max}}^{\nu_{max}}U_p\left(e^{-j\theta}\check{Z}_a\left(\mathbf{z}\left(\mathbf{r},\nu\right)\right)\right)\mathrm{d}\tilde{\nu}}, \tag{5.158}$$

with

$$U_p\left(e^{-j\theta}\check{Z}_a\left(\mathbf{z}\right)\right)=e^{N_p\frac{E_s}{N_0}2\Re\left\{e^{-j\theta}\check{Z}_a\left(\mathbf{z}\right)\right\}}, \tag{5.159}$$

with

$$\check{Z}_a\left(\mathbf{z}\left(\mathbf{r},\nu\right)\right)=\frac{1}{N_p}\sum_{k\in I_p}\check{z}\left(k;\mathbf{r},\nu\right)A_k^*.$$

The DA FIM can thus be computed as in Algorithm 7, with in Step 1 the computation of the APPs according to (5.155) for $i=1$ and (5.156) for $i=2$. Note from (5.157)-(5.159) that the computation of the APPs $p\left(a\left(k\right)=\omega,\theta\left|\mathbf{r},\nu\right.\right)$ and $p\left(a\left(k\right)=\omega,\nu\left|\mathbf{r},\theta\right.\right)$ only involves samples $\check{z}\left(k;\mathbf{r},\nu\right)$ with $k\in I_p$. Also, note from (5.155)-(5.156) that (5.147) and (5.148) vanish for $k\in I_d$, in such a way that also the computation of the log-likelihood functions $\ell'\left(\nu;\mathbf{r}\right)$ and $\ell'\left(\theta;\mathbf{r}\right)$ only involves samples $\check{z}\left(k;\mathbf{r},\nu\right)$ and $\check{z}_\nu\left(k;\mathbf{r},\nu\right)$ with $k\in I_p$. These two facts imply that the complexity of this procedure is linear in the number of pilot symbols N_p (rather than in the number of transmitted symbols N_s, as was the case for the PA NCA FIMs).

Remark

In contrast with the DA FIM related to $L(\mathbf{\Psi};\mathbf{r})$, the DA FIMs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$ do not equal the corresponding DA MFIMs. This is mainly due to the fact that (5.126) involves averaging with respect to either ν or θ, whereas (5.40), (5.42), (5.19) and (5.48) all involve averaging with respect to both ν and θ.

5.8.7 Computing the PA NCA and the DA FIM Related to $L(\nu;\mathbf{r})$: Direct Generation of $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$

In case of the correct observation model, the computation of $\mathbf{s}\left(\mathbf{a}^{(n)},\mathbf{\Psi}^{(n)}\right)$ involves the convolution of the symbol vector $\mathbf{a}^{(n)}$ with the samples of the transmit filter impulse response $p(t)$, while the computation of $\check{\mathbf{z}}(\mathbf{r},\nu)$ and $\check{\mathbf{z}}_{\nu}(\mathbf{r},\nu)$ involves the convolution of the observation vector $\mathbf{r}^{(n)}$ with the samples of the matched filter impulse response (which also equals $p(t)$). We will now show that, for the PA NCA and the DA FIM related to $L(\nu;\mathbf{r})$, this complexity can be reduced by directly generating realizations of $\mathbf{z}(\mathbf{r},\mathbf{\Psi}) = e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\nu)$. This is easy, since when $(\mathbf{a},\nu,\theta)$ is given, $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$ can be decomposed as:

$$\begin{aligned}\mathbf{z}(\mathbf{r},\mathbf{\Psi}) &= e^{-j\theta}\left(\mathbf{a}e^{j\theta}\check{\mathbf{S}}(\nu)+\mathbf{w}\right)\check{\mathbf{S}}^{H}(\nu),\\ &= \mathbf{a}+\mathbf{n},\end{aligned}$$

where $\mathbf{n} = e^{-j\theta}\mathbf{w}\check{\mathbf{S}}^{H}(\nu)$ is a complex-valued AWGN vector with zero mean and covariance matrix $E\left[\mathbf{n}^{H}\mathbf{n}\right]$ given by:

$$\begin{aligned}E\left[\mathbf{n}^{H}\mathbf{n}\right] &= e^{j\theta}\check{\mathbf{S}}(\nu)E\left[\mathbf{w}^{H}\mathbf{w}\right]\check{\mathbf{S}}^{H}(\nu)e^{-j\theta},\\ &= \frac{N_0}{E_s}\mathbf{I}.\end{aligned}$$

Note that, for given $(\mathbf{a},\mathbf{\Psi})$, the statistics of $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$ do not depend on $\mathbf{\Psi}$.

The integrands involved in the computation of (5.147), (5.152) and (5.157) are periodic in the integration variable $\tilde{\theta}$ with a period equal to 2π, so there is no need to adjust the integration interval ($[-\pi,\pi]$) when performing the substitution $\tilde{\theta} = \tilde{\theta}' + \theta$, where θ denotes the actual phase shift and $\tilde{\theta}'$ is the new integration variable. As a result, the derivative $\ell'(\nu;\mathbf{r})$ (5.143) can be expressed as:

$$\ell'(\nu;\mathbf{r}) = \frac{2E_s}{N_0}\sum_{k\in I}\Re\left\{K\left(k;\mathbf{z}(\mathbf{r},\mathbf{\Psi})\right)z_{\nu}(k;\mathbf{r},\mathbf{\Psi})\right\}, \tag{5.160}$$

where $(I,K(k;\mathbf{z})) = (I_s,K_s(k;\mathbf{z}))$ for the PA NCA scenario and $(I,K(k;\mathbf{z})) = (I_p,K_p(k;\mathbf{z}))$ for the DA scenario, with:

$$K_s(k;\mathbf{z}) = \begin{cases} A_k^{*}D_s(\mathbf{z}) & ,k\in I_p\\ V(k;\mathbf{z}) & ,k\in I_d\end{cases}, \tag{5.161}$$

$$D_s(\mathbf{z}) = \frac{\int_{-\pi}^{\pi} e^{-j\tilde{\theta}'} U_s\left(e^{-j\tilde{\theta}'}\mathbf{z}\right) \mathrm{d}\tilde{\theta}'}{\int_{-\pi}^{\pi} U_s\left(e^{-j\tilde{\theta}'}\mathbf{z}\right) \mathrm{d}\tilde{\theta}'},$$

$$V(k;\mathbf{z}) = \frac{\int_{-\pi}^{\pi} \sum_{m=0}^{M-1} G\left(\omega_m, e^{-j\tilde{\theta}'} z(k)\right) \omega_m^* e^{-j\tilde{\theta}'} U_s\left(e^{-j\tilde{\theta}'}\mathbf{z}\right) \mathrm{d}\tilde{\theta}'}{\int_{-\pi}^{\pi} U_s\left(e^{-j\tilde{\theta}'}\mathbf{z}\right) \mathrm{d}\tilde{\theta}'}$$

and

$$K_p(k;\mathbf{z}) = A_k^* D_p(Z_a(\mathbf{z})), \tag{5.162}$$

$$D_p(Z_a(\mathbf{z})) = \frac{\int_{-\pi}^{\pi} e^{-j\tilde{\theta}'} U_p\left(e^{-j\tilde{\theta}'} Z_a(\mathbf{z})\right) \mathrm{d}\tilde{\theta}'}{\int_{-\pi}^{\pi} U_p\left(e^{-j\tilde{\theta}'} Z_a(\mathbf{z})\right) \mathrm{d}\tilde{\theta}'}, \tag{5.163}$$

where

$$Z_a(\mathbf{z}) = e^{-j\theta}\check{Z}_a(\mathbf{z}) = \frac{1}{N_p}\sum_{k\in I_p} A_k^* z(k). \tag{5.164}$$

It follows from (5.160) that, for the PA NCA and the DA scenarios, $\ell'(\nu;\mathbf{r})$ only depends on $(\mathbf{r},\nu)$ through $\mathbf{z}(\mathbf{r},\mathbf{\Psi}) = e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\nu)$ and $z_\nu(k;\mathbf{r},\mathbf{\Psi}) = e^{-j\theta}\check{z}_\nu(k;\mathbf{r},\nu)$, such that the corresponding FIM $J_1 = J_\nu$ (5.126) can be rewritten as follows:

$$J_\nu = \left(\frac{2E_s}{N_0}\right)^2 E_{\mathbf{z}}\left[\sum_{k,l\in I} H_\nu(k,l;\mathbf{z})\right], \tag{5.165}$$

with $\mathbf{z}$ a short-hand notation of $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$, and

$$H_\nu(k,l;\mathbf{z}) = E_{z_\nu(k),z_\nu(l)|\mathbf{z}}\left[\Re\{K(k;\mathbf{z})\, z_\nu(k)\}\,\Re\{K(l;\mathbf{z})\, z_\nu(l)\}\right], \tag{5.166}$$

with $z_\nu(k)$ a short-hand notation of $z_\nu(k;\mathbf{r},\mathbf{\Psi})$. Taking into account the statistical properties of $\mathbf{z}$, $\frac{\partial \mathbf{z}}{\partial \nu}$ and $\frac{\partial \mathbf{z}}{\partial \theta}$ (see Appendix F on page 207), we are able to perform analytically the average in (5.166) over $z_\nu(k)$ and $z_\nu(l)$, conditioned on $\mathbf{z}$, based on $E[z_\nu(k)\, z_\nu(l)\,|\mathbf{z}]$, $E[z_\nu^*(k)\, z_\nu(l)\,|\mathbf{z}]$ (see Appendix M on page 233). One obtains:

$$J_\nu = \left(\frac{4\pi E_s}{N_0}\right)^2 \left(\frac{N_0}{2E_s} E_2 + E_1\right), \tag{5.167}$$

with

$$E_1 = E_{\mathbf{z}}\left[\Im\left\{\sum_{k\in I}(k-\kappa_0)\, K(k;\mathbf{z})\, z(k)\right\}^2\right] \tag{5.168}$$

and $E_2 = 0$ for the simplified observation model and E_2 given by:

$$E_2 = E_{\mathbf{z}}\left[\Re\left\{\sum_{k,l\in I} \mathrm{f}((k-l)T)\, K(k;\mathbf{z})\, K^*(l;\mathbf{z})\right\}\right], \tag{5.169}$$

for the correct observation model, where $\mathrm{f}(t) = \int \left(\frac{\tilde{t}}{T}\right)^2 p\left(\tilde{t}-\frac{t}{2}\right) p\left(\tilde{t}+\frac{t}{2}\right) \mathrm{d}\tilde{t}$. This leads to the algorithm described in Algorithm 8. The resulting numerical complexity is considerably reduced as compared to Algorithm 7, because the convolution of the vectors $\mathbf{a}^{(n)}$ and $\mathbf{r}^{(n)}$ with the samples of the filter impulse response $p(t)$ is avoided in Step 1.

5.8.7.1 Remark

We will now show that a similar reduction of the computational complexity is not possible for the PA NCA FIM related to $L(\theta;\mathbf{r})$.

It follows from (5.146) that the evaluation of the FIM $J_2 = J_\theta$ (5.126) inevitably requires the generation of realizations of $e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\psi)$ for all values ψ in the normalized frequency offset domain $[-\nu_{max},\nu_{max}]$. Let ν denote the actual value of the normalized frequency offset. For given $(\mathbf{a},\nu,\theta)$, $e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\psi)$ can then be decomposed as follows:

$$\begin{aligned} e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\psi) &= e^{-j\theta}\left(\mathbf{a}e^{j\theta}\check{\mathbf{S}}(\nu)+\mathbf{w}\right)\check{\mathbf{S}}^H(\psi), \\ &= \mathbf{m}+\mathbf{n}, \end{aligned}$$

where

$$\mathbf{m}(\mathbf{a},\psi-\nu) = \mathbf{a}\check{\mathbf{S}}(\nu)\check{\mathbf{S}}^H(\psi)$$

and $\mathbf{n} = e^{-j\theta}\mathbf{w}\check{\mathbf{S}}^H(\psi)$ is a complex-valued AWGN vector with zero mean and covariance matrix $E\left[\mathbf{n}^H\mathbf{n}\right]$ given by:

$$\begin{aligned} E\left[\mathbf{n}^H\mathbf{n}\right] &= e^{j\theta}\check{\mathbf{S}}(\psi)E\left[\mathbf{w}^H\mathbf{w}\right]\check{\mathbf{S}}^H(\psi)e^{-j\theta}, \\ &= \frac{N_0}{E_s}\mathbf{I}. \end{aligned}$$

Note that, for given $(\mathbf{a},\nu,\theta)$, the statistics of $e^{-j\theta}\check{\mathbf{z}}(\mathbf{r},\psi)$ depend on $(\psi-\nu)$. For the correct observation model and $(\psi-\nu)\neq 0$, the computation of $\mathbf{m}\left(\mathbf{a}^{(n)},\mathbf{\Psi}^{(n)}\right)$ involves a two-fold convolution of the symbol vector $\mathbf{a}^{(n)}$ with the samples of the transmit filter impulse response $p(t)$, which indicates that the direct generation of realizations of $e^{-j\theta}\check{z}(k;\mathbf{r},\psi)$ does not in any way reduce the computational complexity of the FIM related to $L(\theta;\mathbf{r})$.

5.8.8 Computing the DA FIM Related to $L(\nu;\mathbf{r})$: Direct Generation of $Z_a(\mathbf{z})$ and $Z_b(\mathbf{z})$

The function $K(k;\mathbf{z}) = K_p(k;\mathbf{z})$ (5.162) involved in the computation of the DA FIM J_ν (5.167) via the expectations E_1 (5.168) and E_2 (5.169) decomposes into the conjugate of the actual value of the kth pilot symbol and a function $D_p(.)$ (5.163) that only depends on $Z_a(\mathbf{z})$ (5.164). Taking this into account, E_1 and E_2 can be rewritten as follows:

$$E_1 = (N_p)^2 E\left[\left(\Im\left\{D_p\left(Z_a(\mathbf{z})\right)Z_b(\mathbf{z})\right\}\right)^2\right], \tag{5.170}$$

Algorithm 8
PA NCA & DA FIM related to $L(\nu;\mathbf{r})$ for linear modulation (bis)

1. For $n=1,2,...,N$:
   - if PA NCA estimation:
     - generate $a^{(n)}(k)$, $k\in I_d$ according to the uniform distribution $\Pr[a(k)=\omega]=\frac{1}{M}I[\omega\in\Omega]$;
     - generate $\mathbf{n}^{(n)}=\left(n^{(n)}(k):\ k\in I_s\right)$ according to $\mathcal{N}_c\left(\mathbf{0},\frac{N_0}{E_s}\mathbf{I}\right)$;
     - compute $z^{(n)}(k)=a^{(n)}(k)+n^{(n)}(k)$ for $k\in I_s$, with $a^{(n)}(k)=A_k$ for $k\in I_p$;
     - compute $K\left(k;\mathbf{z}^{(n)}\right)=K_s\left(k;\mathbf{z}^{(n)}\right)$ for $k\in I=I_s$ according to (5.161).
   - if DA estimation:
     - generate $\mathbf{n}_p^{(n)}=\left(n^{(n)}(k):\ k\in I_p\right)$ according to $\mathcal{N}_c\left(0,\frac{N_0}{E_s}\mathbf{I}\right)$;
     - compute $z^{(n)}(k)=A_k+n^{(n)}(k)$, for $k\in I_p$;
     - compute $K\left(k;\mathbf{z}^{(n)}\right)=K_p\left(k;\mathbf{z}^{(n)}\right)$ for $k\in I=I_p$ according to (5.162).
2. Compute E_1 and E_2 as:

$$E_1=\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k\in I}(k-\kappa_0)\,\Im\left\{K\left(k;\mathbf{z}^{(n)}\right)z^{(n)}(k)\right\}\right)^2,$$

$$E_2=\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k,l\in I}\mathrm{f}((k-l)T)\,\Re\left\{K\left(k;\mathbf{z}^{(n)}\right)K^*\left(l;\mathbf{z}^{(n)}\right)\right\}\right).$$

3. If simplified observation model: compute $J_1=J_\nu$ as

$$J_\nu=-\left(\frac{2E_s}{N_0}\right)^2 4\pi^2 E_1.$$

4. If correct observation model: compute $J_1=J_\nu$ as

$$J_\nu=\left(\frac{2E_s}{N_0}\right)^2 4\pi^2\left(\frac{N_0}{2E_s}E_2-E_1\right).$$

$$E_2 = E\left[\left|D_p\left(Z_a\left(\mathbf{z}\right)\right)\right|^2\right]\sum_{k,l\in I_p}\mathrm{f}\left(\left(k-l\right)T\right)A_k^* A_l, \tag{5.171}$$

where

$$Z_b\left(\mathbf{z}\right) = \frac{1}{N_p}\sum_{k\in I_p}\left(k-\kappa_0\right)A_k^* z\left(k\right), \tag{5.172}$$

such that E_1 and E_2 only depend on $\mathbf{z}$ through $Z_a\left(\mathbf{z}\right)$ and $Z_b\left(\mathbf{z}\right)$. This implies that the DA FIM related to $L\left(\nu;\mathbf{r}\right)$ can be numerically evaluated by substituting (5.170) and (5.171) into (5.167) and replacing in (5.170) and (5.171) the statistical average $E\left[.\right]$ by a numerical average over a large number of realizations of $Z_a\left(\mathbf{z}\right)$ and $Z_b\left(\mathbf{z}\right)$ that are computer-generated according to the joint distribution of $Z_a\left(\mathbf{z}\right)$ and $Z_b\left(\mathbf{z}\right)$. It follows directly from the statistical properties of $\mathbf{z}$ (see Appendix F on page 207) that the quantities $\left(Z_a\left(\mathbf{z}\right), Z_b\left(\mathbf{z}\right)\right)$ are jointly Gaussian complex random variables with means $\left(\mu_a, \mu_b\right)$, and covariance matrix $\mathbf{C}$ with entries $C_{1,1} = C_{a,a}$, $C_{1,2} = C_{a,b} = C_{2,1} = C_{b,a}$ and $C_{2,2} = C_{b,b}$ given by

$$\mu_a = 1 \tag{5.173}$$

$$\mu_b = \kappa_{m,p} - \kappa_0 \tag{5.174}$$

$$C_{a,a} = \frac{N_0}{N_p E_s} \tag{5.175}$$

$$C_{a,b} = C_{b,a} = \frac{N_0}{N_p E_s}\left(\kappa_{m,p} - \kappa_0\right) \tag{5.176}$$

$$C_{b,b} = \frac{N_0}{N_p E_s}\left(\sigma_p^2 + \left(\kappa_{m,p} - \kappa_0\right)^2\right) \tag{5.177}$$

where the quantities $\kappa_{m,p} = \frac{1}{N_p}\sum_{k\in I_p} k$ and $\sigma_p^2 = \frac{1}{N_p}\sum_{k\in I_p}\left(k-\kappa_{m,p}\right)^2$ are defined as in section 5.5. This leads to the algorithm described in Algorithm 9. The resulting numerical complexity is slightly reduced as compared to Algorithm 8, because the generation of N_p independent Gaussian complex random variables is replaced by the generation of only 2 jointly Gaussian complex random variables in Step 1. In fact, the computational complexity of this procedure is independent of (and does not increase linearly with) the number of pilot symbols N_p.

5.8.9 Computing the NPA NCA FIMs for Linear Modulation

The NPA NCA FIMs are a special case of the PA NCA FIMs, namely for pilot symbol free transmission (compare (5.16) to (5.17)). The evaluation of J_ν (5.167) for $\left(I_p, I_d\right) = \left(\emptyset, I_p \bigcup I_d\right)$ yields a closed-form expression for the NPA NCA FIM related to $L\left(\nu;\mathbf{r}\right)$, and the procedure for computing the NPA NCA FIMs related to $L\left(\nu;\mathbf{r}\right)$ and $L\left(\theta;\mathbf{r}\right)$ is the same as for the PA NCA FIMs with $\left(I_p, I_d\right) = \left(\emptyset, I_p \bigcup I_d\right)$ (Algorithms 7 and 8).

Algorithm 9 DA FIM related to $L(\nu; \mathbf{r})$ for linear modulation (ter)

1. For $n = 1, 2, ..., N$:
   - generate $\left(n_a^{(n)}, n_b^{(n)}\right)$ according to $\mathcal{N}_c(\mathbf{0}, \mathbf{C})$, with the entries of $\mathbf{C}$ given by (5.175)-(5.177);
   - compute $Z_a^{(n)} = 1 + n_a^{(n)}$ and $Z_b^{(n)} = (\kappa_{m,p} - \kappa_0) + n_b^{(n)}$;
   - compute $D_p\left(Z_a^{(n)}\right)$ according to (5.163).
2. Compute E_1 and E_2 as:

$$E_1 = (N_p)^2 \frac{1}{N} \sum_{n=1}^{N} \left(\Im \left\{ D_p\left(Z_a^{(n)}\right) Z_b^{(n)} \right\} \right)^2,$$

$$E_2 = \left(\frac{1}{N} \sum_{n=1}^{N} \left| D_p\left(Z_a^{(n)}\right) \right|^2 \right) \sum_{k,l \in I_p} \mathrm{f}\left((k-l)T\right) A_k^* A_l.$$

3. If simplified observation model: compute $J_1 = J_\nu$ as:

$$J_\nu = -\left(\frac{2E_s}{N_0}\right)^2 4\pi^2 E_1.$$

4. If correct observation model: compute $J_1 = J_\nu$ as:

$$J_\nu = \left(\frac{2E_s}{N_0}\right)^2 4\pi^2 \left(\frac{N_0}{2E_s} E_2 - E_1\right).$$

5.8.10 Derivation, Numerical Evaluation and Discussion of the CRB Related to $L(\nu;\mathbf{r})$

Assuming a square-root cosine roll-off transmit pulse with a 30% excess bandwidth and a burst with N_p consecutive pilot symbols and $N_d = N_s - N_p$ data symbols, we now compute the DA and the NPA NCA CRBs related to $L(\nu;\mathbf{r})$, as well as their low-SNR asymptotes (ACRB). The ACRBs are obtained analytically. The computation of the CRBs involves a numerical computation. It follows from the Bayesian Cramer-Rao inequality (3.21) that the MSEs pertaining to the estimation of ν are lower-bounded by the inverse of the scalar FIM $J_1 = J_\nu$ related to $L(\nu;\mathbf{r})$. We obtain: $E\left[(\nu-\hat{\nu})^2\right] \geq CRB_\nu$, with:

$$CRB_\nu = \frac{1}{J_\nu}. \tag{5.178}$$

Algorithm 9 will be used to evaluate the DA FIM and Algorithm 8 will be used to evaluate the NPA NCA FIM. Step 2 of these algorithms involves the evaluation of two numerical averages E_1 and E_2. Note that these averages are with respect to either two jointly Gaussian complex-valued scalars (DA mode) or a complex-valued vector of size N_s (for the NPA NCA mode).

The results are presented and discussed in sections 5.8.11 and 5.8.12.

5.8.11 Numerical Evaluation and Discussion of the DA CRB Related to $L(\nu;\mathbf{r})$

In this subsection, numerical results pertaining to the ratio CRB/MCRB related to $L(\nu;\mathbf{r})$ and DA estimation are obtained for a large number (N_p) of consecutive pilot symbols and $\kappa_0 = \kappa_{m,p}$ (i.e. the normalized time instant that corresponds to the center of the pilot part of the symbol burst, see (5.28)). Under these conditions, the variables Z_a (5.164) and Z_b (5.172) are statistically independent, such that the statistical average over $Z_b(\mathbf{z})$ in (5.170) can be computed analytically, which in turn allows to write the ratio CRB/MCRB related to $L(\nu;\mathbf{r})$ and DA estimation as follows:

$$\frac{CRB_\nu|_{\kappa_0=\kappa_{m,p}}}{MCRB_\nu(N_p)|_{\kappa_0=\kappa_{m,p}}} \approx \frac{1}{E_{Z_a}\left[|D_p(Z_a)|\right]}, \quad N_p \gg 1, \tag{5.179}$$

which holds for both observation models. It follows directly from (5.159), (5.163) and the pdf $\mathcal{N}_c\left(1, \frac{N_0}{N_p E_s}\right)$ of $Z_a(\mathbf{z})$ that the ratio CRB/MCRB (5.179) is a function of $N_p E_s/N_0$, rather than of N_p and E_s/N_0 separately. The low-SNR asymptote of the ratio CRB/MCRB (5.179) has been derived in [2]:

$$\frac{ACRB_\nu|_{\kappa_0=\kappa_{m,p}}}{MCRB_\nu(N_p)|_{\kappa_0=\kappa_{m,p}}} \approx \frac{N_0}{3N_p E_s}, \quad N_p \gg 1, \tag{5.180}$$

which holds for both observation models and behaves inversely proportional to $N_p E_s/N_0$. In (5.179) and (5.180), $MCRB_\nu(N_p)$ denotes the DA MCRB related

to $L(\nu;\mathbf{r})$. This MCRB is given by (5.45) (for both observation models) and coincides with the DA MCRB related to $L(\boldsymbol{\Psi};\mathbf{r})$ and pertaining to the estimation of ν when $\kappa_0 = \kappa_{m,p}$. Because the DA MCRBs related to $L(\boldsymbol{\Psi};\mathbf{r})$ equal the corresponding DA CRBs related to $L(\boldsymbol{\Psi};\mathbf{r})$(see section 5.7.10), the ratio CRB/MCRB from (5.179) can also be interpreted as the ratio of the CRB related to $L(\nu;\mathbf{r})$ to the CRB related to $L(\boldsymbol{\Psi};\mathbf{r})$ and pertaining to the estimation of ν. We recall that the bound using $L(\nu;\mathbf{r})$ is always greater than the bound using $L(\boldsymbol{\Psi};\mathbf{r})$, and therefore more tight [49]. This is perfectly in keeping with the fact that the MCRB in the denominator of (5.179) is in general looser than the corresponding CRB in the numerator of (5.179) [50].

Fig. 5.15 shows the numerical evaluation of (5.179) and (5.180) as a function of $N_p E_s/N_0$. We observe that, for small (large) SNR, the ratio CRB/MCRB (5.179) converges to the ratio ACRB/MCRB (5.180) (to 1). For a given number of pilot symbols, the ratio CRB/MCRB is essentially equal to 1 when E_s/N_0 exceeds about (20-10log(N_p)) dB. This means that, for a given N_p, the FIMs related to $L(\nu;\mathbf{r})$ and $L(\boldsymbol{\Psi};\mathbf{r})$ yield virtually the same CRB on the frequency estimation error variance above E_s/N_0 = (20-10log(N_p)) dB.

5.8.12 Numerical Evaluation and Discussion of the NPA NCA CRB Related to $L(\nu;\mathbf{r})$

Results pertaining to the NPA NCA CRB for normalized frequency offset estimation are presented for $\kappa_0 = \kappa_{m,s}$ (i.e., the normalized time instant that corresponds to the center of the symbol burst, see (5.24)). In [2], the following low-SNR asymptotic ratios CRB/MCRB (ACRB/MCRB) have been presented for $N_s \gg 1$, $\kappa_0 = \kappa_{m,s}$ and the NPA NCA scenario. For the simplified observation model:

$$\frac{ACRB_\nu|_{\kappa_0=\kappa_{m,s}}}{MCRB_\nu(N_s)|_{\kappa_0=\kappa_{m,s}}} = \frac{((K-1)!)^2}{N_s(E_s/N_0)^{2K-1}},\ (\text{si}), \tag{5.181}$$

where K is related to the symmetry angle $2\pi/K$ of the constellation ($K = M$ for M-PSK, $K = 4$ for M-QAM, $K = 2$ for M-PAM), and for the correct observation model:

$$\frac{ACRB_\nu|_{\kappa_0=\kappa_{m,s}}}{MCRB_\nu(N_s)|_{\kappa_0=\kappa_{m,s}}} = \frac{N_s^2}{12\mathrm{f}(0)(E_s/N_0)},\ (\text{co}), \tag{5.182}$$

where $\mathrm{f}(0) = \int \left(\frac{t}{T}p(t)\right)^2 \mathrm{d}t$ and which is independent of the constellation (size and type), but is affected by the shape of the transmit pulse. In (5.181) and (5.182), the quantity $MCRB_\nu(N_s)$ denotes the NPA NCA MCRB related to $L(\nu;\mathbf{r})$. This MCRB is given by (5.43), which holds for both observation models and coincides with the NPA NCA MCRB related to $L(\boldsymbol{\Psi};\mathbf{r})$ and pertaining to the estimation of ν for $\kappa_0 = \kappa_{m,s}$. Comparing (5.182) with (5.115), we consequently find that, for $K > 2$, large N_s and the correct observation model, the FIMs related to $L(\nu;\mathbf{r})$ and $L(\boldsymbol{\Psi};\mathbf{r})$ yield the same ACRB for frequency estimation.

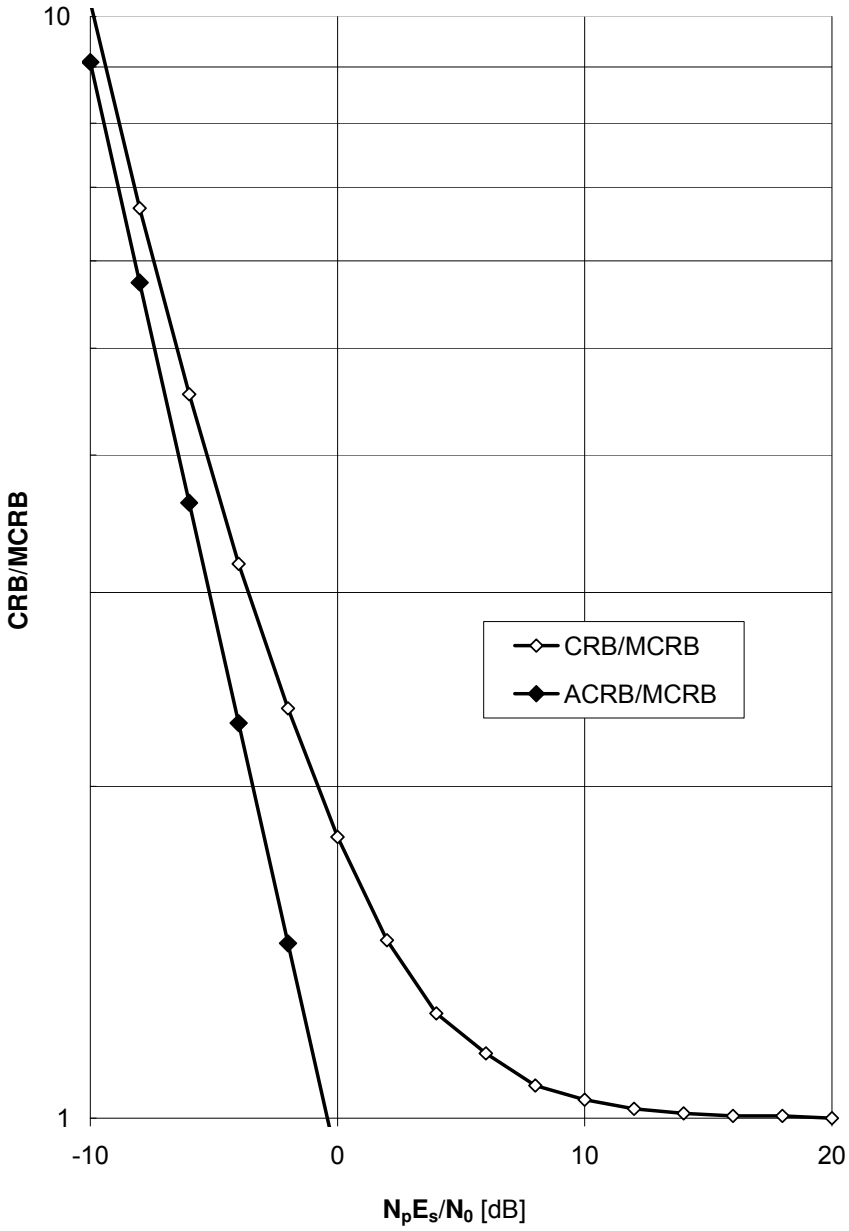

Figure 5.15: Ratio (A)CRB/MCRB related to $L(\nu;\mathbf{r})$, for DA estimation from the observation of N_p consecutive pilot symbol periods.

In Fig. 5.16 we have plotted the ratio:

$$\frac{CRB_{\nu}|_{\kappa_0=\kappa_{m,s}}}{MCRB_{\nu}(N_s)|_{\kappa_0=\kappa_{m,s}}}, \tag{5.183}$$

related $L(\nu;\mathbf{r})$ and NPA NCA estimation as a function of E_s/N_0 for $N_s = 10, 100$, 4-PSK (4-QAM) and for both observation models. The corresponding low-SNR asymptotes (5.181) and (5.182) are also displayed. We find that the behavior of the ratio (5.183) related to $L(\nu;\mathbf{r})$ is comparable to that of the ratios $CRB_{\nu}^{(si)}/MCRB_{\nu}(N_s)$ and $CRB_{\nu}^{(co)}/MCRB_{\nu}(N_s)$ from section 5.7.7 (resulting from the FIM related to $L(\mathbf{\Psi};\mathbf{r})$):

- When E_s/N_0 is sufficiently large, the CRBs converge to the MCRB.
- When E_s/N_0 is sufficiently small, the CRBs are close to the corresponding ACRBs.
- The simplified observation model yields a larger CRB than the correct observation model, but the difference between the two observations models becomes apparent only at very low SNR. For a given value of E_s/N_0, the CRB obtained with the correct observation model approaches the CRB obtained with the simplified observation model, if and when the number of observed samples (N_s) increases. Comparing Fig. 5.16 with Fig. 4.7 from chapter 4 indicates that, even for values of N_s as small as 10, both observation models yield essentially the same ratio CRB/MCRB for all SNR values of practical interest.

The most striking difference between the CRBs related to $L(\nu;\mathbf{r})$ and $L(\mathbf{\Psi};\mathbf{r})$ is that at the SNR values of practical interest, the ratio CRB/MCRB corresponding to $L(\nu;\mathbf{r})$ decreases with N_s, whereas the ratio CRB/MCRB corresponding to $(\mathbf{\Psi};\mathbf{r})$ does not depend on N_s. This is consistent with the observation that for the simplified model, which is relevant for practical values of SNR, the low-SNR ratio ACRB/MCRB corresponding to $L(\nu;\mathbf{r})$ is proportional to N_s^{-1} whereas the low-SNR ratio ACRB/MCRB corresponding to $L(\mathbf{\Psi};\mathbf{r})$ does not depend on N_s.

Assuming the simplified observation model, Fig. 5.17 compares the CRB corresponding to $L(\nu;\mathbf{r})$ and $L(\mathbf{\Psi};\mathbf{r})$, for 4-PSK or 4-QAM. We have verified that, for the ranges of E_s/N_0 and of CRB/MCRB considered in Fig. 5.17, the correct observation model yields virtually the same curves. It should be noted that the bound using $L(\nu;\mathbf{r})$ is always greater than the bound using $L(\mathbf{\Psi};\mathbf{r})$. This is in keeping with the result from the literature (see section 5.2) that the first CRB is a tighter lower bound on the MSE than the second one [49]. We further observe that, for given a E_s/N_0, increasing N_s causes the CRB related to $L(\nu;\mathbf{r})$ to approach the CRB related to $L(\mathbf{\Psi};\mathbf{r})$.

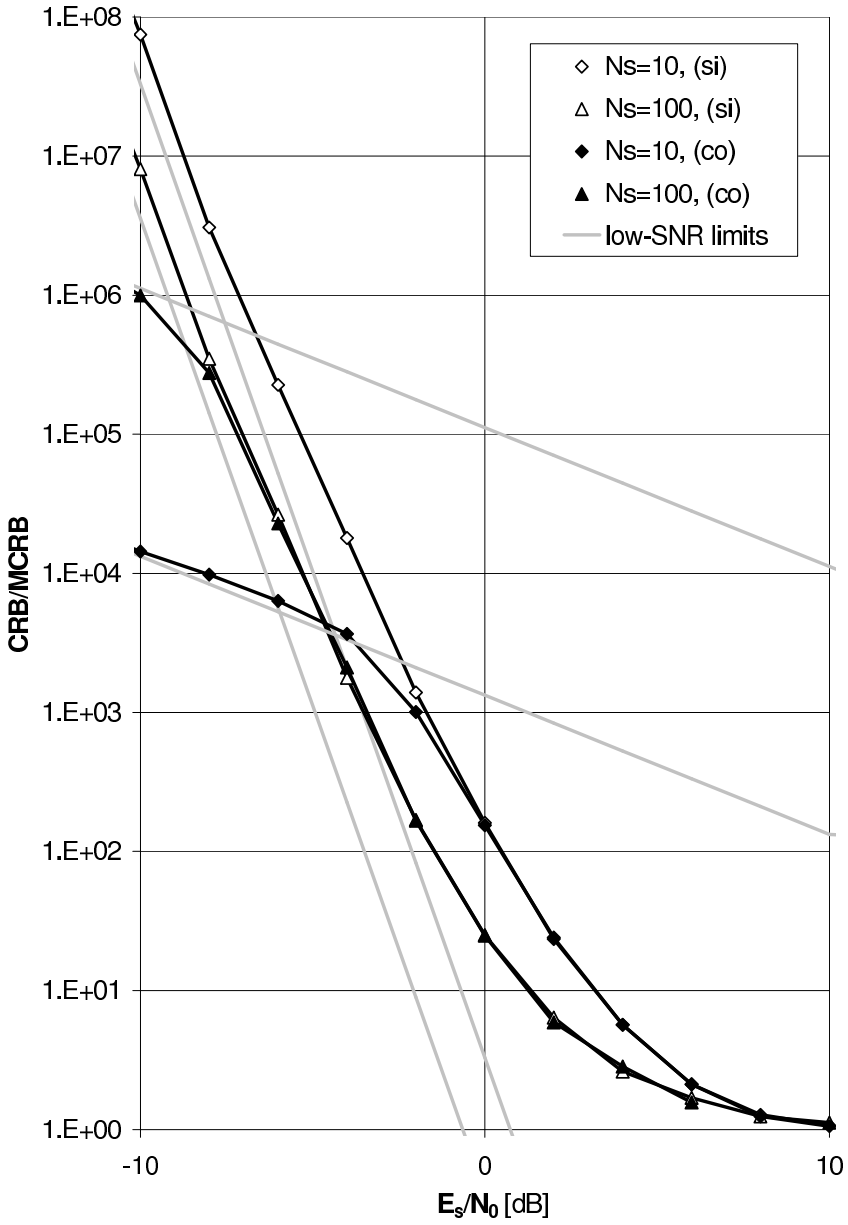

Figure 5.16: Ratio CRB/MCRB related to $L(\nu;\mathbf{r})$, for the NPA NCA estimation mode and a 4-PSK signaling constellation.

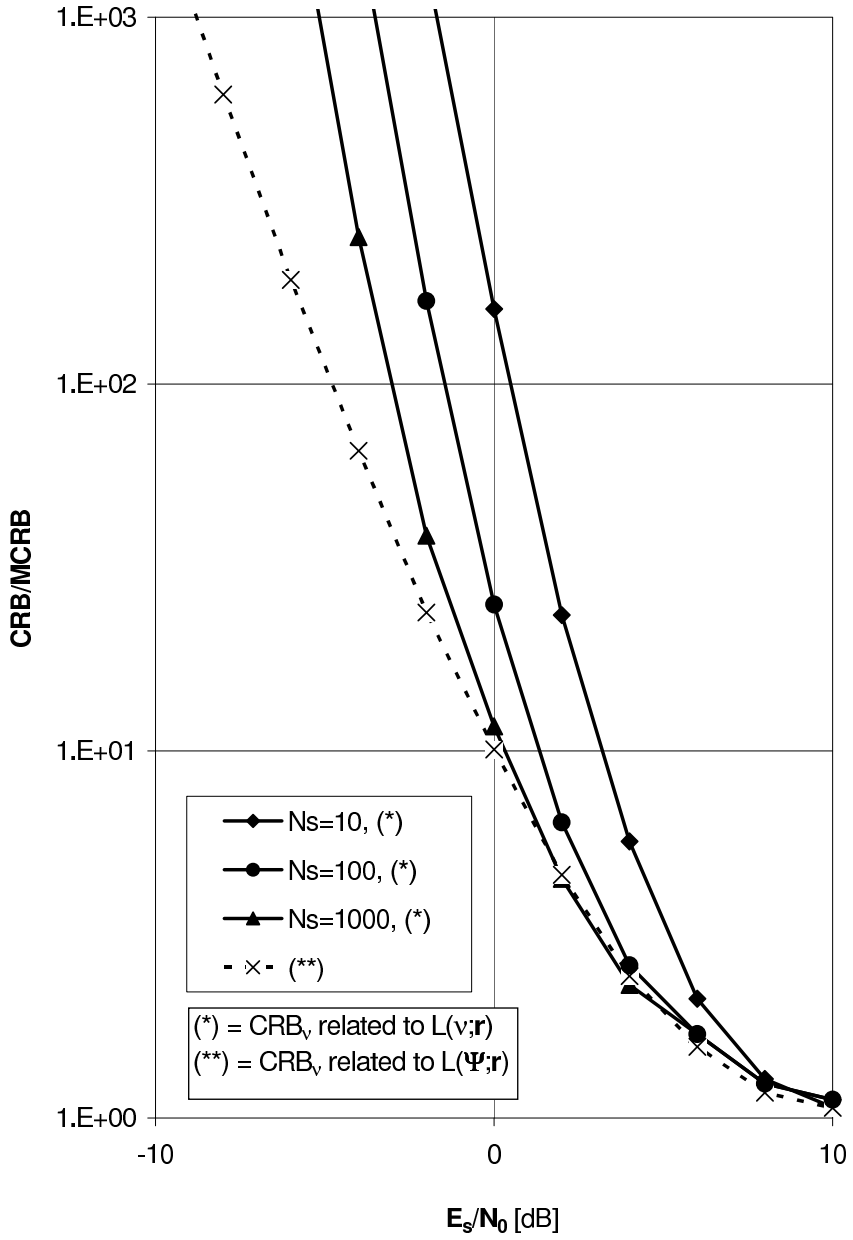

Figure 5.17: Ratio CRB/MCRB related to $L(\nu;\mathbf{r})$ versus ratio CRB/MCRB related to $L(\mathbf{\Psi};\mathbf{r})$, for the NPA NCA estimation mode and a 4-PSK signaling constellation.

5.9 Conclusions and Remarks

In this chapter we have computed Cramer-Rao bounds (CRBs) for the estimation of the (normalized) frequency offset ν and the phase shift θ (at $t = \kappa_0 T$, with T the symbol period) from the noisy observation $\mathbf{r}$ of a linearly-modulated burst signal with an arbitrary square-root Nyquist transmit pulse that contains pilot symbols as well as coded data symbols. Both the correct observation model and an approximation of the correct model used in [55–57], have been considered. Four different estimation modes have been discussed: the pilot-aided code-aided (PA CA) mode, the data-aided (DA) mode, the pilot-aided non-code-aided (PA NCA) mode and the non-pilot-aided non-code-aided (NPA NCA) mode. Bounds have been derived from the 2×2 FIM related to the likelihood function $L(\nu, \theta; \mathbf{r})$, as well as from the scalar FIMs related to the marginal likelihood functions $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ of ν and θ, respectively.

The Fisher information matrix (FIM) related to the likelihood function $L(\nu, \theta; \mathbf{r})$ has been expressed in terms of the marginal symbol a posteriori probabilities. Taking into account the linear modulation, we have derived a numerical procedure that allows the efficient numerical evaluation of the corresponding CRBs. This procedure holds for both observation models and for the different estimation modes. For the DA, the PA NCA and the NPA NCA modes, some additional simplifications have been performed. The DA FIM related to $L(\nu, \theta; \mathbf{r})$ has been evaluated analytically, and coincides with the DA modified FIM (MFIM) related to $L(\nu, \theta; \mathbf{r})$. The PA NCA and the NPA NCA FIMs related to $L(\nu, \theta; \mathbf{r})$ have been expressed in terms of one (for the simplified observation model) or two (for the correct observation model) statistical averages over a two-dimensional Gaussian pdf. These averages depend on the signaling constellation, but not on the pulse shape or on the burst structure; the effect of the pulse shape and of the location of the pilot symbols is analytically accounted for.

The numerical evaluation of the PA CA, the PA NCA and the NPA NCA CRBs related to $L(\nu, \theta; \mathbf{r})$ shows that, at the normal operating signal-to-noise ratio (SNR) of a coded system (say, bit error rate (BER) smaller than 10^{-3}), the PA CA CRB is very close to the corresponding MCRB. Furthermore, the PA CA CRB is less than the PA NCA CRB, which in turn is less than the NPA NCA CRB. This indicates that in order to approach optimal performance, estimators should make clever use of both the code properties and pilot symbols during the estimation process. The numerical results also show that the effect of the constellation on the CRB is small, for both low and high SNR. For moderate SNR, the CRB increases with increasing constellation size. Finally, it was observed that the correct and the simplified model yield CRBs that are substantially different only at low SNR. The influence of the pulse shape is restricted to these low SNR values. The simplified model of the matched filter output samples ignores the inter-symbol-interference (ISI) and the reduction of the useful symbol magnitude, which are caused by the frequency offset νT at the input of the matched filter. For all practical values of the SNR we have verified (details not reported here) that the useful signal magnitude is reduced

by less than 0.01 dB and the ISI power is at least 20 dB below the noise power, provided that $|\nu| < 0.015$ ($|\nu| < 0.030$) for a cosine roll-off transmit pulse with roll-off factor of 20 % (of 100%). Consequently, the simplified observation model is valid as long as the maximum normalized frequency offset is about 1 %.

In general, the FIMs related to the likelihood functions $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ are found much harder to compute than the FIM related to $L(\nu, \theta; \mathbf{r})$. Taking the linear modulation into account and assuming statistical independent (i.e., uncoded) symbols, we have derived numerical procedures for evaluating the FIMs related to $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$, respectively (comparable to that for computing the FIM related to $L(\nu, \theta; \mathbf{r})$). These procedures hold for the DA, the PA NCA and the NPA NCA estimation modes, but not for the PA CA estimation mode. The computation of the PA CA FIMs related to $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ seems practically unfeasible. The procedure for evaluating the DA, PA NCA or NPA NCA FIM related to $L(\nu; \mathbf{r})$ has been further simplified in the case of the correct observation model; whereas a similar reduction of the computational complexity proved to be impossible for the FIM related to $L(\theta; \mathbf{r})$.

The DA CRB related to $L(\nu; \mathbf{r})$ has been numerically computed for κ_0 being the normalized time instant that corresponds to the center of the pilot symbols and assuming a large number of consecutive pilot symbols and the NPA NCA CRB related to $L(\nu; \mathbf{r})$ has been numerically computed for κ_0 being the normalized time instant that corresponds to the center of the symbol burst. Comparing the resulting CRBs to the DA and the NPA NCA CRB for frequency estimation related to $L(\nu, \theta; \mathbf{r})$ shows that the CRBs related to $L(\nu; \mathbf{r})$ are larger than the corresponding CRBs related to $L(\nu, \theta; \mathbf{r})$. This implies that the former CRBs are the tightest bounds, which is in keeping with [49]. For a given SNR, the gap between the first and the second type of CRBs decreases with increasing the observation interval. For the DA CRB, this penalty can be neglected for practical values of SNR, even for moderate observation intervals. For the NPA NCA CRB, considerably longer observation intervals are required in order to make the penalty very small.

The scientific contributions [1–4, 7–14] describe the original work related to this chapter. A part of the work in [3] is related to timing recovery, which falls outside the scope of this doctoral thesis.

6

FF Carrier Synchronizers for Coded Systems

In this chapter we will discuss a variety of methods to estimate the carrier synchronization parameters (frequency offset and phase shift) of a linearly modulated bandpass signal that is corrupted by additive white Gaussian noise (AWGN). We will focus on algorithms that are very close approximations of the true *maximum likelihood* (ML) estimator. We will present simulation results pertaining to the mean square error (MSE) performance of these estimators and we will compare this MSE performance to the associated Cramer-Rao bounds (CRBs) derived in chapter 5. We are particularly concerned with the design of the carrier synchronization section of receivers with a maximum-a-posteriori (MAP) bit detector operating according to the sum-product algorithm (as with turbo and low-density parity-check (LDPC) codes). At the extremely low signal-to-noise ratio (SNR) that is typical of their operation, we search for synchronization schemes that are capable of providing an estimation accuracy that ensures a negligible degradation of the bit error rate (BER) at the output of the detector, as compared to the case of perfect synchronization.

The remainder of this chapter is organized as follows. First, section 6.1

summarizes the system model under consideration. Then true ML carrier estimation is discussed in section 6.2 and section 6.3 describes the classification of the practical ML-based synchronization algorithms. Next, *data-aided* (DA) estimation is revisited in section 6.4, section 6.5 is dedicated to a widely-adopted *non-pilot-aided non-code-aided* (NPA NCA) method, referred to as Kth-power estimation and section 6.6 covers *pilot-aided code-aided* (PA CA) estimation. Finally, section 6.7 provides some concluding remarks.

6.1 Observation Model

As before, we will assume the transmission of a linearly modulated signal over a bandlimited additive white Gaussian noise (AWGN) channel. The transmitted symbol burst consists of N_s symbols and is denoted as a row vector $\mathbf{a} = (a(k) : k \in I_s)$. The symbol indices k belong to a set I_s of N_s consecutive integers. There are N_p pilot symbols and $N_d = N_s - N_p$ data symbols. The quantity $\lambda_p = N_p/N_s\,(\ll 1)$ is referred to as the pilot symbol ratio. For k belonging to the set of indices $I_p \subset I_s$, $a(k)$ is a known pilot symbol with value A_k and $|A_k| = 1$. For k belonging to the set of indices $I_d = I_s \backslash I_p$ the symbol $a(k)$ is an unknown data symbol. The data symbols result from the encoding and mapping of a sequence of N_b information bits. We use an M-fold signaling constellation which is denoted $\Omega = \{\omega_0, \omega_1, ..., \omega_{M-1}\}$ and assume that $M^{-1}\sum_{m=0}^{M-1} |\omega_m|^2 = 1$. The instantaneous phase shift of the received signal vis-à-vis the local reference carrier is denoted as $\Theta(t; \mathbf{\Psi})$, where $\mathbf{\Psi}$ is a parameter vector that fully characterizes $\Theta(t; \mathbf{\Psi})$.

The conceptual block diagram of the receiver is depicted in Fig. 6.1. The received signal is applied to a matched filter, sampled at the optimal time instants kT (with T denoting the symbol period), and fed to the carrier synchronizer. Note that we implicitly assume that carrier phase correction is performed *after* matched filtering. As compared to a receiver with carrier phase correction in front of the matched filter, this yields a remarkable reduction in the receiver complexity because the matched filter output samples $\{r(k)\}$ need to be computed only once, since they are not affected by the carrier phase correction. Assuming that the variations of $\Theta(t; \mathbf{\Psi})$ are slow as compared to the symbol rate $1/T$, this practice causes a negligible performance degradation. In this case, we can use the simplified observation model from section 4.5: $\mathbf{r} = (r(k) : k \in I_s)$, with the samples $r(k)$ given by:

$$r(k) = a(k)\, e^{j\Theta(kT;\mathbf{\Psi})} + w(k). \tag{6.1}$$

In (6.1), $\{w(k)\}$ denotes a sequence of independent and identically distributed (i.i.d.) complex-valued Gaussian random variables with i.i.d. real and imaginary parts, each having a power spectral density equal to $N_0/2E_s$. The quantity E_s/N_0 is referred to as the signal-to-noise ratio (SNR).

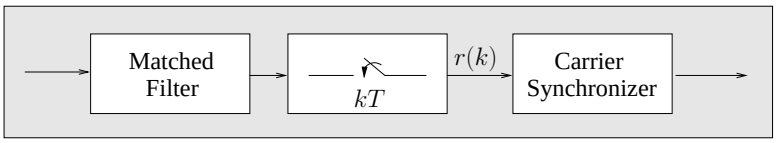

Figure 6.1: Receiver structure.

6.2 ML Estimation

We further assume that the carrier phase shift of the received signal vis-a-vis the receiver's local reference carrier is linear in time, i.e.,

$$\Theta\left(t;\mathbf{\Psi}\right)=2\pi F\left(t-t_0\right)+\theta=2\pi\nu\left(\kappa-\kappa_0\right)+\theta. \tag{6.2}$$

In (6.2), F is the carrier frequency offset and $\theta=\Theta\left(t_0;\mathbf{\Psi}\right)$ is the phase shift at an arbitrary time instant t_0. In order to deal with dimensionless quantities, we introduce the normalized frequency offset $\nu=FT$ and the normalized time variables $\kappa=t/T$ and $\kappa_0=t_0/T$. The quantities ν and θ are uniformly distributed in the intervals $[-\nu_{max},\nu_{max}]$ and $[-\pi,\pi]$, respectively. The column vector $\mathbf{\Psi}$ collects the unknown parameters ν and θ:

$$\mathbf{\Psi}=\left(\begin{array}{c}\Psi_1\\ \Psi_2\end{array}\right)=\left(\begin{array}{c}\nu\\ \theta\end{array}\right). \tag{6.3}$$

The a priori probability density function of $\mathbf{\Psi}$ is denoted as $p\left(\mathbf{\Psi}\right)$. From the observation vector $\mathbf{r}$, we want to recover the value of $\mathbf{\Psi}$.

According to (3.17), the *maximum a posteriori* (MAP) estimate $\hat{\mathbf{\Psi}}\left(\mathbf{r}\right)$ of the carrier synchronization vector $\mathbf{\Psi}$ is the value of $\mathbf{\Psi}$ that maximizes the a posteriori probability $p\left(\mathbf{\Psi}\left|\mathbf{r}\right.\right)$. We have:

$$\hat{\mathbf{\Psi}}\left(\mathbf{r}\right)=\arg\max_{\tilde{\mathbf{\Psi}}}p\left(\tilde{\mathbf{\Psi}}\left|\mathbf{r}\right.\right), \tag{6.4}$$

where $p\left(\tilde{\mathbf{\Psi}}\left|\mathbf{r}\right.\right)$ denotes the a posteriori probability $p\left(\mathbf{\Psi}\left|\mathbf{r}\right.\right)$ of $\mathbf{\Psi}$ for a given value of $\mathbf{r}$, evaluated at $\mathbf{\Psi}=\tilde{\mathbf{\Psi}}$. This MAP estimator is known to be asymptotically optimal in the sense that it achieves the MSE performance predicted by the CRB when the number of observed matched filter output samples (N_p) is large (see chapter 3).

It follows directly from Bayes' rule that:

$$p\left(\mathbf{\Psi}\left|\mathbf{r}\right.\right)=\frac{L\left(\mathbf{\Psi};\mathbf{r}\right)p\left(\mathbf{\Psi}\right)}{p\left(\mathbf{r}\right)}, \tag{6.5}$$

where $L\left(\mathbf{\Psi};\mathbf{r}\right)$ denotes the likelihood function of $\mathbf{\Psi}$, resulting from $\mathbf{r}$; we have:

$$L\left(\mathbf{\Psi};\mathbf{r}\right)=p\left(\mathbf{r}\left|\mathbf{\Psi}\right.\right).$$

Because the a priori probability $p(\mathbf{\Psi})$ is uniform, the MAP estimation rule (6.4) reduces to the *maximum likelihood* (ML) estimation rule (see (3.19)), which comes down to maximizing the likelihood function $L(\mathbf{\Psi};\mathbf{r})$, or, equivalently, the log-likelihood function $\ell(\mathbf{\Psi};\mathbf{r})$ of $\mathbf{\Psi}$; we have:

$$\ell(\mathbf{\Psi};\mathbf{r}) = \ln p(\mathbf{r}\,|\mathbf{\Psi}) \tag{6.6}$$

and

$$\hat{\mathbf{\Psi}}(\mathbf{r}) = \arg\max_{\tilde{\mathbf{\Psi}}} \ell\left(\tilde{\mathbf{\Psi}};\mathbf{r}\right), \tag{6.7}$$

where $\ell\left(\tilde{\mathbf{\Psi}};\mathbf{r}\right)$ denotes the log-likelihood function $\ell(\mathbf{\Psi};\mathbf{r})$ of $\mathbf{\Psi}$ resulting from $\mathbf{r}$, evaluated at $\mathbf{\Psi} = \tilde{\mathbf{\Psi}}$.

Conceptually, the derivation of the ML carrier synchronization algorithms is straightforward. The difficulty lies in the fact that the statistics of the observation vector $\mathbf{r}$ depend not only on the carrier synchronization parameter vector $\mathbf{\Psi}$, but also on the symbol vector $\mathbf{a}$. It follows directly from (6.1) that the joint conditional probability density function $p(\mathbf{r}|\,\mathbf{a},\mathbf{\Psi})$ of $\mathbf{r}$, given the symbol vector $\mathbf{a}$ and the carrier synchronization parameter vector $\mathbf{\Psi}$, is given by:

$$p(\mathbf{r}|\,\mathbf{a},\mathbf{\Psi}) = \prod_{k\in I_s} p(r(k)|\,a(k),\mathbf{\Psi}), \tag{6.8}$$

where $p(r(k)|\,a(k),\mathbf{\Psi})$ is the marginal conditional probability density function of the matched filter output sample $r(k)$ at instant kT, given the symbol $a(k)$ and $\mathbf{\Psi}$, which in turn is (within a factor not depending on $a(k)$ and $\mathbf{\Psi}$) given by:

$$\begin{aligned} &p(r(k)|\,a(k),\mathbf{\Psi}) \\ &\propto \exp\left(\frac{E_s}{N_0}\left(2\Re\left\{a^*(k)\,z(k;r(k),\mathbf{\Psi}) - |a(k)|^2\right\}\right)\right), \end{aligned} \tag{6.9}$$

with

$$z(k;r(k),\mathbf{\Psi}) = \check{z}(k;r(k),\nu)\,e^{-j\theta} \tag{6.10}$$

and

$$\check{z}(k;r(k),\nu) = r(k)\,e^{-j2\pi\nu(k-\kappa_0)}. \tag{6.11}$$

Evaluating the log-likelihood function $\ell(\mathbf{\Psi};\mathbf{r})$ from (6.6) requires averaging $p(\mathbf{r}|\,\mathbf{a},\mathbf{\Psi})$ from (6.8) over the nuisance parameter vector $\mathbf{a}$:

$$\ell(\mathbf{\Psi};\mathbf{r}) \quad = \quad \ln\sum_{\tilde{\mathbf{a}}} \Pr[\mathbf{a}=\tilde{\mathbf{a}}]\,p(\mathbf{r}\,|\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi}), \tag{6.12}$$

where $\tilde{\mathbf{a}} = (\tilde{a}(k) : k \in I_s)$ denotes a trial value of the symbol vector $\mathbf{a}$. The sum in (6.12) ranges over all possible symbol vectors $\tilde{\mathbf{a}}$ with a priori distribution $\Pr[\mathbf{a}=\tilde{\mathbf{a}}]$. When all 2^{N_b} information bit sequences are equally likely,

$$\Pr[\mathbf{a}=\tilde{\mathbf{a}}] = 2^{-N_b} I[\tilde{\mathbf{a}} \in \zeta_{\mathcal{A}}],\ \forall \tilde{\mathbf{a}} \in \Omega^{N_s}, \tag{6.13}$$

where $\zeta_{\mathcal{A}}$ denotes the set of legitimate symbol vectors $\mathbf{a}$, i.e., the set of symbol vectors that result from the encoding of all possible information bit sequences of size N_b, followed by bit-to-symbol mapping and pilot symbol insertion. The function $I[.]$ is the indicator function, i.e., for a predicate P:

$$I[P] = \begin{cases} 0, \text{ if } P \text{ is false} \\ 1, \text{ if } P \text{ is true} \end{cases}. \tag{6.14}$$

As the summation in (6.12) gives rise to a computational burden that is exponential in N_b, ML estimation as defined in (6.7) cannot be used in practice and one has to resort to approximative techniques. The various carrier synchronization parameter estimation algorithms are the result of applying these techniques, which can either be systematic or ad hoc [34].

6.3 Estimation Modes

As explained in section 5.4, the operation of the various carrier synchronization parameter estimators is commonly classified into four different estimation modes: the *pilot-aided code-aided* (PA CA), the *data-aided* (DA), the *pilot-aided non-code-aided* (PA NCA) and the *non-pilot-aided non-code-aided* (NPA NCA) estimation mode.

1. Estimators that take into account the full a priori information (6.13) are said to operate in a PA CA mode.

2. DA operation means computing the estimate $\hat{\mathbf{\Psi}}(\mathbf{r})$ as if only the pilot symbols $(a(k) : k \in I_p)$ were transmitted, i.e., by assuming that the data symbols $(a(k) : k \in I_d)$ are zero:

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}}] = \prod_{k \in I_p} I[\tilde{a}(k) = A_k] \prod_{k \in I_d} I[\tilde{a}(k) = 0], \forall \tilde{\mathbf{a}} \in \Omega^{N_s}, \tag{6.15}$$

   where A_k is the actual value of the pilot symbol $a(k) \in I_p$. This approach neglects the received signal energy associated with the data symbols.

3. PA NCA algorithms rely on the assumption that all possible combinations of random data symbols $(a(k) : k \in I_d)$ are equiprobable:

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}}] = M^{-N_d} \prod_{k \in I_p} I[\tilde{a}(k) = A_k], \forall \tilde{\mathbf{a}} \in \Omega^{N_s}, \tag{6.16}$$

   where again A_k is the actual value of the pilot symbol $a(k) \in I_p$ and M is the order of the signalling constellation. PA NCA algorithms ignore the structure that is enforced upon the data symbols by the channel code and assume uncoded transmission with pilot symbols.

4. The NPA NCA approach consists of ignoring both the structure of the channel code and the fact that the receiver knows the value of the pilot symbols. This comes down to the assumption that all possible symbol vectors $\mathbf{a}$ are a priori equiprobable, i.e.,

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}}] = M^{-N_s}, \forall \tilde{\mathbf{a}} \in \Omega^{N_s}, \tag{6.17}$$

where M is the order of the signalling constellation. NPA NCA algorithms drop all the available statistical a priori information about the transmitted symbol vector.

The MSEs resulting from estimation algorithms that rely on a particular hypothesis about $\Pr[\mathbf{a} = \tilde{\mathbf{a}}]$ are lower-bounded by the CRBs that are computed according to (3.25) but with the same hypothesis about $\Pr[\mathbf{a} = \tilde{\mathbf{a}}]$. CRBs pertaining to the PA CA, DA, PA NCA and NPA NCA estimation of $\mathbf{\Psi}$ have been obtained in chapter 5. We have seen that, for all practical values of λ_p and E_s/N_0, the DA CRBs are larger than the PA CA, the PA NCA or the NPA NCA CRBs. This is due to the fact that the DA estimation mode benefits only from the received signal energy associated with the pilots. We have also seen that the PA CA, the PA NCA and the NPA CRBs are equal at sufficiently high SNRs. This means that, as long as the SNR is high enough, operating in PA CA, PA NCA or NPA NCA mode will lead to the same estimator performance limit. At lower SNRs, however, we have found that there is a gap between the (theoretical) performances that can be achieved by PA CA estimators on the one hand, and PA NCA and NPA NCA estimators on the other. When E_s/N_0 decreases, a point $(E_s/N_0)_{thr}$ is reached where the CRBs resulting from PA CA, PA NCA and NPA NCA estimators begin to diverge from their high-SNR limit. For the PA CA scenario, $(E_s/N_0)_{thr}$ corresponds to a coded SER of about 10^{-3}. For PA NCA and NPA NCA operation, $(E_s/N_0)_{thr}$ corresponds to an uncoded SER of about 10^{-3}, and consequently exceeds $(E_s/N_0)_{thr}$ for PA CA operation by an amount equal to the coding gain. It follows that PA NCA and NPA NCA estimators might fail to provide reliable estimates at the small operating SNR of the powerful turbo code at hand. In such cases, PA CA estimators are needed to achieve accurate synchronization. Unfortunately, taking into account all the available statistical information about $\mathbf{a}$ in the estimation process often turns out to be very complex (essentially because of the complexity of the encoding rule).

6.4 DA Estimation

6.4.1 Algorithm Derivation

Let us assume that only the pilot symbols have been transmitted, or equivalently, that the a priori probability mass function $\Pr[\mathbf{a}]$ of the symbol vector $\mathbf{a}$ is given by (6.15). Substituting this into (6.12) yields the DA log-likelihood function $\ell_{DA}(\mathbf{\Psi}; \mathbf{r})$. Using (6.8) and (6.9), we obtain within a constant not depending

on $\mathbf{\Psi}$:

$$\begin{aligned}
&\ell_{DA}\left(\mathbf{\Psi};\mathbf{r}\right)\\
&=\ln\prod_{k\in I_p}p\left(r\left(k\right)|\,a\left(k\right)=A_k,\mathbf{\Psi}\right), &(6.18)\\
&=2\frac{E_s}{N_0}\sum_{k\in I_p}\Re\left\{A_k^*z\left(k;r\left(k\right),\mathbf{\Psi}\right)\right\}. &(6.19)
\end{aligned}$$

The DA ML carrier parameter estimates are determined by maximizing the log-likelihood function from (6.19) over $\mathbf{\Psi}$. Discarding the constant factor $2E_s/N_0$ from the maximization, we obtain the following estimation rule:

$$\left(\hat{\nu}\left(\mathbf{r}\right),\hat{\theta}\left(\mathbf{r}\right)\right) = \arg\max_{\left(\tilde{\nu},\tilde{\theta}\right)}B_{DA}\left(\tilde{\nu},\tilde{\theta};\mathbf{r}\right), \tag{6.20}$$

where $\left(\tilde{\nu},\tilde{\theta}\right)$ denotes a trial value of (ν,θ) and $B_{DA}\left(\nu,\theta;\mathbf{r}\right)$ is referred to as the objective function. We have:

$$\begin{aligned}
B_{DA}\left(\nu,\theta;\mathbf{r}\right) &= \Re\left\{Y_{DA}\left(\nu;\mathbf{r}\right)e^{-j\theta}\right\}, &(6.21)\\
&= \left|Y_{DA}\left(\nu;\mathbf{r}\right)\right|\cos\left(\arg\left\{Y_{DA}\left(\nu;\mathbf{r}\right)\right\}-\theta\right), &(6.22)
\end{aligned}$$

with

$$Y_{DA}\left(\nu;\mathbf{r}\right)=\sum_{k\in I_p}A_k^*\tilde{z}\left(k;r\left(k\right),\nu\right)=\sum_{k\in I_p}A_k^*r\left(k\right)e^{-j2\pi\nu\left(k-\kappa_0\right)}. \tag{6.23}$$

Note from (6.23) that from the N_s observed samples, only the N_p samples that correspond to the transmission of pilot symbols are used for the estimation. DA carrier synchronization is therefore intrinsically suboptimal. As no approximations are involved in (6.20)-(6.23), the estimates resulting from (6.20) are true ML estimates, hence converging to the DA CRB for large values of N_p or E_s/N_0 (see section 3.4.2).

The maximum in (6.20) is to be found by first computing, for every value $\tilde{\nu}$ of ν, the value of θ that maximizes the objective function $B_{DA}\left(\tilde{\nu},\theta;\mathbf{r}\right)$; the result is a function of $\tilde{\nu}$ that we denote as $\hat{\theta}\left(\mathbf{r};\tilde{\nu}\right)$. We have:

$$\hat{\theta}\left(\mathbf{r};\tilde{\nu}\right) = \arg\max_{\tilde{\theta}}B_{DA}\left(\tilde{\nu},\tilde{\theta};\mathbf{r}\right)=\arg\left\{Y_{DA}\left(\tilde{\nu};\mathbf{r}\right)\right\}. \tag{6.24}$$

Subsequently, the DA ML normalized frequency offset estimate $\hat{\nu}\left(\mathbf{r}\right)$ is found by maximizing the function $B_{DA}\left(\tilde{\nu},\hat{\theta}\left(\mathbf{r};\tilde{\nu}\right);\mathbf{r}\right)$, which is obtained by replacing in (6.22) the parameter θ with $\hat{\theta}\left(\mathbf{r};\tilde{\nu}\right)$, over $\tilde{\nu}$. Finally, the DA phase shift estimate $\hat{\theta}\left(\mathbf{r}\right)$ is obtained by evaluating the function $\hat{\theta}\left(\mathbf{r};\tilde{\nu}\right)$ at $\tilde{\nu}=\hat{\nu}\left(\mathbf{r}\right)$. When we write this down in a more formal way, we obtain:

$$\begin{cases}\hat{\nu}\left(\mathbf{r}\right)=\arg\max_{\tilde{\nu}}B_{DA}\left(\tilde{\nu},\hat{\theta}\left(\mathbf{r};\tilde{\nu}\right);\mathbf{r}\right)\\ \hat{\theta}\left(\mathbf{r}\right)=\hat{\theta}\left(\mathbf{r};\hat{\nu}\left(\mathbf{r}\right)\right)\end{cases}. \tag{6.25}$$

Using (6.24) and taking into account (6.22), (6.25) becomes:

$$\begin{cases} \hat{\nu}(\mathbf{r}) = \arg\max_{\tilde{\nu}} |Y_{DA}(\tilde{\nu};\mathbf{r})| \\ \hat{\theta}(\mathbf{r}) = \arg\{Y_{DA}(\hat{\nu}(\mathbf{r});\mathbf{r})\} \end{cases}. \tag{6.26}$$

In this way, the two-dimensional search in (6.20) over ν and θ has been reduced to a one-dimensional search. For normalized frequency offset estimation we need only to maximize the absolute value of the function $Y_{DA}(\nu;\mathbf{r})$ from (6.23). Once an estimate of ν is made, the estimate of θ follows from straightforward computation using the second line of (6.26). The instantaneous phase shift estimate is then computed as:

$$\Theta\left(t;\hat{\mathbf{\Psi}}(\mathbf{r})\right) = \hat{\theta}(\mathbf{r}) + 2\pi\hat{\nu}(\mathbf{r})\left(\frac{t}{T} - \kappa_0\right). \tag{6.27}$$

The most difficult and time-consuming part of the algorithm (6.26) is locating the maximum of $|Y_{DA}(\tilde{\nu};\mathbf{r})|$. This part is essentially a search routine. It can be circumvented by using other, more ad-hoc, DA frequency estimation methods (see for example [32]). Within the scope of this doctoral thesis, we will not dig into this any deeper. What we will do, however, is calculate $|Y_{DA}(\tilde{\nu};\mathbf{r})|$ for a discrete set of N (with N even) equispaced frequencies $\{\nu_n\}$ defined by $\nu_n = \frac{n}{N}$ for $n = -\frac{N}{2}, -\frac{N}{2}+1, ..., -\frac{N}{2}-1$ and identify the value of ν that maximizes $|Y_{DA}(\tilde{\nu};\mathbf{r})|$ over this discrete set. Here, it is tacitely assumed that $|\nu| \ll 0.5$. It follows from the periodicity property of the discrete Fourier transform (DFT) and (6.23) that the set $\{|Y_{DA}(\nu_n;\mathbf{r})|\}$ is the norm of the DFT of a vector of samples $\{A_k^* r(k)\}$ padded with zeros at all time instants where no pilot symbol is available, i.e.,

$$\begin{aligned} \mathbf{y} &= (y(0), y(1), ..., y(N-1)), \\ &= \left(\acute{r}(k_0), \mathbf{0}, \acute{r}(k_1), \mathbf{0}, ..., \acute{r}\left(k_{N_p-1}\right), \mathbf{0}\right), \end{aligned}$$

where $\acute{r}(k) = A_k^* r(k)$, $\left\{k_0, k_1, ..., k_{N_p-1}\right\}$ is an explicit representation of the index set I_p and $\mathbf{0}$ represents an all zero vector. The length of each zero vector is equal to the number of sampling periods between the preceding and succeeding pilot symbol. The right-most zero vector is optional and its length determines the size N of the vector $\mathbf{y}$ and hence also the resolution of the search. Fast Fourier transform (FFT) techniques [58] make the calculation of the set $\{|Y_{DA}(\nu_n;\mathbf{r})|\}$ efficient. Fig. 6.2 shows a block diagram of this frequency search. If the resolution of the search is sufficiently high, this procedure will yield a good approximate location of the global maximum of $\ell_{DA}(\mathbf{\Psi};\mathbf{r})$ from (6.19), and thus provide the DA ML estimates.

6.4.2 Numerical MSE Results and Discussion

We will now simulate the DA estimation of the normalized frequency offset ν and the value θ of the phase shift $\Theta(t;\mathbf{\Psi})$ at $t = \kappa_0 T$ (according to (6.26)). We will assume that the symbol vector $\mathbf{a}$ contains two blocks of $N_p/2$ consecutive

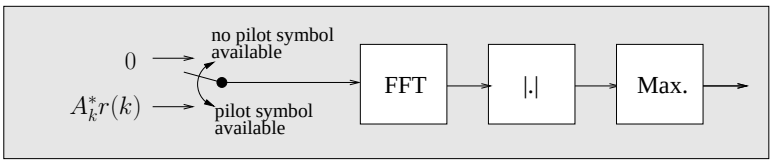

Figure 6.2: DA frequency search.

pilot symbols spaced with S data symbols, and we let the normalized time instant κ_0 correspond to the center $\kappa_{m,p}$ of the pilot part of the symbol burst, i.e., $\kappa_0 = \kappa_{m,p}$, with

$$\kappa_{m,p} = \frac{1}{N_p} \sum_{k \in I_p} k. \tag{6.28}$$

When $\kappa_0 = \kappa_{m,p}$, the DA FIM is diagonal, the DA CRB for the estimation of ν is given by (see chapter 5):

$$DACRB_\nu = -\frac{1}{E_{\mathbf{r},\mathbf{\Psi}}\left[\frac{\partial^2 \ell_{DA}(\mathbf{\Psi};\mathbf{r})}{\partial \nu^2}\right]}, \tag{6.29}$$

$$\approx \frac{3N_0}{2\pi^2 N_p^3 \left(1 + 3\epsilon_{Sp}\left(\epsilon_{Sp} + 1\right)\right) E_s} \tag{6.30}$$

and the DA CRB for the estimation of θ achieves its minimum value, which is given by (see chapter 5):

$$DACRB_\theta = -\frac{1}{E_{\mathbf{r},\mathbf{\Psi}}\left[\frac{\partial^2 \ell_{DA}(\mathbf{\Psi};\mathbf{r})}{\partial \theta^2}\right]} = \frac{N_0}{2N_p E_s}. \tag{6.31}$$

In (6.30), the approximation holds for $(N_p)^2 \gg 1$, and the quantity ϵ_{Sp} denotes the *spacing ratio* S/N_p.

The CRB from (6.30) is reported in Fig. 6.3 for $N_p = \{100, 200\}$[1] and $\epsilon_{Sp} \equiv S/N_p = \{0, 0.5, 1, 1.5, 2, 2.5\}$, along with the simulated MSE $E_{\mathbf{r}}\left[\left(\nu - \hat{\nu}\left(\mathbf{r}\right)\right)^2\right]$ resulting from the first line of (6.26). For the sake of accuracy, the size N of the FFT was taken relatively large ($N = 2^{19}$ for $N_p = 100$ and $N = 2^{20}$ for $N_p = 200$) and a minimum of 10^5 simulations was run. For each simulation a new normalized frequency offset ν was taken from a random uniform distribution over $[-0.1, 0.1]$. We observe that the MSE reaches the corresponding CRB at high SNR. For a given value of N_p, an increase in the spacing ratio ϵ_{Sp} decreases the CRB from (6.30), thus improving the accuracy of the DA frequency estimate from (6.26) at high SNR. The same is true for a given value of ϵ_{Sp} and an

[1]Later in this chapter we will consider bursts of length $N_s = 1000$ symbols. For these bursts $N_p = 100, 200$ is an acceptable value for the number of pilot symbols.

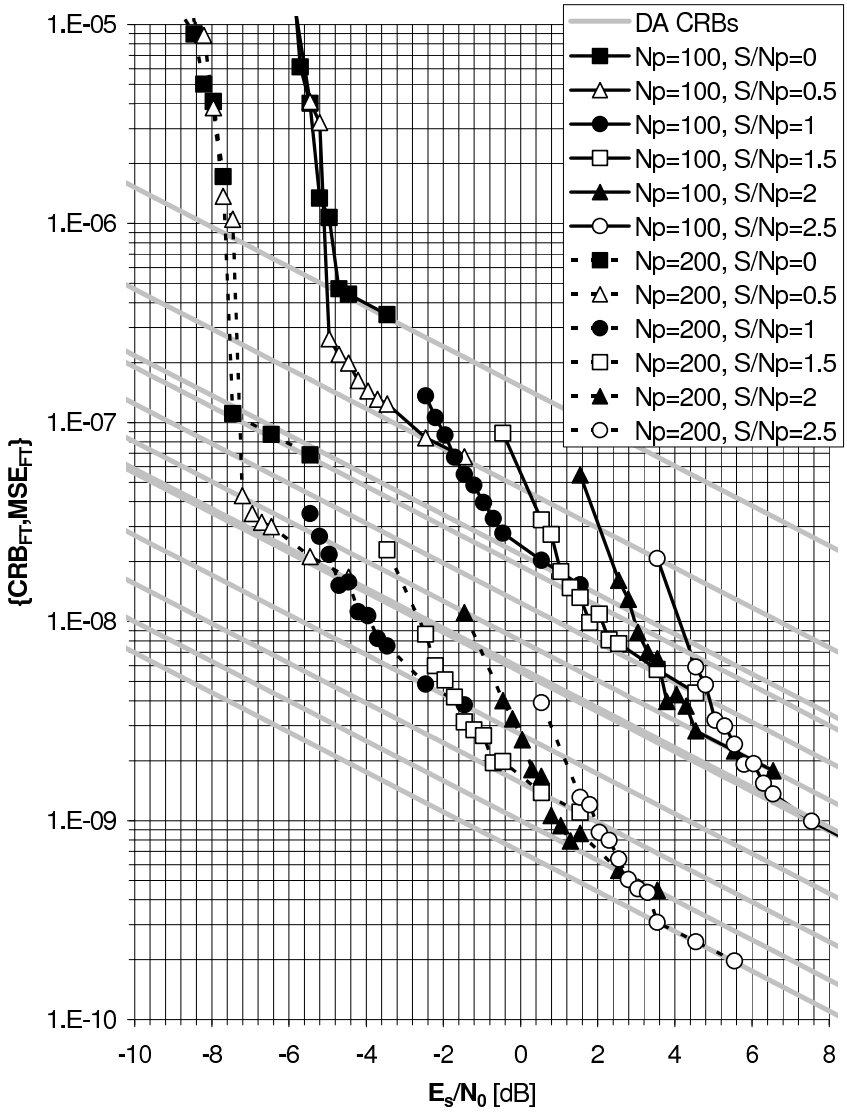

Figure 6.3: MSE related to the DA estimation of the normalized frequency offset $\nu = FT$.

increase in the number of pilot symbols N_p. As (6.30) is inversely proportional to the third power of N_p, increasing N_p by a factor 2 decreases the DA CRB by a factor $8 = \left(2^3\right)$. We also observe that, when the SNR is below a certain threshold value, the MSE performance dramatically degrades across a narrow SNR interval, yielding an MSE that is much larger than that predicted by the DA CRB. For a given value of N_p, the threshold is very low for $\epsilon_{Sp} = 0$, but tends to increase as the two pilot symbol blocks are separated by data symbols ($\epsilon_{Sp} > 0$). This imposes a practical limit on how much the two pilot symbol blocks can be separated. When the separation is very large, the SNR threshold is so large that an MSE close to the CRB is impossible except in virtually noise-free conditions. This is in spite of the fact that a larger separation reduces the CRB related to frequency estimation. Our results also indicate that, for a given value of ϵ_{Sp}, increasing N_p by a factor 2 decreases the threshold by approximately $3\,(= 10\log_{10} 2)$ dB.

These observations can be explained as follows [59, 60]. According to (6.29), the DA CRB_ν decreases when $\left|E_{\mathbf{r},\mathbf{\Psi}}\left[\partial^2 \ell_{DA}\left(\mathbf{\Psi};\mathbf{r}\right)/\partial\nu^2\right]\right|$ increases. Let $Y_0\left(\tilde{\nu}\right)$ denote the noiseless version of $Y_{DA}\left(\tilde{\nu};\mathbf{r}\right)$. In Appendix N on page 235 we point out that the curvature of $|Y_0\left(\tilde{\nu}\right)|$ at $\tilde{\nu} = \nu$ is proportional to $E_{\mathbf{r},\mathbf{\Psi}}[\partial^2 \ell_{DA}\left(\mathbf{\Psi};\mathbf{r}\right)/\partial\nu^2]$, and increases with increasing N_p and increasing S. The larger the curvature of $|Y_0\left(\tilde{\nu}\right)|$ at $\tilde{\nu} = \nu$, the more accurate the frequency estimate in the presence of (a small amount of) noise. The high-SNR behavior of the MSE related to frequency estimation is consistent with these observations. We also point out in Appendix N that the objective function $Y_{DA}\left(\tilde{\nu};\mathbf{r}\right)$ can be decomposed as the sum of $Y_0\left(\tilde{\nu}\right)$ and a zero-mean statistical fluctuation with a standard deviation of $\sigma_Y = \sqrt{N_p N_0/E_s}$. The maximum of $|Y_0\left(\tilde{\nu}\right)|$ is achieved at $\tilde{\nu} = \nu$. The maximum of $|Y_{DA}\left(\tilde{\nu};\mathbf{r}\right)|$ is usually close to $\tilde{\nu} = \nu$. Occasionally, however, $|Y_{DA}\left(\tilde{\nu};\mathbf{r}\right)|$ will be so much affected by noise that its highest peak will occur at a frequency that lies far from the true normalized frequency offset. When this happens, the DA frequency estimator (6.26) makes large errors, referred to as *outliers*. The occurrence of outliers has a deleterious effect on the corresponding MSE performance, which becomes much larger than the DA CRB_ν. Obviously, the probability of having an outlier increases when the fluctuation of $Y_{DA}\left(\tilde{\nu};\mathbf{r}\right)$ increases (this is characterized by an increasing value of σ_Y), or when, at a position far removed from the true normalized frequency offset, the difference between $|Y_0\left(\tilde{\nu}\right)|$ and its maximum $|Y_0\left(\nu\right)|$ decreases. We show in Appendix N that $|Y_0\left(\tilde{\nu}\right)|$ exhibits several secondary maxima. The larger the spacing ratio ϵ_{Sp}, the smaller the difference in magnitude between the main peak and the secondary peaks of $|Y_0\left(\tilde{\nu}\right)|$. For a given value of ϵ_{Sp}, the difference in magnitude between the main peak and the secondary peaks of $|Y_0\left(\tilde{\nu}\right)|$ is observed to be proportional to N_p. Taking into account that σ_Y is independent of S and proportional to $\sqrt{N_p}$ and $\sqrt{E_s/N_0}$, it may be concluded that the occurrence of outliers increases as E_s/N_0 decreases, ϵ_{Sp} increases, or N_p decreases. This is reflected in the value of the SNR threshold below which the MSE curves for DA frequency offset estimation start deviating from the corresponding CRBs in Fig. 6.3.

The simulated MSE $E\left[\left(\theta-\hat{\theta}\left(\mathbf{r}\right)\right)^2\right]$ resulting from the second line of (6.26) is reported in Fig. 6.4 for $N_p = 100$ and $\epsilon_{Sp} = \{0, 1, 2\}$. The corresponding DA CRB (6.31) is also shown. It is assumed that frequency offset estimation is performed prior to phase shift estimation and for this two scenarios are considered:

1. Perfect frequency offset estimation, i.e., $\hat{\nu}\left(\mathbf{r}\right) \equiv \nu$. It follows from (6.23), (6.11) and (6.1)-(6.2) that in this case the statistical properties of $\hat{\theta}\left(\mathbf{r}\right)$ depend on the value of N_p and not on the value of ϵ_{Sp}.
2. DA frequency offset estimation. In this case, the quantity $\hat{\nu}\left(\mathbf{r}\right)$ results from locating the maximum of $\left|Y_{DA}\left(\tilde{\nu};\mathbf{r}\right)\right|$ (see first line of (6.26)).

A minimum of 10^3 (case scenario 1) or 10^5 (case scenario 2) simulations have been run to ensure accuracy. In each simulation, a new normalized frequency offset ν and a new phase shift θ were taken from a random uniform distribution over $[-0.1, 0.1]$ and $[-\pi, \pi]$, respectively. The phase estimation error was measured modulo 2π, i.e., in the interval $[-\pi, \pi]$. We make the following observations. When $\hat{\nu}\left(\mathbf{r}\right) \equiv \nu$ (scenario 1), the MSE resulting from the second line of (6.26) (which does not depend on ϵ_{Sp}) closely approaches the DA CRB for the whole range of E_s/N_0 considered in Fig. 6.4. When $\hat{\nu}\left(\mathbf{r}\right)$ results from DA estimation (scenario 2), the MSE resulting from the second line of (6.26) only achieves the DA CRB when the DA frequency estimator operates above threshold (compare to Fig. 6.3). Below this threshold, the performance rapidly degrades across a narrow SNR interval, yielding a MSE that is much larger than the one predicted by the DA CRB. This shows that outlier normalized frequency offset estimates have a detrimental effect on the accuracy of the phase shift estimate resulting from the second line of (6.26).

6.5 NPA NCA Estimation: Kth-Power Approximation

6.5.1 Algorithm Derivation

Assuming equiprobable symbol vectors, i.e.,

$$\Pr\left[\mathbf{a}\right] \equiv M^{-N_s}, \quad \forall \mathbf{a} \in \Omega^{N_s}, \tag{6.32}$$

the log-likelihood function from (6.12) reads (within a constant not depending on $\mathbf{\Psi}$):

$$\ell_{NPA\,NCA}\left(\mathbf{\Psi};\mathbf{r}\right) = \ln \prod_{k\in I_s}\left(\sum_{\omega_i\in\Omega} e^{\left(\frac{E_s}{N_0}\Re\left\{2\omega_i^* z(k;r(k),\mathbf{\Psi})-|\omega_i|^2\right\}\right)}\right), \tag{6.33}$$

which will be referred to as the NPA NCA log-likelihood function. In arriving at (6.33) we have made use of (6.8) and (6.9).

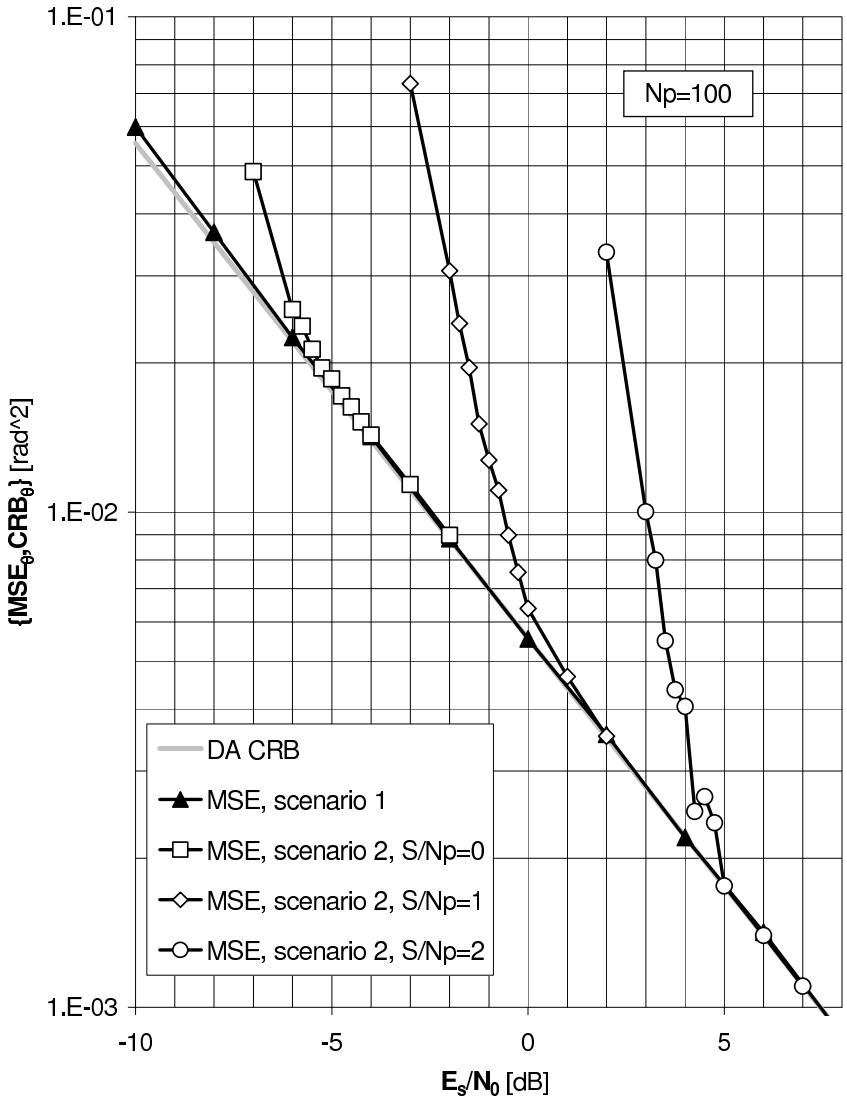

Figure 6.4: MSE related to the DA estimation of the phase shift θ at the center of the pilot part of the symbol burst ($N_p = 100$).

We obtain a low-SNR approximation of (6.33) by expanding the exponential function into a Taylor series, computing the sum over the set of constellation points, computing the product over the set of indices I_s and keeping only the (ν,θ)-dependent terms that correspond to the smallest powers of E_s/N_0. Taking into account the rotational symmetry of the constellation, the low-SNR limit of (6.33) is (within a factor not depending on (ν,θ)) found to be [61]:

$$\ell_{NPA\,NCA,0}\left(\mathbf{\Psi};\mathbf{r}\right)=\sum_{k\in I_s}\Re\left\{z^K\left(k;r\left(k\right),\mathbf{\Psi}\right)\right\}, \tag{6.34}$$

with K related to the symmetry angle $2\pi/K$ of the constellation (K=2 for M-PAM, K=4 for M-QAM and K=M for M-PSK). The procedure in which a maximum of $\ell_{NPA\,NCA,0}\left(\mathbf{\Psi};\mathbf{r}\right)$ is located is commonly referred to as *Kth-power estimation* [59,62,63].

Let us introduce the integer parts, k_ν and k_θ, and the fractional parts, δ_ν and δ_θ, of the normalized frequency offset ν and the phase shift θ, respectively, defined by [64]:

$$\nu=\delta_\nu+\frac{k_\nu}{K},$$
$$\theta=\delta_\theta+\frac{2\pi k_\theta}{K},$$

where $|\delta_\nu|<\frac{1}{2K}$, $|\delta_\theta|<\frac{\pi}{K}$ and $k_\nu,k_\theta\in\{0,1,...,K-1\}$. Then, the Kth-power estimator is to be interpreted as an estimate of the fractional parts δ_ν and δ_θ rather than the 'total' normalized frequency offset ν and the 'total' phase shift θ. The reason for this is the following. Due to the periodicity of the log-likelihood function (6.33) and its approximation (6.34) with respect to the frequency and phase variables, the Kth-power estimator exhibits a $\frac{1}{K}$ normalized frequency estimation ambiguity and a $\frac{2\pi}{K}$ phase shift estimation ambiguity. This can be explained by the fact that in the case of equiprobable symbol vectors, $(\mathbf{a},\nu,\theta)$ and $\left(\check{\mathbf{a}},\check{\nu},\check{\theta}\right)$ yield the same matched filter output samples $\left(r\left(k\right):\,k\in I_s\right)$ when:

$$\check{a}\left(k\right)=a\left(k\right)e^{-j\frac{2\pi}{K}(km+n)},\quad k\in I_s,$$
$$\check{\nu}=\nu+\frac{m}{K},$$
$$\check{\theta}=\theta+\frac{2\pi n}{K},$$

with m and n integer.

It is interesting to see that the implementation and operation of the 'fractional' Kth-power estimator is very similar to that of the DA estimator from section 6.4. Indeed, the joint maximum of (6.34) over ν and θ can be computed as:

$$\hat{\delta}_\nu\left(\mathbf{r}\right)=\arg\max_{\tilde{\nu}\in\left[-\frac{1}{2K},\frac{1}{2K}\right]}\left|Y_{K\text{th}-\text{power}}\left(\tilde{\nu};\mathbf{r}\right)\right|, \tag{6.35}$$

$$\hat{\delta}_\theta\left(\mathbf{r}\right)=\frac{1}{K}\arg\left\{Y_{K\text{th}-\text{power}}\left(\hat{\delta}_\nu\left(\mathbf{r}\right);\mathbf{r}\right)\right\}, \tag{6.36}$$

with

$$Y_{K\text{th}-\text{power}}(\nu;\mathbf{r}) = \sum_{k\in I_s} \check{z}^K(k;r(k),\nu) = \sum_{k\in I_s} r^K(k)\, e^{-j2\pi K\nu(k-\kappa_0)}. \quad (6.37)$$

Again, the search for the maximum of (6.35) could be circumvented by using other and more ad-hoc Kth-power frequency estimation methods (see for example [32]), but instead we chose to locate the maximum of $(|Y_{K\text{th}-\text{power}}(\nu_n;\mathbf{r})|: n = -\frac{N}{2},...,\frac{N}{2}-1)$ with $\nu_n = \frac{n}{KN}$ and N even. Taking into account the periodicity property of the DFT, it follows from (6.37) that $(|Y_{K\text{th}-\text{power}}(\nu_n;\mathbf{r})|: n = -\frac{N}{2},...,\frac{N}{2}-1)$ is the norm of the DFT of a vector with the following N entries: the N_s samples $\left\{r^K(k): k\in I_s\right\}$, followed by $N-N_s$ zeros.

The fractional Kth-power estimator (6.35)-(6.36) must be combined with an ambiguity resolution algorithm that produces estimates $\hat{k}_\nu(\mathbf{r})$ and $\hat{k}_\theta(\mathbf{r})$ of the integer parts k_ν and k_θ of the normalized frequency offset ν and the phase shift θ. A pragmatic approach to the frequency ambiguity consists of applying the constraint $|\nu| \ll \frac{1}{2K}$ to ν and skipping the frequency ambiguity resolution alltogether, i.e., putting $\hat{k}_\nu(\mathbf{r})$ identically equal to zero or, equivalently,

$$\hat{\nu}(\mathbf{r}) \equiv \hat{\delta}_\nu(\mathbf{r}), \quad (6.38)$$

where $\hat{\delta}_\nu(\mathbf{r})$ is the fractional Kth-power estimate resulting from (6.35). This severely limits the operating range of the Kth-power frequency estimator, but mostly avoids the occurrence of large estimation errors due to the unresolved frequency offset ambiguity. Given $\hat{\nu}(\mathbf{r})$, a conventional DA phase ambiguity resolution algorithm based upon the pilot symbols is [65]:

$$\hat{k}_\theta(\mathbf{r}) = \arg \max_{\tilde{k}\in\{0,1,...,K-1\}} \Re\left\{Y_{DA}(\hat{\nu}(\mathbf{r});\mathbf{r})\, e^{-j\hat{\delta}_\theta(\mathbf{r})} e^{-j\frac{2\pi\tilde{k}}{K}}\right\}, \quad (6.39)$$

where the function $Y_{DA}(.;\mathbf{r})$ is defined as in (6.23) and $\hat{\delta}_\theta(\mathbf{r})$ is the fractional Kth-power estimate resulting from (6.36). We obtain:

$$\hat{\theta}(\mathbf{r}) \equiv \hat{\delta}_\theta(\mathbf{r}) + \frac{2\pi}{K}\hat{k}_\theta(\mathbf{r}), \quad (6.40)$$

where $\hat{\delta}_\theta(\mathbf{r})$ is the fractional Kth-power estimate resulting from (6.36) and where $\hat{k}_\theta(\mathbf{r})$ is the estimate (6.39) of the discrete parameter k_θ.

The Kth-power estimator uses all the received matched filter output samples, but operates in the NPA NCA mode, i.e., ignoring all the available statistical information about the transmitted symbol vector. It is therefore intrinsically suboptimal. The MSEs resulting from (6.35) and (6.36) are lower-bounded by the NPA NCA CRB.

Remarks

1. Whereas in the case of M-PAM and M-QAM the MSE resulting from (6.35) and (6.36) is always substantially larger than the NPA NCA CRB,

it can be shown that for M-PSK constellations the MSEs for increasing E_s/N_0 converge to this CRB [62]. This indicates that for M-PSK constellations the Kth-power synchronizer is one of the best (most accurate) NPA NCA estimators for large SNR.

2. Sometimes the Kth-power algorithm is generalized, in the sense that the quantities $\check{z}^K(k; r(k), \nu)$ in (6.37) are replaced by

$$\mathcal{F}\left(\left|\check{z}\left(k; r\left(k\right), \nu\right)\right|\right) e^{j\left(K \cdot \arg\left\{\check{z}\left(k; r(k), \nu\right)\right\}\right)}$$

where $\mathcal{F}(.)$ is an arbitrary function. By selecting different $\mathcal{F}(.)$ we obtain a slightly different performance and the best alternative depends on the SNR and the number of constellation points. (See, e.g., [63] by Viterbi and Viterbi.)

3. Just like the DA estimator, the Kth-power estimator also suffers from a threshold effect: when the SNR is below a threshold value, the MSE is much larger than the NPA NCA CRB due to the occurrence of outliers. In general, the SNR threshold for the Kth-power NPA NCA estimator is much higher than for a DA estimator using the same number of observations. This is because the Kth-power nonlinearity operating on $\check{z}(k; r(k), \nu)$ increases the noise level. In order to improve the performance of the Kth-power synchronizer below the threshold, a two-stage coarse-fine DA/Kth-power estimator has been proposed in [60]. First a DA estimator is used to coarsely locate the normalized frequency offset (according to the first line of (6.26)), and then the Kth-power estimator (which is potentially more accurate than the DA synchronizer because it makes use of the energy of all received samples) attempts to improve the estimate within the uncertainty of the coarse estimator (according to (6.35)). In fact, the search range of the Kth-power estimator is restricted to the area around the peak of the DA log-likelihood function from (6.19). This excludes a large percentage of potential outliers from the search range of the Kth-power estimator. When we assume that the MSE of the initial DA normalized frequency offset estimate equals the DA CRB_ν related to (6.19), the uncertainty range can be determined as $\pm\alpha\sqrt{DACRB_\nu}$, where α should be carefully chosen. When the parameter α increases, the search region increases, as well as the probability of estimating an outlier normalized frequency offset, which may result in a degradation of the performance at low SNR (outlier effect). However, if α decreases, the search region decreases, as well as the probability of comprising the correct normalized frequency offset, which in turn may result in a degradation of the performance at high SNR. A major disadvantage of this two-stage DA/Kth-power algorithm is that it does not exploit the knowledge of the pilot symbols nor the structure of the channel code in the Kth-power fine estimation step. Therefore, its MSE is still lower-bounded by the NPA NCA CRB.

6.5.2 Numerical MSE Results and Discussion

Numerical results pertaining to the conventional Kth-power algorithm are shown in Fig. 6.5 and Fig. 6.6. The number of transmitted symbols N_s is 250 or 1000, the signaling constellation is 2-PSK (in which case K equals 2) or 4-PSK (in which case K equals 4). We choose $\mathcal{F}(x) = x^2$ and let the normalized time instant κ_0 correspond to the center of the symbol burst, i.e.,

$$\kappa_0 = \kappa_{m,s} \equiv \frac{1}{N_s} \sum_{k \in I_s} k. \tag{6.41}$$

The middle N_p=100 symbols are assumed to be known pilot symbols. Normalized frequency offset estimation is performed according to (6.35),(6.38) and prior to phase shift estimation, which in turn is performed according to (6.36),(6.39)-(6.40). The size N of the FFT is set to N = 2^{17} for $N_s = 250$ and N = 2^{19} for $N_s = 1000$. A total of 10^4 simulations has been run. In each simulation, a new normalized frequency offset ν and a new phase shift θ were taken from a random uniform distribution over $[-0.1, 0.1]$ and $[-\pi, \pi]$, respectively. Fig. 6.5 shows the resulting MSE for normalized frequency offset estimation. Fig. 6.6 corresponds to phase shift estimation. For the sake of comparison, the MSE resulting from (6.36) with $\hat{\nu}(\mathbf{r}) = \nu$ (i.e., perfect frequency offset estimation) is also shown. The bold grey curves represent the NPA NCA CRBs of interest. For the computation of these CRBs we refer the reader to chapter 5. We come to the following conclusions.

We can define an SNR threshold, above which the Kth-power algorithm achieves near optimal NPA NCA CRB performance for both normalized frequency offset and phase shift estimation. Below this threshold, the performance of the Kth-power frequency estimator dramatically degrades across a narrow SNR interval due to the insurgence of outliers. This also affects the performance of the phase shift estimator, yielding a MSE that is significantly larger than that of estimators with perfect frequency offset estimation (in dashed), operating very closely to the NPA NCA CRB. For a given value of N_s, the SNR threshold for the Kth-power estimator is much lower for 2-PSK than for 4-PSK (about 7.25 dB for $N_s = 250$, and about 7.75 dB for $N_s = 1000$). We also observe that, for a given value of $M = K$, the SNR threshold significantly decreases when the number of observed samples N_s quadruples from 250 to 1000 (about 3.25 dB for 2-PSK and $K = 2$, and about 2.75 dB for 4-PSK and $K = 4$).

These observations can be explained as follows. We point out in Appendix O on page 241 that the objective function $Y_{K\text{th}-\text{power}}(\tilde{\nu}; \mathbf{r})$ can be decomposed as the sum of $Y_0(\tilde{\nu})$ and a zero-mean statistical fluctuation with a standard deviation that increases with K and N_0/E_s, and is proportional to $\sqrt{N_s}$. We also show in Appendix O that $|Y_0(\tilde{\nu})|$ exhibits several secondary maxima. The difference in magnitude between the main peak and the secondary peaks of $|Y_0(\tilde{\nu})|$ is observed to be independent of K and N_0/E_s, and proportional to N_s. Taking this into account, it may be concluded that the occurrence of outliers increases as E_s/N_0 decreases, K increases, or N_s decreases. This is reflected

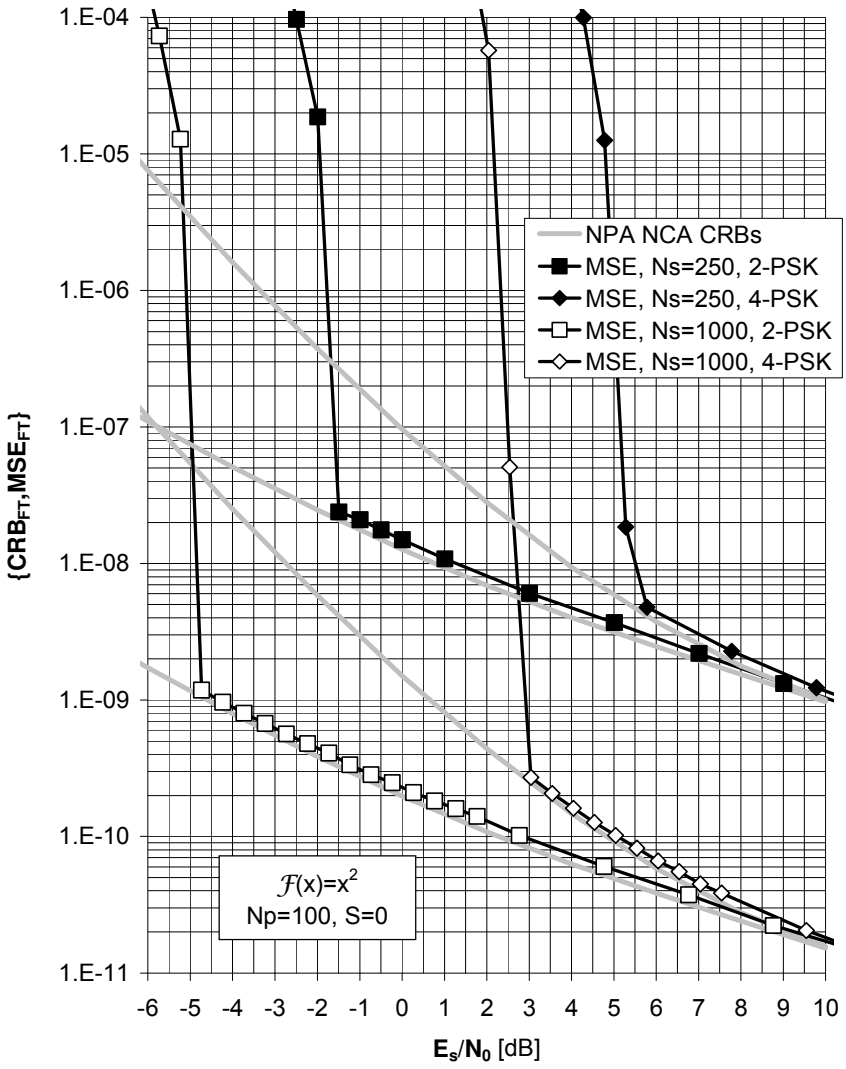

Figure 6.5: MSE for conventional Kth-power estimation of the normalized frequency offset $\nu = FT$ ($\mathcal{F}(x) = x^2$, $N_p = 100$, $S = 0$).

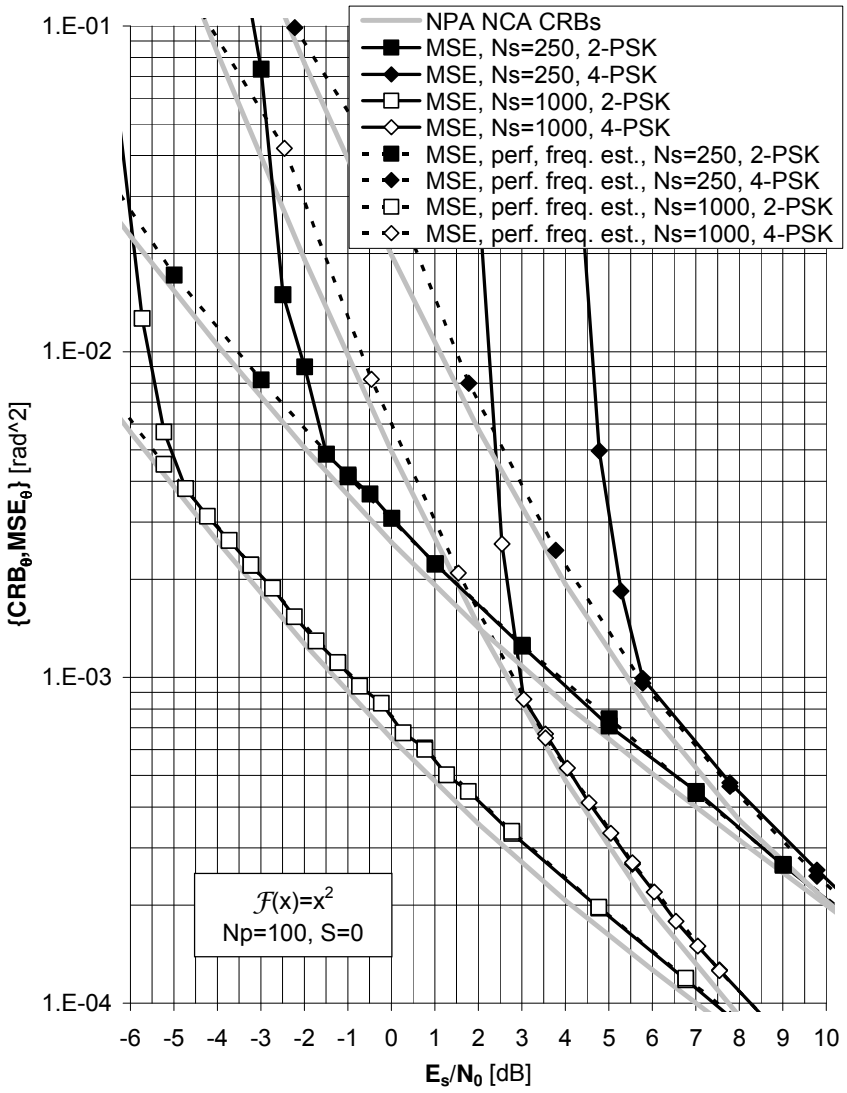

Figure 6.6: MSE for conventional Kth-power estimation of the phase shift θ at the center of the symbol burst ($\mathcal{F}(x) = x^2$, $N_p = 100$, $S = 0$).

in the value of the SNR threshold below which the MSE curves for Kth-power normalized frequency offset estimation start deviating from the corresponding CRBs in Fig. 6.5.

Fig. 6.7 and Fig. 6.8 illustrate the benefits of the two-stage DA/Kth-power frequency estimation procedure. The number of transmitted symbols N_s is 1000 and the signaling constellation is 4-PSK. Again, we choose $\mathcal{F}(x) = x^2$, $\kappa_0 = \kappa_{m,s}$ and $N_p = 100$. The parameter α, which determines the search range of the Kth-power fine normalized frequency offset estimation step is set equal to 3. In contrast to the previous set-up we also consider what happens when the pilot part of the symbol burst is divided into two blocks of $N_p/2 = 50$ consecutive symbols that are symmetrically located from the center of the symbol vector $\mathbf{a}$ and separated by $S > 0$ unknown data symbols.

At high SNR, the performance of the 2-stage DA/Kth-power frequency estimator is seen to match that of the conventional Kth-power frequency estimator. A threshold is still evident, but the performance below the SNR threshold degrades less rapidly than with the conventional Kth-power frequency estimator. The performance of the phase shift estimator improves correspondingly. The MSE of the 2-stage DA/Kth-power frequency estimator is lower when the two pilot symbol blocks are separated by $S = 100$ data symbols than when they are transmitted consecutively (case $S = 0$). This can be explained by the fact that for the values of the SNR being considered, the coarse DA normalized frequency offset estimation step (see Fig. 6.3):

1. performs close to the corresponding DA CRB for $S = 0$, 100;

2. is significantly more accurate for $S = 100$ than for $S = 0$.

6.6 PA CA and PA NCA Estimation

Both the DA and the (DA/)Kth-power algorithm are intrinsically suboptimal, in the sense that the resulting MSEs in many cases cannot approach the true (PA CA) CRB related to the actual log-likelihood function $\ell(\mathbf{\Psi}; \mathbf{r})$ from (6.12). DA estimation only allows to take benefit from the received signal energy associated with the pilot symbols, so that the resulting MSE performance is lower-bounded by the DA CRB. The Kth-power estimator drops all statistical information regarding the transmitted symbols, so that its MSE is lower-bounded by the NPA NCA CRB. In order to achieve a MSE close to the PA CA CRB, within the carrier estimation process the estimator should actually exploit both the knowledge of the pilot symbols and the structure of the channel code.

We will now discuss two different ways of deriving PA CA carrier synchronizers. First we look at the conventional *hard-decision-directed* (HDD) method, which uses a detected symbol vector $\hat{\mathbf{a}}$ as if it were the true symbol vector $\mathbf{a}$. Subsequently we will explore an alternative method that treats the maximiza-

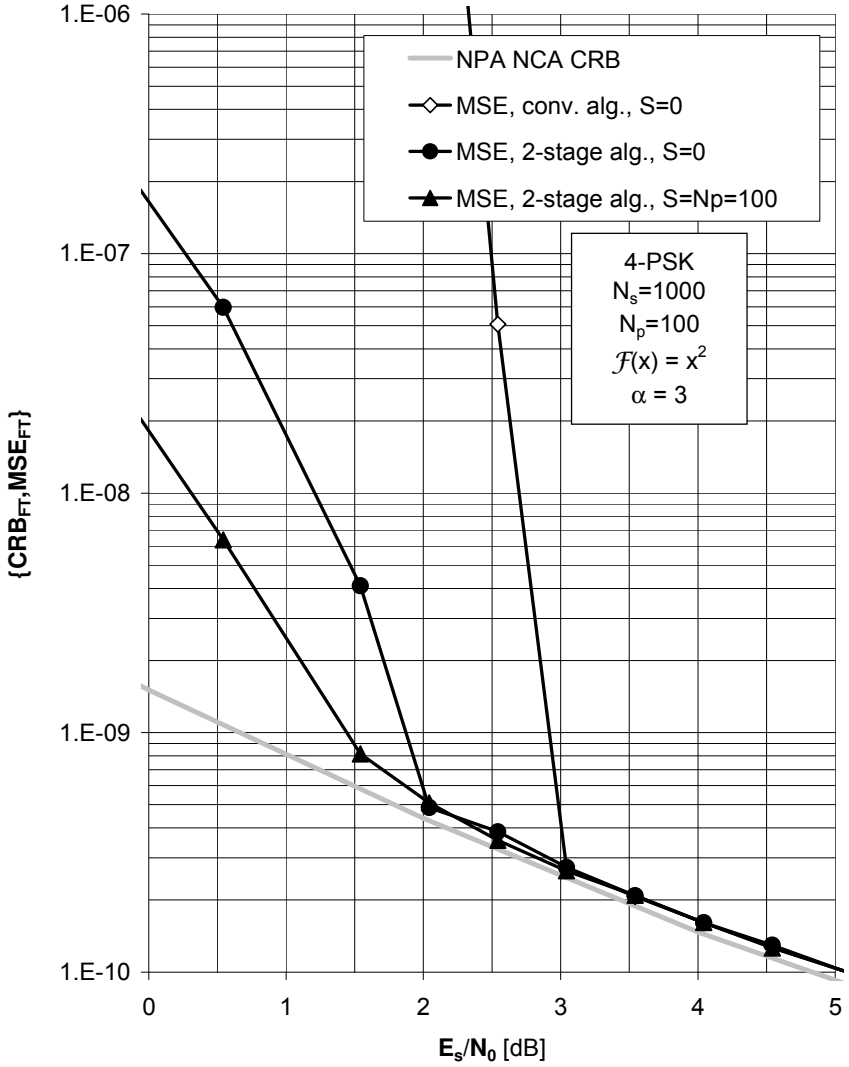

Figure 6.7: MSE for two-stage DA/Kth-power estimation of the normalized frequency offset $\nu = FT$ (4-PSK, $\alpha = 3$, $\mathcal{F}(x) = x^2$, $N_s = 1000$, $N_p = 100$).

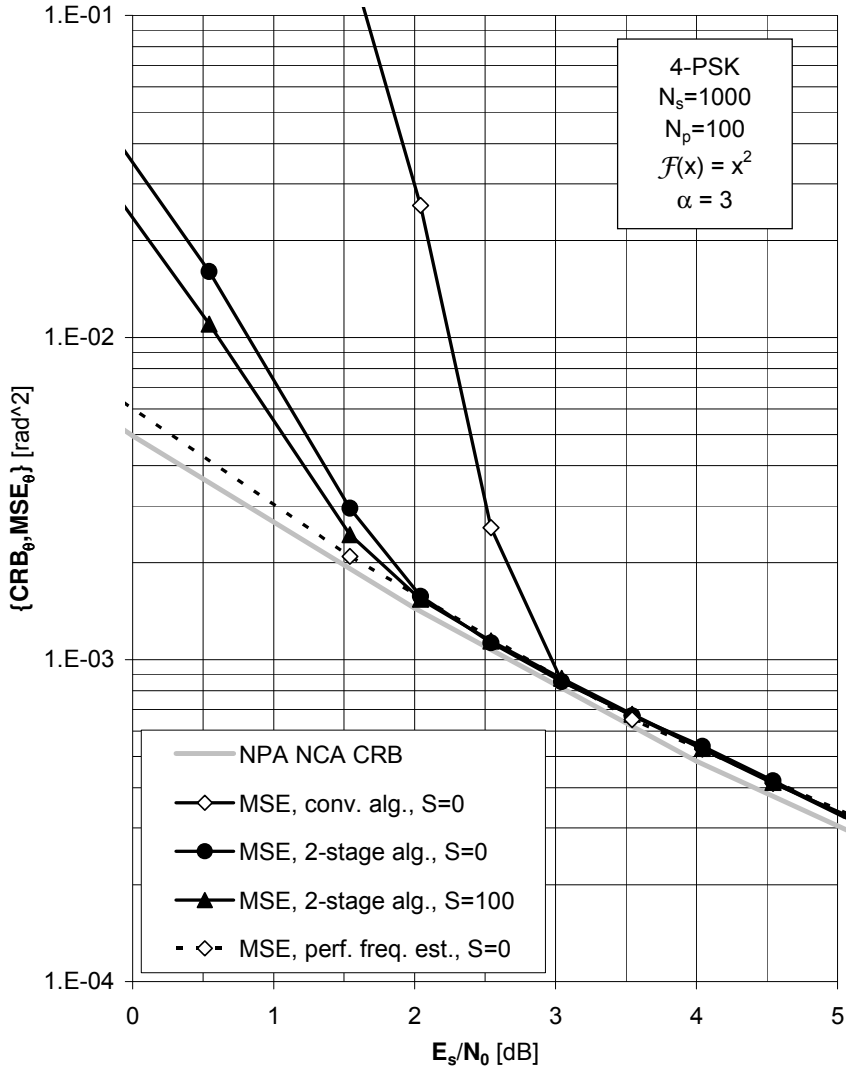

Figure 6.8: MSE for 2-stage DA/Kth-power estimation of the phase shift θ at the center of the symbol burst (4-PSK, $\alpha = 3$, $\mathcal{F}(x) = x^2$, $N_s = 1000$, $N_p = 100$).

tion of (6.12) by (iteratively) solving the set of ML equations:

$$\left.\frac{\partial \ell\left(\mathbf{\Psi};\mathbf{r}\right)}{\partial \Psi_i}\right|_{\mathbf{\Psi}=\hat{\mathbf{\Psi}}(\mathbf{r})} = 0, \qquad i = 1,2, \tag{6.42}$$

with $\Psi_1 = \nu$ and $\Psi_2 = \theta$. This will lead to a general theoretical framework for iterative synchronization structures that exchange (soft) information between the MAP bit detector (decoder) and the synchronization parameter estimator. From this theoretical basis, we derive a new algorithm that yields a more elegant way of extending the DA algorithm than the conventional HDD technique. We will denote this algorithm as *soft-decision-directed* (SDD).

6.6.1 Hard-Decision-Directed Approximation

The conventional HDD synchronizers use hard symbol decisions, computed with the help of the decoder/detector. Some examples of PA CA HDD synchronization techniques can be found in [66, 67].

Let us denote by $\hat{\mathbf{a}}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right) = \left(\hat{a}\left(k;\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right) : k \in I_s\right)$ an estimate of the symbol vector $\mathbf{a}$, produced by the detector, based on the observation vector $\mathbf{r}$ and a previous estimate $\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)$ of $\mathbf{\Psi}$. When the probability that $\hat{\mathbf{a}}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right)$ equals the true symbol vector $\mathbf{a}$, is high, the a posteriori probability (APP) $\Pr\left[\mathbf{a}|\,\mathbf{r}\right]$ of $\mathbf{a}$, given $\mathbf{r}$, can be very well approximated as follows:

$$\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}|\,\mathbf{r}\right] \approx \begin{cases} 1, & \tilde{\mathbf{a}} = \hat{\mathbf{a}}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right) \\ 0, & \text{otherwise} \end{cases}. \tag{6.43}$$

(See also the first part of Appendix E on page 203.) Algorithms that are based on (6.43) are referred to as HDD.

Using (6.43), we obtain for the APP $p\left(\mathbf{\Psi}\,|\mathbf{r}\right)$ of the carrier synchronization parameter vector $\mathbf{\Psi}$, given $\mathbf{r}$:

$$\begin{aligned} p\left(\mathbf{\Psi}\,|\mathbf{r}\right) &= \sum_{\tilde{\mathbf{a}}} p\left(\mathbf{a}=\tilde{\mathbf{a}},\mathbf{\Psi}\,|\mathbf{r}\right), \\ &= \sum_{\tilde{\mathbf{a}}} p\left(\mathbf{\Psi}\,|\mathbf{r},\mathbf{a}=\tilde{\mathbf{a}}\right)\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}|\,\mathbf{r}\right], \\ &\approx p\left(\mathbf{\Psi}\,\middle|\,\mathbf{r},\mathbf{a}=\hat{\mathbf{a}}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right)\right) \end{aligned}$$

and because of the uniform a priori distribution of $\mathbf{\Psi}$, the following holds:

$$\begin{aligned} p\left(\mathbf{r}\,|\mathbf{\Psi}\right) &\propto p\left(\mathbf{\Psi}\,|\mathbf{r}\right), \\ &\approx p\left(\mathbf{\Psi}\,\middle|\,\mathbf{r},\mathbf{a}=\hat{\mathbf{a}}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right)\right), \\ &\propto p\left(\mathbf{r}\,\middle|\,\mathbf{a}=\hat{\mathbf{a}}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right),\mathbf{\Psi}\right), \end{aligned} \tag{6.44}$$

where $\propto$ denotes equality up to a factor not depending on $\mathbf{\Psi}$ and $p\left(\mathbf{r}\,|\mathbf{a},\mathbf{\Psi}\right)$ is the conditional probability density function of $\mathbf{r}$ given $(\mathbf{a},\mathbf{\Psi})$ from (6.8). Taking the

logarithm of (6.44) yields the HDD approximation of the log-likelihood function $\ell(\mathbf{\Psi};\mathbf{r})$, which we will denote as $\ell_{HDD}(\mathbf{\Psi};\mathbf{r})$. Using (6.8)- (6.9), we obtain:

$$\begin{aligned}
&\ell(\mathbf{\Psi};\mathbf{r}) \approx \ell_{HDD}(\mathbf{\Psi};\mathbf{r}) \\
&= \ln p\left(\mathbf{r}\left|\mathbf{a}=\hat{\mathbf{a}}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r})\right),\mathbf{\Psi}\right.\right), \\
&= \ln \prod_{k\in I_s} p\left(r(k)|\, a(k)=\hat{a}\left(k;\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r})\right),\mathbf{\Psi}\right), \\
&\propto 2\frac{E_s}{N_0}\sum_{k\in I_s}\Re\left\{\hat{a}\left(k;\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r})\right)^* z(k;r(k),\mathbf{\Psi})\right\}. \qquad (6.45)
\end{aligned}$$

The HDD ML-based carrier parameter estimates (6.7) are determined by maximizing the right-hand side of (6.45) over $\mathbf{\Psi}$. This two-dimensional maximum can be computed as follows:

$$\begin{cases} \hat{\nu}(\mathbf{r}) = \arg\max_{\tilde{\nu}} \left|\sum_{k\in I_s}\hat{a}^*\left(k;\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r})\right)\check{z}(k;r(k),\tilde{\nu})\right|, \\ \hat{\theta}(\mathbf{r}) = \arg\left\{\sum_{k\in I_s}\hat{a}^*\left(k;\mathbf{r},\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r})\right)\check{z}(k;r(k),\hat{\nu}(\mathbf{r}))\right\}. \end{cases} \qquad (6.46)$$

As we see, the implementation and operation of the HDD version of the ML-based estimator is very similar to that of the DA ML estimator from section 6.4. This is because the functions that have to be maximized, in both case have essentially the same structure ((6.46) versus (6.26),(6.23)).

It is now possible to use an iterative method. Starting from an initial estimate of the carrier parameters, the detector makes a decision about the transmitted symbols. This decision is then used to compute a new estimate of the carrier parameters and so forth. To obtain a reliable initial estimate one may use the DA algorithm, the Kth-power or the two-stage DA/Kth-power algorithm.

We note that, in the case of the iterative method, the convergence of the HDD synchronizer is usually critical at low SNR: detected data symbols are unreliable due to the low SNR and erroneous data symbol values are fed back into the synchronizer, which fails to provide accurate estimates of the synchronization parameters, thus making the subsequently detected data symbols even worse.

6.6.2 Soft-Decision-Directed Approximation: Turbo Synchronization

Instead of exploiting hard decisions on the transmitted data symbols (which implies a loss of information about the decision reliability), the synchronizer can also be fed with so-called *soft* symbol decisions carrying both the symbol decision and its reliability. A number of examples can be found in [68, 69]. This approach is referred to as *turbo synchronization* since, in agreement with the standard turbo decoding algorithm, it is based on the exchange of soft (rather than hard) information. Earlier approaches for turbo synchronization aimed at plugging the soft-outputs of an iterative decoder into the structure of a conventional decision-directed synchronizer [68,69]. Although somewhat ad-hoc,

these approaches led to considerable performance improvement as compared to conventional synchronization methods. What was still missing, however, was a general framework to justify the architecture of these synchronizers and to suggest the systematic development of new algorithms.

Below we will give a mathematical interpretation of turbo synchronization. The basic idea is to consider turbo synchronization as an iterative solution to the ML synchronization problem. Subsection 6.6.2.1 will first provide a general formulation of iterative ML estimation of the unknown carrier parameters in the presence of unknown data symbols. The practical implementation of the resulting SDD estimator will then be addressed in subsection 6.6.2.3.

6.6.2.1 General Idea

In order to find the ML solution, let us solve the ML equations (6.42), or, in vector notation:

$$\frac{\partial \ell\left(\mathbf{\Psi};\mathbf{r}\right)}{\partial \mathbf{\Psi}} = \mathbf{0}, \tag{6.47}$$

where $\frac{\partial .}{\partial \mathbf{\Psi}} = \left(\frac{\partial .}{\partial \Psi_1} \; \frac{\partial .}{\partial \Psi_2}\right)^T = \left(\frac{\partial .}{\partial \nu} \; \frac{\partial .}{\partial \theta}\right)^T$ and $\mathbf{0} = (0\,0)^T$.

In section 5.7.4, it was demonstrated that, due to the particular structure of the digital data-modulated signal, the derivatives $\partial \ell\left(\mathbf{\Psi};\mathbf{r}\right)/\partial \Psi_i$, $i = 1, 2$ can be expressed as follows:

$$\frac{\partial \ell\left(\mathbf{\Psi};\mathbf{r}\right)}{\partial \Psi_i} = \frac{2E_s}{N_0} \sum_{k \in I_s} \Re\left\{ \mu^*\left(k; \mathbf{z}\left(\mathbf{r},\mathbf{\Psi}\right)\right) \frac{\partial z\left(k; r\left(k\right),\mathbf{\Psi}\right)}{\partial \Psi_i} \right\}, \tag{6.48}$$

where

$$\mu\left(k; \mathbf{z}\left(\mathbf{r},\mathbf{\Psi}\right)\right) = \sum_{l=0}^{M-1} \Pr\left[a\left(k\right) = \omega_l \left| \mathbf{z}\left(\mathbf{r},\mathbf{\Psi}\right)\right.\right] \omega_l \tag{6.49}$$

denotes the a posteriori expectation of the kth symbol $a\left(k\right)$ conditioned on $\mathbf{r}$ and $\mathbf{\Psi}$, with

$$\Pr\left[a\left(k\right) = \omega_l \left| \mathbf{z}\left(\mathbf{r},\mathbf{\Psi}\right)\right.\right] = \Pr\left[a\left(k\right) = \omega_l \left| \mathbf{r},\mathbf{\Psi}\right.\right] \tag{6.50}$$

the marginal symbol APPs of $a\left(k\right)$, given $\mathbf{r}$ and $\mathbf{\Psi}$. In most practical scenarios, the marginal symbol APPs from (6.50), for $k \in I_s$ and $l = 0, 1, ..., M-1$, can be directly computed in an efficient way by applying the sum-product algorithm to a factor graph representing a suitable factorization of the joint APP $p\left(\mathbf{a},\mathbf{b}\left|\mathbf{r},\mathbf{\Psi}\right.\right)$. This yields a computational complexity for evaluating (6.48) that increases linearly with the number of transmitted symbols N_s. In fact, receivers that perform maximum-a-posteriori bit detection according to the sum-product algorithm (e.g. turbo or LDPC receivers, see chapter 4) are at the same time fully-equipped to compute the symbol APPs involved in (6.49). More details can be found in chapter 4.

It remains, however, that finding the solution of (6.47) is not trivial, since $\mathbf{\Psi}$ appears in both factors of the summation in (6.48). We therefore try an iterative method that produces a sequence of values $\hat{\mathbf{\Psi}}^{(i)}\left(\mathbf{r}\right) = \left(\hat{\nu}^{(i)}\left(\mathbf{r}\right) \hat{\theta}^{(i)}\left(\mathbf{r}\right)\right)^T$, which

will hopefully converge to the desired solution. In particular, we use the previous estimate $\hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})$ to evaluate the first factor of the summation in (6.48), and we find the current estimate $\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})$ by solving the resulting simplified equation that follows:

$$\sum_{k\in I_s} \Re\left\{ \mu^*\left(k; \mathbf{z}\left(\mathbf{r}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right)\right) \left.\frac{\partial z\left(k; r\left(k\right), \mathbf{\Psi}\right)}{\partial \mathbf{\Psi}}\right|_{\mathbf{\Psi}=\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})} \right\} = \mathbf{0}. \tag{6.51}$$

If the sequence of estimates $\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})$ resulting from (6.51) converges to a finite value, that value will provide a solution of the ML equation (6.47) [70].

Attention is drawn to the fact that the first factor between the braces in (6.51) does not depend on $\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})$. We can therefore bring the derivative back out of the braces and obtain the equivalent equation:

$$\frac{\partial}{\partial \mathbf{\Psi}} \sum_{k\in I_s} \Re\left\{ \mu^*\left(k; \mathbf{z}\left(\mathbf{r}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right)\right) z\left(k; r\left(k\right), \mathbf{\Psi}\right)\right\}\Bigg|_{\mathbf{\Psi}=\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})} = \mathbf{0}. \tag{6.52}$$

Consequently, the new estimate $\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})$ at the ith iteration is determined by the following maximization with respect to $\mathbf{\Psi}$:

$$\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r}) = \arg\max_{\mathbf{\Psi}} \Lambda\left(\mathbf{\Psi}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right), \tag{6.53}$$

$$\begin{aligned} &\Lambda\left(\mathbf{\Psi}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right) \\ &= \sum_{k\in I_s} \Re\left\{ \mu^*\left(k; \mathbf{z}\left(\mathbf{r}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right)\right) z\left(k; r\left(k\right), \mathbf{\Psi}\right)\right\}. \end{aligned} \tag{6.54}$$

Taking into account (6.10), the corresponding solution is:

$$\hat{\nu}^{(i)}(\mathbf{r}) = \arg\max_{\tilde{\nu}} \left| \sum_{k\in I_s} \mu^*\left(k; \mathbf{z}\left(\mathbf{r}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right)\right) \check{z}\left(k; r\left(k\right), \tilde{\nu}\right)\right|, \tag{6.55}$$

$$\hat{\theta}^{(i)}(\mathbf{r}) = \arg\left\{ \sum_{k\in I_s} \mu^*\left(k; \mathbf{z}\left(\mathbf{r}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right)\right) \check{z}\left(k; r\left(k\right), \hat{\nu}^{(i)}(\mathbf{r})\right)\right\}. \tag{6.56}$$

The obtained solution describes an iterative synchronization procedure that can be referred to as PA CA SDD. What we call *soft decisions* here, are in fact the a posteriori symbol average values $\mu\left(k; \mathbf{z}\left(\mathbf{r}, \hat{\mathbf{\Psi}}^{(i-1)}(\mathbf{r})\right)\right)$ of each channel symbol. By analogy with the estimation of continuous-valued parameters, the conditional mean $\mu\left(k; \mathbf{z}\left(\mathbf{r}, \mathbf{\Psi}\right)\right)$ can be considered as the *soft* (since taking values in some continuous result space) *minimum mean square error decision* regarding $a\left(k\right)$ based upon the observed samples $\mathbf{r}$ and a trial value of the carrier synchronization parameter vector $\mathbf{\Psi}$ (see section 3.4.1).

6.6.2.2 Expectation-Maximization Interpretation

Formulation (6.55)-(6.56) can also be derived by means of the *expectation-maximization* (EM) algorithm [70,71]. Consider $\mathbf{r}$ as the *incomplete* observation and $\mathbf{q} \equiv \left(\mathbf{r}^T \mathbf{a}^T\right)^T$ as the complete observation. The EM algorithm states that the sequence $\left\{\hat{\boldsymbol{\Psi}}^{(i)}(\mathbf{r})\right\}$ defined by:

1. expectation step (E-step):

$$Q\left(\boldsymbol{\Psi}, \hat{\boldsymbol{\Psi}}^{(i-1)}(\mathbf{r})\right) = E_{\mathbf{a}}\left[\ln p\left(\mathbf{q}\,|\boldsymbol{\Psi}\right)\Big|\mathbf{r}, \hat{\boldsymbol{\Psi}}^{(i-1)}(\mathbf{r})\right], \tag{6.57}$$

2. maximization step (M-step):

$$\hat{\boldsymbol{\Psi}}^{(i)}(\mathbf{r}) = \arg\max_{\boldsymbol{\Psi}} Q\left(\boldsymbol{\Psi}, \hat{\boldsymbol{\Psi}}^{(i-1)}(\mathbf{r})\right), \tag{6.58}$$

converges to the ML estimate, provided that the initial estimate $\hat{\boldsymbol{\Psi}}^{(0)}(\mathbf{r})$ is sufficiently close to the global maximum of the log-likelihood function [72]. Otherwise, convergence to a different stationary point could occur. To make (6.57)-(6.58) equivalent to (6.55)-(6.56), we observe that, by using Bayes' rule and in view of the fact that the distribution of $\mathbf{a}$ does not depend on the parameter vector to be estimated,

$$\begin{aligned} p\left(\mathbf{q}\,|\boldsymbol{\Psi}\right) &= p\left(\mathbf{r}, \mathbf{a}\,|\boldsymbol{\Psi}\right) = p\left(\mathbf{r}\,|\mathbf{a}, \boldsymbol{\Psi}\right)\Pr\left[\mathbf{a}\,|\boldsymbol{\Psi}\right], \\ &= p\left(\mathbf{r}\,|\mathbf{a}, \boldsymbol{\Psi}\right)\Pr\left[\mathbf{a}\right]. \end{aligned} \tag{6.59}$$

Therefore, substituting (6.59) in (6.57), we get:

$$\begin{aligned} Q\left(\boldsymbol{\Psi}, \hat{\boldsymbol{\Psi}}^{(i-1)}(\mathbf{r})\right) &= \sum_{\mathbf{a}} \Pr\left[\mathbf{a}\Big|\mathbf{r}, \hat{\boldsymbol{\Psi}}^{(i-1)}(\mathbf{r})\right] \ln p\left(\mathbf{r}\,|\mathbf{a}, \boldsymbol{\Psi}\right) \\ &\quad + \underbrace{\sum_{\mathbf{a}} \Pr\left[\mathbf{a}\Big|\mathbf{r}, \hat{\boldsymbol{\Psi}}^{(i-1)}(\mathbf{r})\right] \ln \Pr\left[\mathbf{a}\right]}_{\tau}. \end{aligned} \tag{6.60}$$

The second term τ in (6.60) does not depend on $\boldsymbol{\Psi}$, and as far as the M-step is concerned, it can be dropped. Consequently, it follows from:

$$\ln p\left(\mathbf{r}\,|\mathbf{a}, \boldsymbol{\Psi}\right) \propto \frac{E_s}{N_0} 2\Re\left\{\mathbf{z}\left(\mathbf{r}, \boldsymbol{\Psi}\right)\mathbf{a}^H\right\},$$

that the estimation procedure given by (6.55)-(6.56) and the EM algorithm, defined by (6.60)-(6.58), are in fact equivalent and yield the same sequence of estimates.

It should be noted that the EM approach was one of the first attempts to provide a general theoretical framework for turbo synchronization [5]. Since then, other frameworks based on gradient methods [73, 74] or the sum-product algorithm [75] have been proposed in the literature.

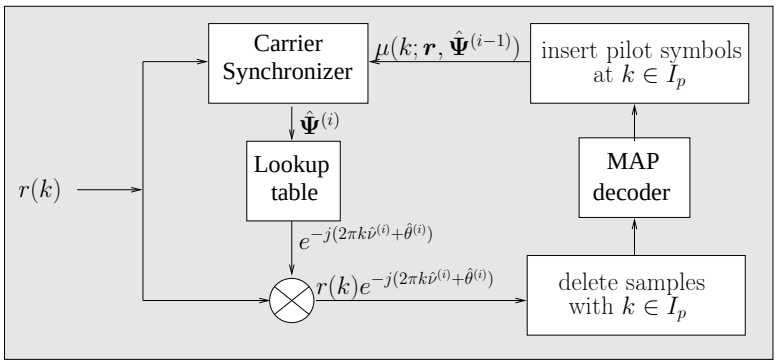

Figure 6.9: Interaction between a MAP decoder and the iterative CA synchronizer.

6.6.2.3 Practical Implementation

As shown above, the estimation of the synchronization parameter vector $\boldsymbol{\Psi}$ via the EM algorithm requires only the knowledge of marginal symbol APPs from (6.50), for $k \in I_s$ and $l = 0, 1, ..., M-1$. From this, it follows that ML synchronizers operating according to the EM algorithm and MAP bit detectors operating according to the sum-product algorithm are in fact complementary, since the marginal probabilities needed by the former can be provided by the latter. The interaction between a MAP decoder and the iterative CA synchronizer is shown in Fig. 6.9.

A major drawback of the proposed estimation algorithm is its computational overhead: for each estimator iteration the marginal APPs for all data symbols need to be computed. When the factor graph of $\Pr[\mathbf{b}\,|\mathbf{r}, \boldsymbol{\Psi}]$ contains cycles, the computation of these APPs according to the sum-product algorithm is in itself an iterative process (as happens with turbo codes and LDPC codes). In principle, for each estimator iteration we should therefore perform enough decoder iterations to achieve convergence of the APPs. This leads to the joint synchronization and detection algorithm described in Algorithm 10. The parameter N denotes the number of performed EM iterations. Step 2a means that the state information in the factor graph is reinitialized to zero at each EM iteration. Step 2b means that the sum-product algorithm will iterate until the marginals reach a steady-state value. These last operations enable to assume that the information provided by the sum-product algorithm will be a good approximation of the true marginal symbol APPs. Step 2c is simply performed by solving (6.55) and (6.56).

To reduce the computational complexity, we resort to the approximate implementation described in Algorithm 11 [5, 15, 68]: after each update of $\hat{\boldsymbol{\Psi}}(\mathbf{r})$, we perform only a single sum-product iteration, but we maintain the state information within the factor graph from one estimator iteration to the next.

Algorithm 10 EM algorithm

1. $\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r}) = \mathbf{\Psi_0}$;
2. For $n = 1, 2, ..., N$:
   (a) reset factor graph state information;
   (b) repeat:
   perform 1 sum-product iteration;
   until:
   marginal APP convergence;
   (c) compute $\hat{\mathbf{\Psi}}^{(n)}(\mathbf{r})$ according to (6.55) and (6.56).

Algorithm 11 EM algorithm approximation

1. $\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r}) = \mathbf{\Psi_0}$;
2. For $n = 1, 2, ..., N$:
   (a) perform 1 sum-product iteration;
   (b) compute $\hat{\mathbf{\Psi}}^{(n)}(\mathbf{r})$ according to (6.55) and (6.56).

Estimator and sum-product iterations are therefore intertwined, which drastically reduces the computational overhead related to the iterative estimation process (6.55)-(6.56). The only extra computational load with respect to a perfectly synchronized MAP bit detector is then the evaluation of equations (6.55) and (6.56). This leads to a very slight complexity increase in comparison with the computational requirements of the iterative detection/decoding algorithms implemented in the receiver.

Note that, strictly speaking, Algorithm 11 is no longer an EM algorithm. Indeed, performing only one turbo iteration at each EM iteration may lead to poor approximations of marginal symbol APPs, especially for the first EM iterations. Moreover, non-resetting the factor graph state information implies that the information computed by the sum-product algorithm will be a function of all the previous estimates $\hat{\mathbf{\Psi}}^{(n-1)}(\mathbf{r})$, $\hat{\mathbf{\Psi}}^{(n-2)}(\mathbf{r})$, ..., $\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r})$ and not solely of $\hat{\mathbf{\Psi}}^{(n-1)}(\mathbf{r})$ as required by the EM algorithm.

Consequences of both these approximations have been studied through simulation results in [15] for the particular case of phase shift estimation in a bit-interleaved coded-modulation system with iterative demodulation and decoding at the receiver. These simulations show that considering more than one internal decoder iteration per synchronizer iteration does not help the overall decoding process, and that the system reaches global (i.e., detection/ synchronization) ML performance after convergence. In particular, it was observed that

Algorithm 10 and 11 exhibit the same performance at EM iteration 10, while Algorithm 10 enables faster convergence in term of EM iterations than Algorithm 11. The latter observation may be explained by the crude approximation made on the posterior probabilities in Algorithm 11. Indeed, the soft information provided by the decoder in the first iterations is a rough approximation of the true marginal APPs. The phase estimates computed with the Algorithm 11 therefore rely in the first iterations on soft information that is less reliable than with Algorithm 10. Note however that, even if more EM steps are needed, Algorithm 11 reaches the same performance as Algorithm 10 at considerably reduced complexity and delay.

Algorithm 11 actually corresponds to the one introduced in an ad hoc fashion by Lottici [68] for the particular case of carrier recovery in a turbo-coded scheme. We provided a mathematical interpretation of Lottici's algorithm by showing that it can actually be regarded as an EM algorithm approximation.

6.6.2.4 Special Cases, Variations and Related Algorithms

1. When the symbol $a(k)$ is a known pilot symbol, the corresponding marginal symbol APP $\Pr[a(k)|\mathbf{z}(\mathbf{r},\mathbf{\Psi})]$ reverts to a Kronecker distribution, yielding $\mu(k;\mathbf{z}(\mathbf{r},\mathbf{\Psi})) = a(k)$. As a result, the SDD algorithm can be seen as an elegant way of extending the DA estimation approach (see also [76]).

2. When all data symbol sequences are equiprobable (this is equivalent to uncoded transmission), the marginal APPs of the data symbols $a(k)$, $k \in I_d$ reduce to:

$$\begin{aligned}\Pr[a(k)|\mathbf{z}(\mathbf{r},\mathbf{\Psi})] &= \Pr[a(k)|\mathbf{r},\mathbf{\Psi}], \\ &= \frac{p(r(k)|a(k),\mathbf{\Psi})}{\sum_{a(k)} p(r(k)|a(k),\mathbf{\Psi})}, \\ &= \Pr[a(k)|r(k),\mathbf{\Psi}], \\ &= \Pr[a(k)|z(k;r(k),\mathbf{\Psi})], \end{aligned} \tag{6.61}$$

   because $p(r(k)|a(k),\mathbf{\Psi})$ from (6.9) depends on $r(k)$ and $\mathbf{\Psi}$ only through $z(k;r(k),\mathbf{\Psi})$. As (6.61) depends only on $z(k;r(k),\mathbf{\Psi})$ (instead of all $\mathbf{z}(\mathbf{r},\mathbf{\Psi})$), we will denote the corresponding symbol decisions as

$$\mu(k;z(k;r(k),\mathbf{\Psi})).$$

   When the data symbol sequences are not equiprobable (as is the case with coded transmission) and we were to replace the marginal symbol APPs $\Pr[a(k)|\mathbf{z}(\mathbf{r},\mathbf{\Psi})]$ in (6.49) by $\Pr[a(k)|z(k;r(k),\mathbf{\Psi})]$ we end up with an iterative estimator which does not take into account the code properties. This synchronization algorithm is denoted PA NCA SDD.

3. If we replace in (6.51) $\mu\left(k;\mathbf{z}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})\right)\right)$ by:

$$\hat{a}\left(k;\mathbf{z}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})\right)\right) = \arg\max_{\omega\in\Omega} \Pr\left[a(k)=\omega \left| \mathbf{z}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(i)}(\mathbf{r})\right)\right.\right], \tag{6.62}$$

we make hard instead of soft decisions with respect to the transmitted symbols. This results in a conventional HDD estimator (see section 6.6.1). Such estimators are quite popular and provide a convenient (though ad-hoc) way of exploiting code-properties. Of course, at sufficiently high SNR, $\Pr\left[a\left(k\right)=\omega\left|\mathbf{z}\left(\mathbf{r},\hat{\mathbf{\Psi}}^{(i)}\left(\mathbf{r}\right)\right)\right.\right]$ will tend to a Kronecker distribution, so that (6.62) and (6.49) will yield the same performance. We denote the estimator resulting from (6.62) by PA CA HDD.

4. The HDD algorithm also exists in a non-code-aided version, which computes hard symbol decision according to:

$$\begin{aligned}&\hat{a}\left(k;z\left(k;r\left(k\right),\hat{\mathbf{\Psi}}^{(i)}\left(\mathbf{r}\right)\right)\right)\\&=\arg\max_{\omega\in\Omega}\Pr\left[a\left(k\right)=\omega\left|z\left(k;r\left(k\right),\hat{\mathbf{\Psi}}^{(i)}\left(\mathbf{r}\right)\right)\right.\right].\end{aligned}\tag{6.63}$$

We denote this estimator as PA NCA HDD.

6.6.2.5 Properties and Features

- The SDD and HDD estimators take advantage of both the (N_p) pilot symbols and the (N_d) data symbols. The MSE of the PA CA SDD and the PA CA HDD estimates is lower-bounded by the true PA CA CRB. The MSE of the PA NCA SDD and the PA NCA HDD estimates is lower-bounded by the PA NCA CRB.

- The normal operating SNR of the receiver is usually above the threshold (caused by outlier estimates) of the SDD and the HDD estimators. This is because a larger amount of samples is used than with the DA method ($N_s \gg N_p$), and because there is no noise enhancement as opposed to the Kth-power method.

- In the absence of pilot symbols the PA NCA SDD and the PA NCA HDD estimators exhibit a $\frac{1}{K}$ normalized frequency offset estimation ambiguity and a $\frac{2\pi}{K}$ phase shift estimation ambiguity, where K is related to the symmetry angle $\frac{2\pi}{K}$ of the constellation (K=2 for M-PAM, K=4 for M-QAM and K=M for M-PSK). These ambiguities can be prevented by using a sufficiently accurate initial estimate $\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)=\left(\hat{\nu}^{(0)}\left(\mathbf{r}\right)\hat{\theta}^{(0)}\left(\mathbf{r}\right)\right)^{T}$.

- In the presence of coding and/or pilot symbols a rotated sequence may no longer be a legitimate one and consequently there might be fewer frequency and phase ambiguities with CA estimators than with NCA estimators.

- The convergence of the iterative joint decision-directed (DD) synchronization and detection processes towards low values of the MSE and the BER requires a sufficiently accurate initial estimate $\hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)$. The range within which the initial estimation errors must lie is commonly referred to as the

acquisition range. The acquisition range for phase shift estimation generally amounts to about half the symmetry angle of the constellation, but the acquisition range for normalized frequency offset estimation is often quite small (see e.g., [5, 76]). This can be easily explained if we consider the following fact. For a given symbol vector $\mathbf{a}$ and a given initial estimate $\hat{\mathbf{\Psi}}^{(0)}$, the samples $z\left(k; r\left(k\right), \hat{\mathbf{\Psi}}^{(0)}\right)$, $k \in I_s$ can be decomposed as:

$$z\left(k; r\left(k\right), \hat{\mathbf{\Psi}}^{(0)}\left(\mathbf{r}\right)\right) \quad = \quad a\left(k\right) e^{-j e_{\Theta}^{(0)}(kT)} + n\left(k\right), \tag{6.64}$$

with

$$e_{\Theta}^{(0)}\left(kT\right) = 2\pi\left(\nu - \hat{\nu}^{(0)}\left(\mathbf{r}\right)\right)\left(k - \kappa_0\right) + \left(\theta - \hat{\theta}^{(0)}\left(\mathbf{r}\right)\right)$$

and $\{n\left(k\right)\}$ AWGN samples. The normalized frequency offset error causes a constant-speed phase rotation of the samples (6.64). When $(\nu\text{-}\hat{\nu}^{(0)}\left(\mathbf{r}\right))$ exceeds some threshold value, the resulting phase increase over the observation interval is too large for the sum-product algorithm to produce reliable symbol APPs, even after several iterations. This hinders joint convergence of the normalized frequency offset estimate and the symbol decisions. The actual limits on the acquisition range depend on the burst size, the SNR, the pilot symbols and the channel code, whereas the probability that the initial estimation errors fall within these limits depends on the algorithm that produces $\hat{\mathbf{\Psi}}^{(0)}$.

- The convergence speed of the MSE is also an important aspect of the iterative joint DD synchronization and detection techniques. In [76] it was shown that both the normalized frequency offset acquisition range and the convergence speed of the PA CA SDD estimator (6.55)-(6.56) can be increased by including pilot symbols in the transmitted burst.

6.6.3 Numerical Results and Discussion

We now derive performance results for a specific channel-coded system. We take 4-PSK as the signal constellation, and choose for the channel code a rate-1/2 turbo code. The turbo encoder encompasses the parallel concatenation of two identical binary 16-state rate-1/2 recursive systematic convolutional encoders with generator polynomials $(21)_8$ and $(37)_8$ in octal notation, via a pseudo random interleaver with block length $N_b = 900$ information bits, and an appropriate puncturing pattern so that the block at the turbo encoder output comprises 1800 coded bits. This binary turbo code is followed by conventional Gray-mapped 4-PSK modulation, giving rise to a block of $N_d = 900$ random data symbols. Finally, $N_p = 100$ pilot symbols are added, yielding a transmitted burst of $N_s = 1000$ symbols and a pilot symbol ratio of $\lambda_p = 10\%$. We further assume that the pilot symbols are organized into two blocks of $N_p/2 = 50$ consecutive pilot symbols with $S = \epsilon_{Sp} N_p$ data symbols in between, and symmetrically positioned about the center of this burst. The average energy per information bit E_b is

related to the average symbol energy E_s by

$$\frac{E_b}{N_0} = \frac{1}{1-\lambda_p}\frac{E_s}{N_0},$$

or, in dB,

$$\left.\frac{E_b}{N_0}\right|_{dB} \approx \left.\frac{E_s}{N_0}\right|_{dB} + 0.46\,\text{dB}.$$

The presence of the pilot symbols causes a power loss of 0.46 dB as compared to a system without pilot symbols.

The BER in the case of perfect carrier synchronization, after i iterations of the associated turbo detector, is shown in Fig. 6.10 as a function of E_b/N_0. The dashed curve corresponds to the BER after 30 iterations for a perfectly synchronized system without pilot symbols. These results are obtained from simulations performed until at least 100 frame errors are counted. We observe that the decoder needs 10 to 15 iterations to converge, and that the normal operating SNR of the system (say, corresponding to a BER of less than 10^{-3}) is situated above $E_b/N_0 \approx 1.75$ dB, or equivalently, at

$$E_s/N_0 \geq 1.3\,\text{dB}.$$

The case of synchronized detection will be considered next. In order to obtain reliable statistical results simulations will be performed until at least 10^5 bursts are transmitted and at least 200 frame errors are counted. For each burst a new normalized frequency offset ν and a new phase shift θ are taken from a random uniform distribution over $[-0.1, 0.1]$ and $[-\pi, \pi]$, respectively. The normalized time instant κ_0 is assumed to correspond to the center of the symbol burst, which, because of the symmetric burst structure, also coincides with the center of the pilot part of the symbol burst:

$$\kappa_0 = \kappa_{m,s} \equiv \frac{1}{N_s}\sum_{k\in I_s} k = \kappa_{m,p} \equiv \frac{1}{N_p}\sum_{k\in I_p} k.$$

The initial estimate $\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r}) = \left(\hat{\nu}^{(0)}(\mathbf{r})\,\hat{\theta}^{(0)}(\mathbf{r})\right)^T$ required to start the iterations of the DD synchronizers is obtained from the DA or the two-stage DA/4th-power method. These schemes will be referred to as 'DA init.' and 'DA/4th-power init.', respectively. In the latter case, the parameter α, which determines the search range of the 4th-power fine frequency offset estimation step is set equal to 3. The accuracy of $\hat{\mathbf{\Psi}}^{(0)}(\mathbf{r})$ can be read from Fig. 6.3 and Fig. 6.4 for the DA method, and from Fig. 6.7 and Fig. 6.8 for the DA/4th-power method. To ensure that the (coarse) DA frequency offset estimation during the initialization step performs close to the corresponding DA CRB, we impose the following upper bound on the separation S between the two pilot symbol blocks (see Fig. 6.3):

$$S \leq 100.$$

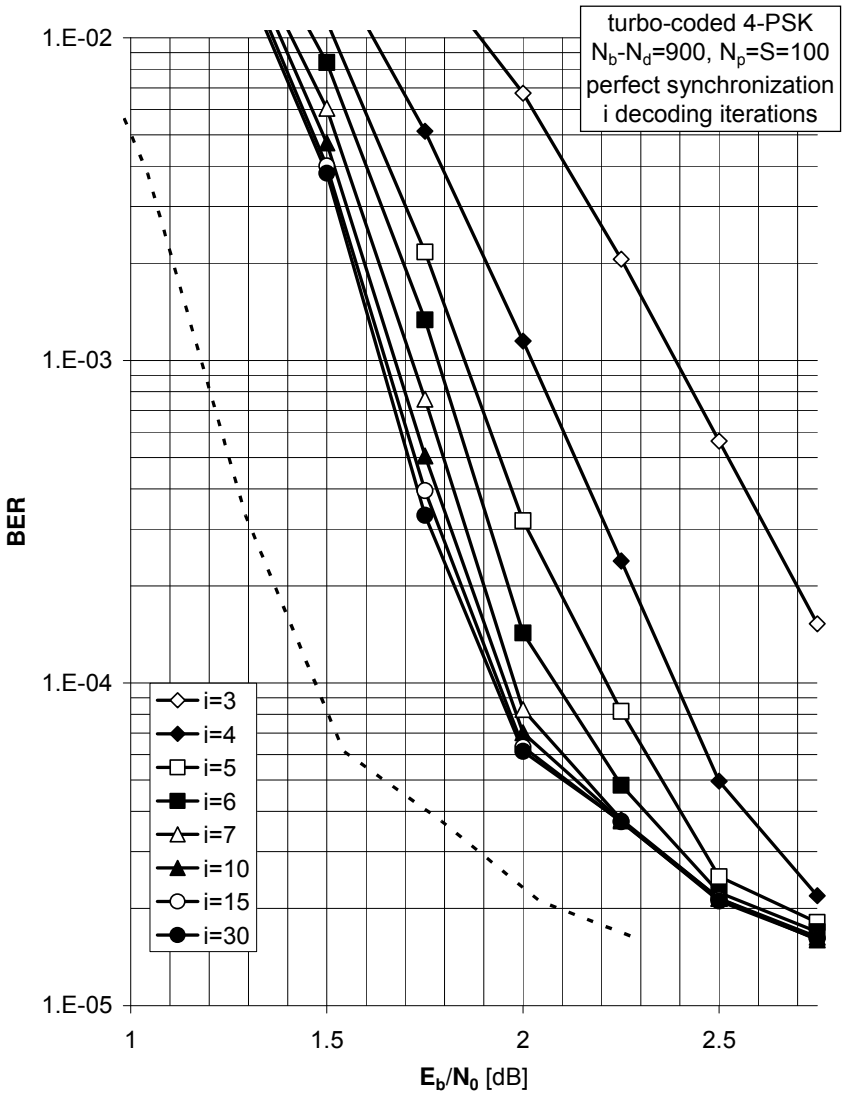

Figure 6.10: BER resulting from perfectly synchronized turbo decoder with $(M, N_b, N_d, N_p) = (4, 900, 900, 100)$, after i=1,2,...,30 decoding iterations.

We first consider PA NCA DD carrier parameter estimation. Fig. 6.11 shows, for $E_b/N_0 = 2$ dB or, equivalently, $E_s/N_0 \approx 1.54$ dB, the MSEs

$$E_{\mathbf{r}}\left[\left(\nu - \hat{\nu}^{(i)}\left(\mathbf{r}\right)\right)^2\right]$$

and

$$E_{\mathbf{r}}\left[\left(\theta - \hat{\theta}^{(i)}\left(\mathbf{r}\right)\right)^2\right],$$

resulting from the PA NCA SDD and the PA NCA HDD estimators, after DA initialization and i synchronizer iterations ($i = 0$ to $i = 50$). Two values for S are considered, namely $S = 50$ and $S = 100$. Iteration $i = 0$ corresponds to the DA initialization step. Several observations can be made about this figure:

- The larger separation of $S = 100$ data symbols yields the same initial phase estimation accuracy but a higher initial frequency estimation accuracy than the smaller separation of $S = 50$ data symbols.

- The MSEs resulting from the PA NCA HDD and the PA NCA SDD estimators converge to a lower steady-state value when the separation S between the pilot symbol blocks increases. This indicates the importance of an accurate initial frequency estimate $\hat{\nu}^{(0)}\left(\mathbf{r}\right)$ in order to achieve good steady-state estimator performance. The more accurate $\hat{\nu}^{(0)}\left(\mathbf{r}\right)$, the higher the probability that $\left(\nu - \hat{\nu}^{(0)}\left(\mathbf{r}\right)\right)$ will fall within the limits of the normalized frequency offset acquisition range and that joint convergence to the correct data symbol and carrier parameter values will occur.

- For a given initial estimation accuracy (given S), the use of hard decisions instead of soft decisions speeds up the convergence but leads to a significant degradation of the steady-state MSE.

Plots not reported here, similar to Fig. 6.11 but for other values of E_b/N_0, indicate that the above observations hold for all $E_b/N_0 \geq 0.5\,\text{dB}$. Convergence is seen to happen faster when the SNR increases, which is due to an increased accuracy of the initial estimates.

Finally, Fig. 6.12 and Fig. 6.13 compare the MSEs after 50 synchronizer iterations with the associated PA NCA CRBs as a function of E_s/N_0. For the estimation of θ the PA NCA CRB does not depend on S. For the estimation of ν the PA NCA CRBs for $S = 50$ and $S = 100$ virtually coincide for the considered range of E_s/N_0. We observe that, for a given value of S, the MSEs resulting from the PA NCA SDD and the PA NCA HDD estimator converge to the PA NCA CRBs when the SNR E_s/N_0 is sufficiently high.

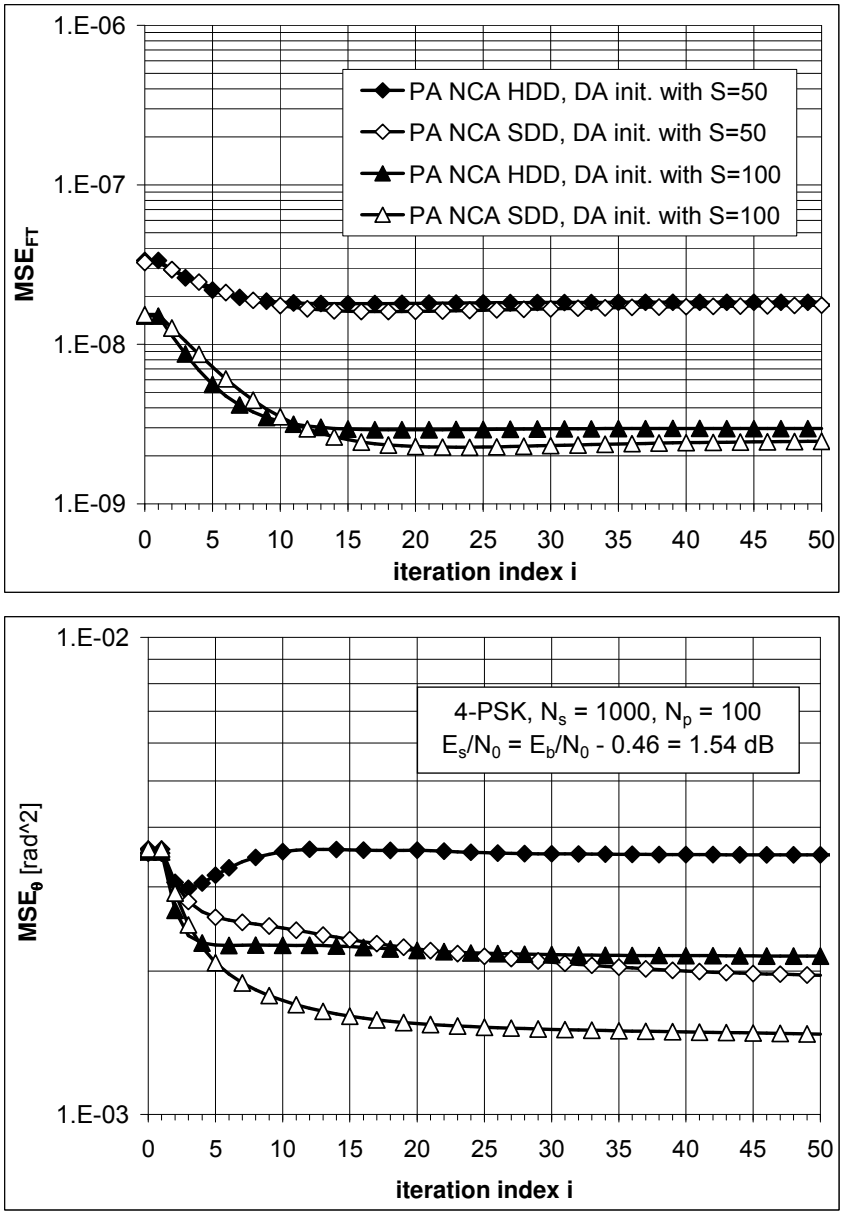

Figure 6.11: MSE resulting from PA NCA DD carrier parameter estimation, with DA initialization, at $E_s/N_0 = E_b/N_0$ - 0.46 = 1.54 dB (4-PSK, N_s = 1000, $N_p = 100$).

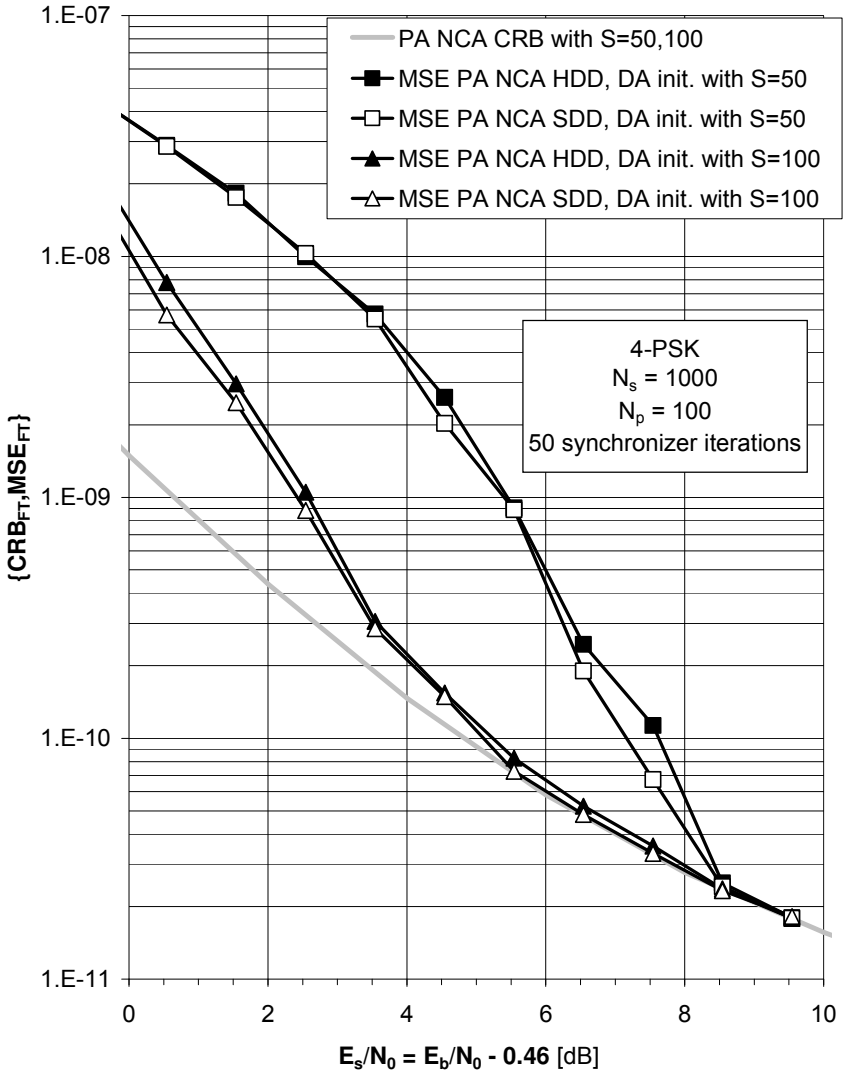

Figure 6.12: MSE for the iterative PA NCA DD estimation of the normalized frequency offset $\nu = FT$, as obtained with DA initialization followed by 50 estimation iterations (4-PSK, $N_s = 1000$, $N_p = 100$).

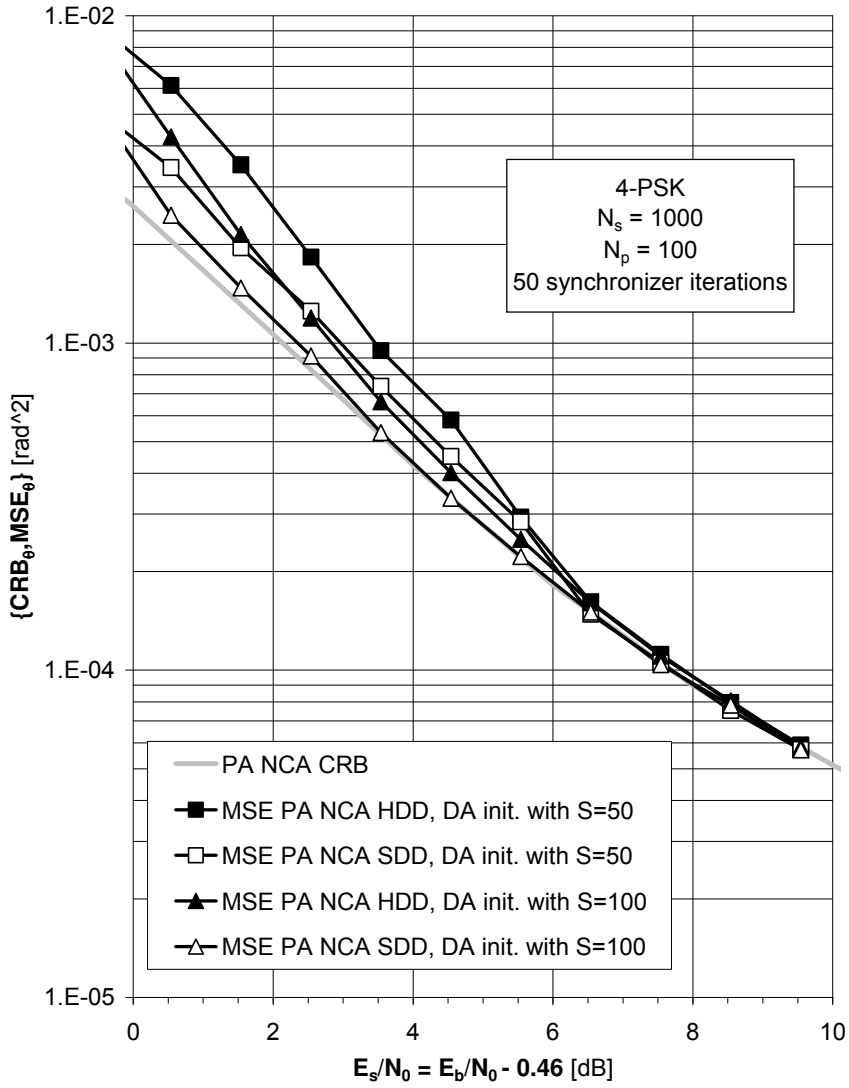

Figure 6.13: MSE for the iterative PA NCA estimation of the phase shift θ at the center of the symbol burst, as obtained with DA initialization followed by 50 estimation iterations (4-PSK, $N_s = 1000$, $N_p = 100$).

Remark

At this point, we have presented simulation results for the following five techniques:

- DA,
- 4th-power,
- DA/4th-power,
- PA NCA HDD,
- PA NCA SDD.

All these techniques perform carrier synchronization prior to and independently of the decoding. This implies that the derived MSEs do not depend on the particular channel encoding and mapping rules that are being considered. It also implies that the BER performance and the overall computational complexity of systems applying these techniques can be meaningfully compared based *solely* on the MSE performance and the computational complexity of the synchronization unit alone.

We will now investigate the performance and the complexity of carrier synchronizers that apply these techniques: the above list gives the carrier parameter estimation techniques in order of increasing complexity and Fig. 6.14 shows the associated MSEs, zooming in on the normal operating range of the turbo-coded system at hand. It follows that only two techniques merit further consideration:

1. DA,
2. DA/4th-power.

The first because it exhibits the lowest computational complexity. The second because it yields the highest frequency offset estimation accuracy and the second highest phase shift estimation accuracy.

The PA NCA SDD technique yields the highest phase shift estimation accuracy but is not retained. This is because the computational complexity of the iterative PA NCA SDD technique is several times larger than that of the two-stage DA/4th-power technique, while the average resulting instantaneous phase shift error:

$$E_{\mathbf{r}}\left[\frac{1}{N_s}\sum_{k\in I_s}\left(\Theta\left(kT;\mathbf{\Psi}\right)-\Theta\left(kT;\hat{\mathbf{\Psi}}^{(\infty)}\left(\mathbf{r}\right)\right)\right)^2\right]$$

is not significantly smaller.

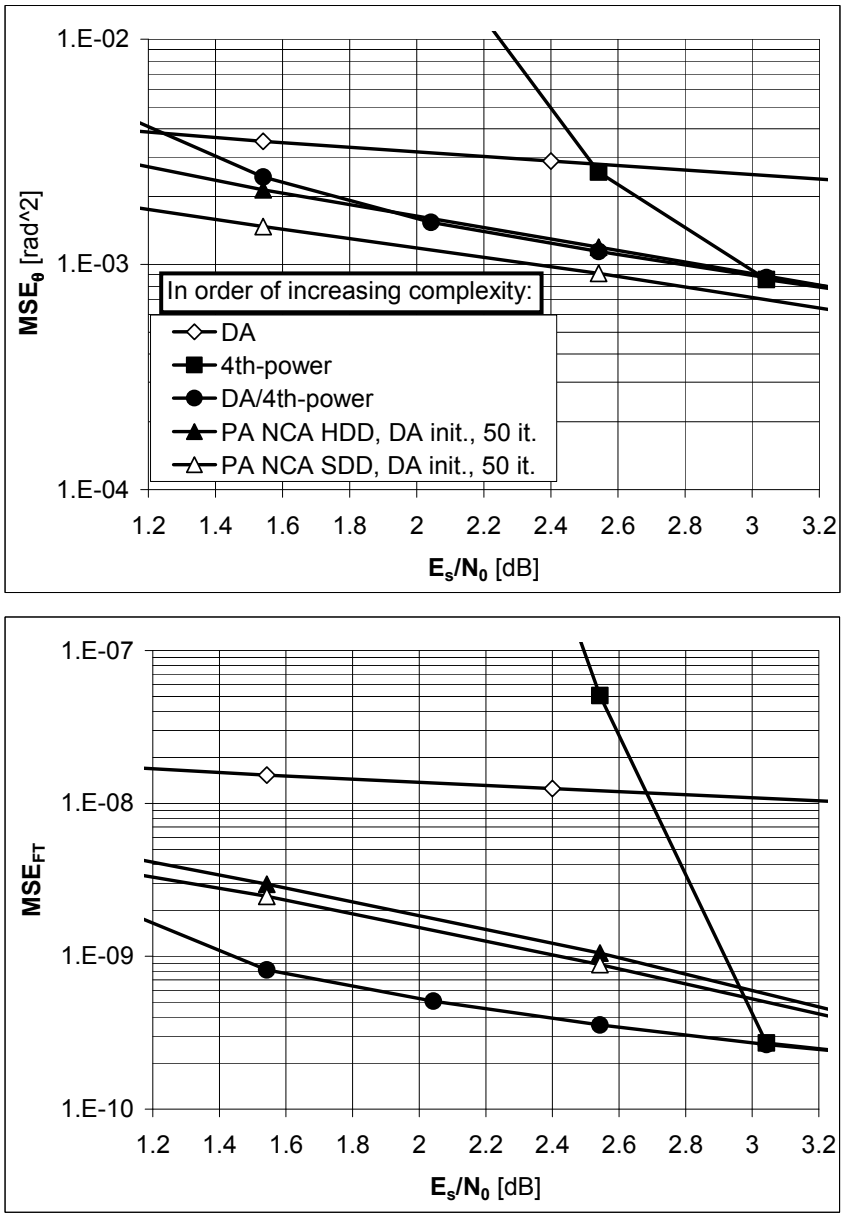

Figure 6.14: MSE for the DA, the 4th-power, the DA/4th-power, the PA NCA HDD and the PA NCA SDD estimation of the carrier parameters θ and $\nu = FT$ (4-PSK, $N_s = 1000$, $N_p = 100$). In the case of PA NCA HDD and PA NCA SDD estimation, 50 iterations are performed and DA initialization is used.

Let us now discuss CA estimation. In this case, carrier synchronization is performed jointly with the decoding. The resulting MSE performance therefore depends on the specific channel encoding and mapping rules that are being used. Fig. 6.15 and Fig. 6.16 show, for $E_b/N_0 = 2.25$ dB or, equivalently, $E_s/N_0 \approx 1.8$ dB, the MSEs

$$E_{\mathbf{r}}\left[\left(\nu - \hat{\nu}^{(i)}\left(\mathbf{r}\right)\right)^2\right], \; i = 0, 1, 2, ...29$$

and

$$E_{\mathbf{r}}\left[\left(\theta - \hat{\theta}^{(i)}\left(\mathbf{r}\right)\right)^2\right], \; i = 0, 1, 2, ...29$$

and the BER

$$E_{\mathbf{r}}\left[\frac{1}{N_b}\sum_{k=0}^{N_b-1}\left(b\left(k\right) - \hat{b}^{(i)}\left(k;\mathbf{r}\right)\right)^2\right], \; i = 1, 2, ..., 30,$$

resulting from a system with DA or two-stage DA/4th-power initial carrier parameter estimation followed by i joint PA CA SDD or PA CA HDD synchronizer/decoder iterations. The separation is fixed at $S = 100$. The MSE values for $i = 0$ result from the initial DA or DA/4th-power estimates that are used to start the iterations. At the given SNR, two-stage DA/4th-power initialization yields a significantly higher estimation accuracy than plain DA initialization. The BER after i conventional turbo decoder iterations in the case of perfect, DA and DA/4th-power carrier synchronization is also shown. The following observations can be made about these figures:

- The BER of the perfectly synchronized system and the BER of the systems with DA or DA/4th-power carrier synchronization exhibit about the same convergence speed: more or less 10 iterations are required to achieve convergence.
- The systems with CA synchronization require a considerably higher number of iterations to converge: up to 30 iterations or more.
- The simplest system, i.e., the one with code-unaware DA synchronization, exhibits a very large BER degradation as compared to the case of perfect synchronization: the obtained BER is no less than about 540 times as large.
- The most accurate NCA algorithm, i.e., NPA NCA two-stage DA/4th-power synchronization, gives rise to a BER that is still about 3.2 times as high as in the case of perfect synchronization.
- The CA estimators converge to significantly lower steady-state MSE and BER values with DA/4th-power initialization than with the DA initialization.

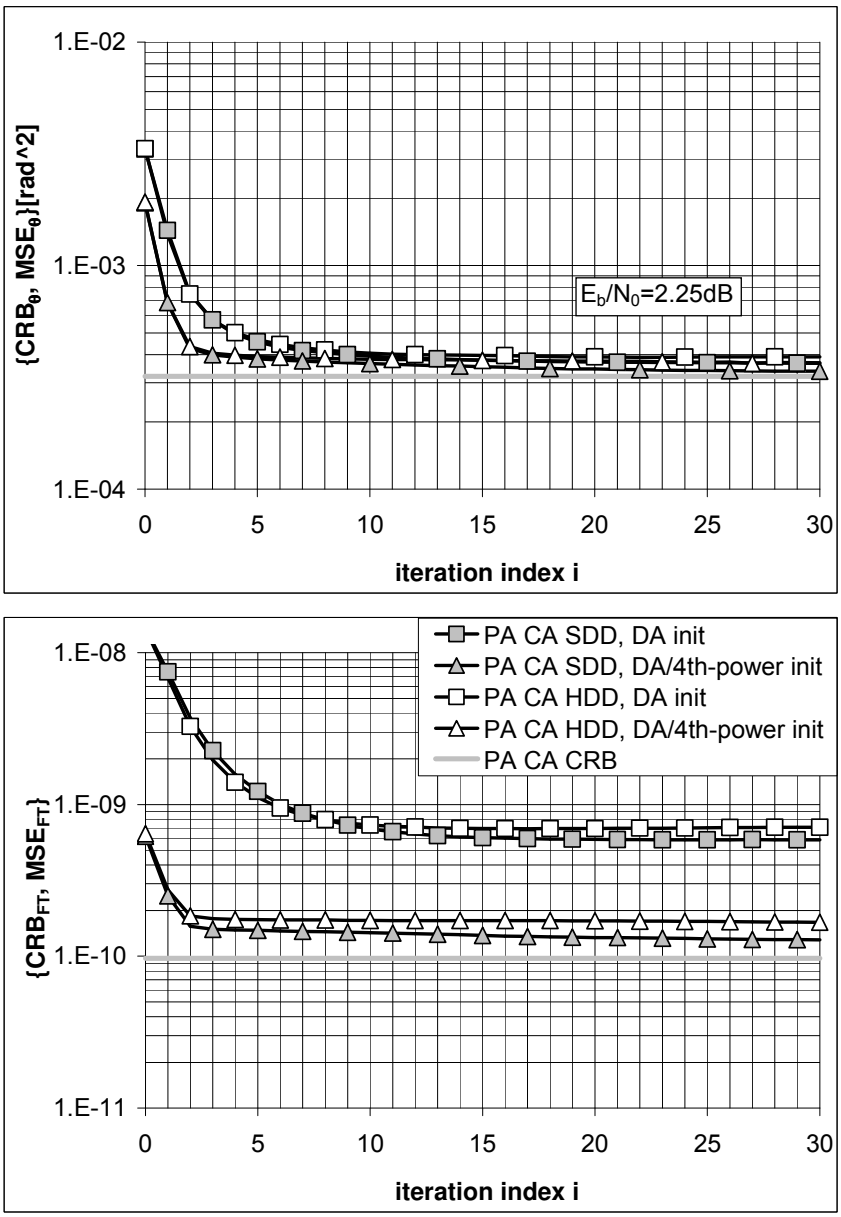

Figure 6.15: MSE resulting from PA CA DD carrier parameter estimation, with DA or DA/4th-power initialization, at $E_s/N_0 = E_b/N_0$ - 0.46 = 1.8 dB (4-PSK, $N_s = 1000$, $N_p = 100$).

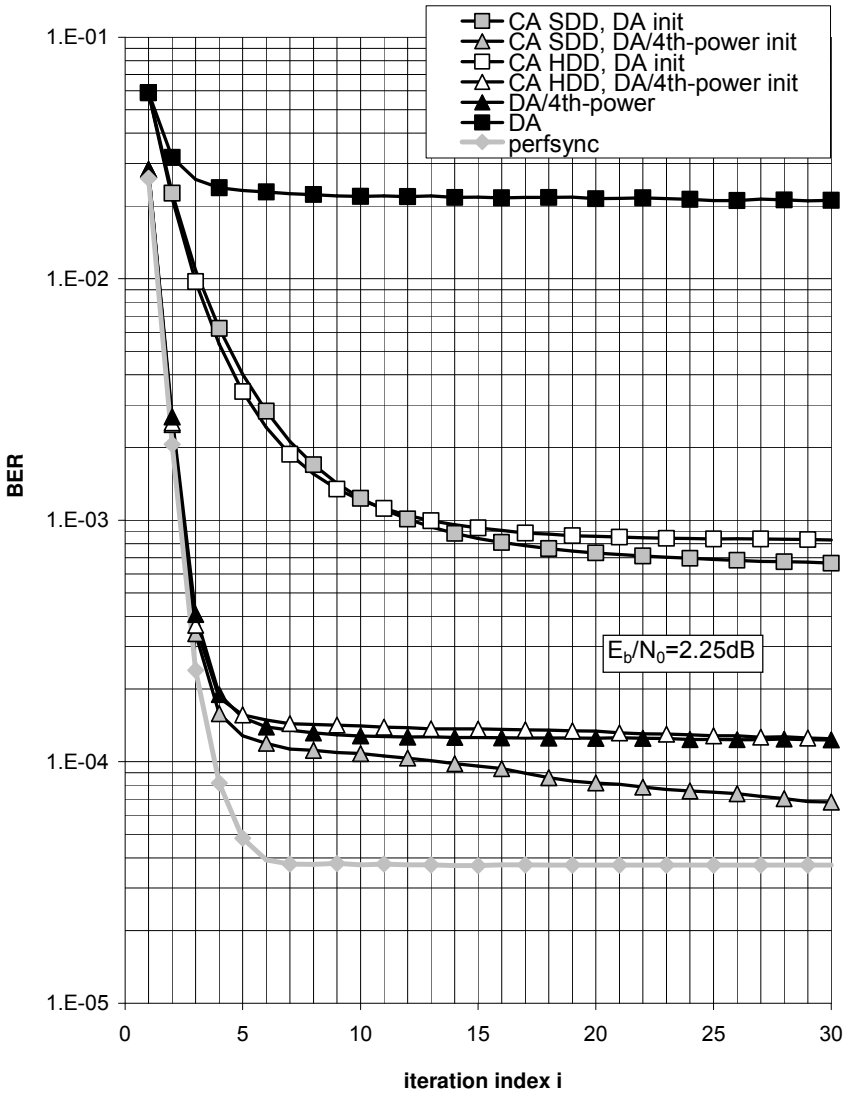

Figure 6.16: BER at $E_s/N_0 = E_b/N_0$ - 0.46 = 1.8 dB (4-PSK, $N_s = 1000$, $N_p = 100$), as obtained with perfect, DA, DA/4th-power, PA CA HDD and PA CA SDD carrier synchronization. The iterations of the PA CA HDD and PA CA SDD carrier synchronizers are intertwined with those of the detector.

- When the less accurate DA initial estimates are used, using a PA CA SDD or PA CA HDD synchronizer instead of a conventional DA/4th-power synchronizer significantly degrades the BER performance, although it yields about the same MSE for frequency offset estimation and substantially reduces the MSE for phase shift estimation. This clearly indicates the importance of a sufficiently accurate initial estimate to ensure the *joint* convergence of the intertwined synchronization and detection processes (EM algorithm approximation, Algorithm 11).

- For a given initialization, the PA CA estimator with soft decisions converges slower and to lower steady-state MSE and BER values than the PA CA estimator using hard decisions. When using DA/4th-power initialization and performing 30 iterations, the use of soft decisions instead of hard decisions reduces the BER with about a factor of 1.8.

- The PA CA HDD synchronizer with DA/4th-power initialization and 30 iterations yields virtually the same BER as the conventional DA/4th-power synchronizer: in this case, there is also no real benefit in using the more complex turbo synchronization approach, although the MSE at the normal operating SNR of the detector is significantly reduced.

- After 30 iterations, the PA CA SDD synchronizer with DA/4th-power initialization yields a BER which is about 1.8 times lower than in the case of DA/4th-power synchronization and only about 1.8 times as high as in the case of perfect synchronization.

Similar results for other values of E_b/N_0 are still being computed.

6.7 Conclusions and Remarks

The content of this chapter is based on [3–5, 11–17]. We have compared the CRBs from chapter 5 against the mean square estimation errors resulting from several types of practical ML-based carrier synchronizers [5, 60, 63, 77]. It was verified that the DA CRBs form a lower bound on the MSE performance of synchronizers that use only the pilot symbols, while the (N)PA NCA CRBs return a lower bound for the MSE performance of synchronizers that make no use of the code structure, and the NPA (N)CA CRBs lower-bound the performance of synchronizers that do not exploit the a priori knowledge about the pilot symbols. In order to approach the true PA CA CRB, during the estimation process estimators should make clever use of both the code properties and the pilot symbols.

We have been particularly concerned with the design of the carrier synchronization subsystem of receivers for powerful iteratively-decodable channel codes such as turbo and LDPC codes [35, 36, 38]: how to derive accurate carrier synchronization parameter estimates from the received signal at those (extremely low) SNRs typical of such codes? Earlier attempts of carrier synchronization in

the low-SNR regime focused on the traditional DA and (N)PA NCA algorithms. We have seen, however, that sometimes these algorithms are simply incapable of providing an estimation accuracy that is sufficiently high to ensure a negligible degradation of the BER at the output of the turbo or LDPC detector, as compared to the case of perfect synchronization. PA CA carrier synchronization parameter estimation appears as a natural solution to improve the carrier synchronization, but its implementation is at times critical and has led to numerous contributions in the technical literature, see e.g. [6] and references therein.

The MAP decoding of advanced codes (such as turbo and LDPC codes) involves the maximization of the APPs of the individual information bits. These APPs are obtained by a marginalization of the joint APP of all information bits and all coded symbols. The marginalization is accomplished by passing messages, computed by means of the sum-product algorithm, on a factor graph that represents a factorization of the joint APP. When the synchronization parameters are known, the factor graph has the information bits and coded symbols as variable nodes, and the synchronization parameters appear as parameters in some of the function nodes. As the factor graph contains cycles, the message-passing process becomes iterative, and the marginal APPs obtained after convergence are approximations of the true marginal APPs. In the presence of unknown synchronization parameters, we can distinguish two strategies for the detection of the advanced codes.

- In the case of synchronized detection, we use the same decoder as for known synchronization parameters. The decoder uses estimates (rather than the true values) of the synchronization parameters, which are updated after each decoding iteration. Hence, the synchronization process is also iterative. The joint iterative synchronization and detection process is initialized by conventionally obtained synchronization parameter estimates. The estimates are updated based on information provided by the decoder. This type of synchronization is ofter referred to as *turbo synchronization* (e.g., [68, 69, 73, 74]).

- Alternatively, the receiver performs decoding based on the iterative computation of (an approximation of) the APPs in the presence of the unknown synchronization parameters . This involves message passing in a factor graph that has the information bits, the coded symbols and the synchronization parameters as variable nodes. Strictly speaking, the receiver does not perform synchronization (because no explicit synchronization parameter estimates are used in the decoding process), but rather computes (as a by-product of the decoding process) the marginal a posteriori distributions of the synchronization parameters. Note that (unlike synchronized detection) this approach in general does not allow to use the decoder that assumes the synchronization parameters to be known (see e.g. [78–80]). As an exception, we mention the use of the sum-product algorithm in [75, 81]. Approximating by Dirac functions the messages leaving the variable nodes that represent the synchronization parameters, the iterative decoder that

assumes the synchronization parameters to be known can still be used. In this approach, the soft information to be provided by the decoder consists of the so-called extrinsic probabilities of the coded symbols.

In this doctoral thesis we restricted ourselves to the approach of synchronized detection. An important contribution is a new theoretical framework for turbo synchronization, yielding a new PA CA soft-decision-directed (SDD) turbo carrier synchronization algorithm with a good performance complexity trade-off, which requires from the decoder the marginal APPs of the coded symbols. It was observed that this iterative PA CA SDD estimator significantly outperforms conventional DA and (N)PA NCA algorithms in terms of MSE performance and operates very closely to the true CRB at the normal operating SNR of the detector, provided that an accurate initial estimate is available.

Although the MSE at the normal operating SNR of the detector is significantly reduced by using a PA CA synchronizer instead of a conventional DA or (N)PA NCA synchronizer, it highly depends on the specific transmission system considered whether or not this reduction in MSE yields a considerable improvement in BER performance. In [77], code-unaware algorithms for carrier phase, carrier frequency and timing estimation that operate on a turbo-coded 4-PSK signal give rise to a BER degradation of only 0.05 dB as compared to a perfectly synchronized system: in this case, there is no need to use PA CA synchronization to further reduce the already very small BER degradation. On the other hand, in section 6.6.3 a different turbo-coded 4-PSK system was considered, where the most accurate conventional algorithm, i.e., the NPA NCA two-stage DA/4th-power carrier synchronizer, yields a BER degradation of about 0.25 dB at a BER of 10^{-3}, whereas PA CA carrier synchronization reduces this BER degradation to about 0.1 dB only.

7

Conclusions and Ideas for Future Research

This chapter summarizes the main conclusions (see section 7.1) and formulates ideas for future research (see section 7.2).

7.1 Summarizing Conclusions and Remarks

Since the invention of turbo codes [38], receivers with synchronized maximum-a-posteriori (MAP) bit detectors operating according to the *iterative* sum-product algorithm (as with turbo codes, LDPC codes and BICM schemes) increasingly moved into the focus of research [36, 39, 43, 44]. Provided that the receiver's local reference carrier and clock are adequately *synchronized* with the incoming signal, such systems enable a reliable communication at lower signal-to-noise ratios (SNR) than ever before. Of course, this also sets high requirements as to the synchronizer's design, which in turn has stimulated international research into more effective synchronization structures [66–69, 75, 78–80]. It is in this context that our doctoral research is to be situated.

In this thesis we restricted our attention to carrier synchronization. We

considered the transmission of linearly-modulated burst-mode signals over a bandlimited additive white Gaussian noise (AWGN) channel. Each transmitted signal was obtained by applying a sequence of N_s constellation symbols, at a rate of $1/T$ symbols per second, to a square-root Nyquist transmit filter. The constellation symbols were either known pilot symbols or unknown random data symbols. The latter resulted from the encoding and mapping of a sequence of information bits. The carrier phase shift of the received signal vis-a-vis the receiver's local reference carrier was assumed to be linear in time over the duration of the burst. Consequently, it could be fully specified by the following two unknown but constant carrier synchronization parameters: the normalized frequency offset $\nu = FT$ and the reference phase shift θ. Carrier synchronization then involved estimating the value of ν and θ from $\mathbf{r}$, where $\mathbf{r}$ denotes a suitable vector representation of the received baseband signal.

In chapter 5, we derived Bayesian Cramer-Rao bounds (CRB) pertaining to the estimation of (ν,θ). The latter are a fundamental lower bound on the achievable mean square estimation error (MSE), and as such serve as a useful benchmark for practical parameter estimators [47]. The approach taken differed considerably from previous studies in three important aspects:

1. We did not (a priori) impose the independence nor the identical distribution of the transmitted symbols. This allowed us to derive CRBs for channel-coded and/or pilot-symbol-aided transmissions. Instead, we considered four different estimation modes: the pilot-aided code-aided (PA CA) mode, the data-aided (DA) mode, the pilot-aided non-code-aided (PA NCA) mode and the non-pilot-aided non-code-aided (NPA NCA) mode. Each estimation mode corresponds to a particular hypothesis about the a priori probability mass function of the transmitted symbol sequence. The PA CA mode takes into account the full a priori information, but the DA, the PA NCA and the NPA NCA modes use only a simplifying approximation thereof.

2. We used the correct continuous-time model of the received signal. Where possible, we compared our results from this correct model with existing results obtained from a simplified discrete-time model of the received signal. The latter model is commonly used but it ignores the reduction of the useful signal and the occurrence of inter-symbol-interference caused by a nonzero normalized frequency offset ν.

3. We derived CRBs not only from the 2×2 Fisher information matrix (FIM) related to the joint likelihood function $L(\nu, \theta; \mathbf{r}) = p(\mathbf{r}\,|\nu, \theta)$ of (ν, θ), but also from the scalar FIMs related to the marginal likelihood functions $L(\nu; \mathbf{r}) = p(\mathbf{r}\,|\nu)$ and $L(\theta; \mathbf{r}) = p(\mathbf{r}\,|\theta)$ of ν and θ, respectively. In [49], it was shown that the CRBs using $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ are always greater than or equal to the CRBs obtained from $L(\nu, \theta; \mathbf{r})$, and thus more tight.

Taking into account the linear modulation, we first expressed the 2×2 FIM related to $L(\nu, \theta; \mathbf{r})$ in terms of the marginal symbol a posteriori probabilities,

which in turn resulted in an algorithm that allows for an efficient numerical evaluation of the corresponding CRBs. This algorithm holds for both observation models and for the four estimation modes that are being considered. It involves the numerical evaluation of a statistical expectation with respect to a complex-valued vector of size N_s. This average does depend on the encoding rule, the pilot-symbol-insertion rule and the mapping rule, but not on the pulse shape. The effect of the pulse shape is analytically accounted for. For the DA, the PA NCA and the NPA NCA modes, additional simplifications were performed. The DA FIM related to $L(\nu, \theta; \mathbf{r})$ was evaluated analytically. For the computation of the PA NCA and the NPA NCA FIMs related to $L(\nu, \theta; \mathbf{r})$ a reduced-complexity algorithm was presented, which only involves the numerical evaluation of one (for the simplified observation model) or two (for the correct observation model) statistical expectations pertaining to a complex-valued scalar (for PSK) or a real-valued scalar (for PAM and QAM). These averages depend on the mapping rule, but not on the pulse shape or the pilot-symbol-insertion rule. The effect of the pulse shape and of the location of the pilot symbols is analytically accounted for.

In general, the scalar FIMs related to the likelihood functions $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ are much harder to compute than the 2×2 FIM related to $L(\nu, \theta; \mathbf{r})$. Taking into account the linear modulation and assuming statistically independent symbols, we nevertheless derived efficient numerical procedures for evaluating the FIMs related to $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ (comparable to the procedure for computing the FIM related to $L(\nu, \theta; \mathbf{r})$). These procedures hold for the DA, the PA NCA and the NPA NCA estimation modes, but unfortunately not for the PA CA estimation mode. The computation of the PA CA FIMs related to $L(\nu; \mathbf{r})$ and $L(\theta; \mathbf{r})$ seems practically unfeasible. The procedure for evaluating the DA, PA NCA or NPA NCA FIM related to $L(\nu; \mathbf{r})$ has been simplified further in the case of the correct observation model, whereas a similar reduction of the computational complexity is shown to be impossible for the FIM related to $L(\theta; \mathbf{r})$.

Numerical results were obtained for the DA, the NPA NCA, the PA NCA and the PA CA CRBs related to $L(\nu, \theta; \mathbf{r})$, as well as for the DA and the NPA NCA CRBs related to $L(\nu; \mathbf{r})$. This led to the following findings:

- At the normal operating SNR of a coded system (say, bit error rate (BER) smaller than 10^{-3}), the PA CA CRBs related to $L(\nu, \theta; \mathbf{r})$ are very close to the DA CRBs that would be obtained if all transmitted symbols were known pilot symbols. Furthermore, the PA CA CRBs related to $L(\nu, \theta; \mathbf{r})$ are less than the PA NCA CRBs related to $L(\nu, \theta; \mathbf{r})$, which are in turn less than the NPA NCA CRBs related to $L(\nu, \theta; \mathbf{r})$. This indicates that in order to approach optimal performance, estimators should make clever use of both the code properties and pilot symbols during the estimation process.

- The correct and the simplified model yield CRBs that only differ substantially at a low SNR. The influence of the pulse shape is restricted to these

low SNR values. For all practical values of the SNR we have verified that the useful signal magnitude is reduced by less than 0.01 dB and the ISI power is at least 20 dB below the noise power, provided that $|\nu| < 0.015$ ($|\nu| < 0.030$) for a cosine roll-off transmit pulse with roll-off factor of 20 % (of 100 %). It can therefore be concluded that the simplified observation model is valid as long as the maximum normalized frequency offset is about 1 %.

- The CRBs related to $L(\nu; \mathbf{r})$ are larger than the corresponding CRBs related to $L(\nu, \theta; \mathbf{r})$, which implies that the former CRBs are the tightest bounds. This is in keeping with [49]. For a given SNR, the gap between the former CRBs and the latter decreases as the observation interval increases. For the DA CRB, this penalty can be neglected for practical values of SNR and for moderate observation intervals. For the NPA NCA CRB, considerably longer observation intervals are required in order to significantly reduce the penalty.

In chapter 6 we analyzed the performance of several practical feedforward carrier synchronization parameter estimation algorithms and we compared the resulting MSEs with the corresponding CRBs. It was verified that the DA CRBs are a lower bound on the MSE performance of synchronizers that only use the pilot symbols, while the (N)PA NCA CRBs provide a lower bound for the MSE performance of synchronizers that do not make use of the code structure and the NPA (N)CA CRBs provide a lower bound for the performance of synchronizers that do not exploit the a priori knowledge about the pilot symbols. Our main contribution is a new theoretical framework for iterative synchronization structures that exchange information between the MAP bit detector and the synchronization parameter estimator. Based on this work, we derived a new soft-decision-directed (SDD) carrier synchronization algorithm with a good performance complexity trade-off. This new algorithm makes clever use of both the code properties and the pilot symbols during the estimation process. The algorithm significantly outperforms the conventional carrier synchronization algorithms and it operates very closely to the PA CA CRBs at the normal operating SNR of systems with powerful channel codes. Although using a PA CA synchronizer instead of a conventional DA or (N)PA NCA synchronizer substantially reduces the MSE at the detector's normal operating SNR, it highly depends on the transmission system that is being considered whether or not this reduction in MSE will yield a considerable improvement in BER performance. In [77], code-unaware algorithms for carrier phase, carrier frequency and timing estimation which operate on a turbo-coded 4-PSK signal give rise to a BER degradation of only 0.05 dB as compared to a perfectly synchronized system: in this case, there is no need to use PA CA synchronization to reduce the already very small BER degradation any further. On the other hand, in section 6.6.3 a different turbo-coded 4-PSK system was considered, in which the most accurate conventional algorithm, namely the NPA NCA two-stage DA/4th-power carrier synchronizer, yields a BER degradation of about 0.25 dB at a BER of 10^{-3},

whereas PA CA carrier synchronization reduces this BER degradation to only about 0.1 dB.

7.2 Work-in-Progress and Ideas for Future Research

The research that is presented in this thesis has so far resulted in 6 journal articles [1–6] and 14 contributions in international conference proceedings [7–20]. Apart from these, 11 published papers (4 journal papers and 7 conference papers) [21–31] provide additional findings in relation to feedback carrier synchronization and timing recovery, two aspects that were not addressed in this thesis. The latter contributions are to be considered as work-in-progress and they form an extensive breeding ground for future research. More details and some preliminary results will be briefly presented in subsections 7.2.1 and 7.2.2.

7.2.1 Random Carrier Phase Shift Process

Let us assume that the received signal is applied to a matched filter, and sampled at the optimal instants kT , where T denotes the symbol period. We then obtain a sequence of matched filter output samples $\{r\,(kT)\}$. The carrier synchronization algorithms described in this thesis estimate the parameters (ν, θ) *directly* from these matched filter output samples, *before* a frequency and phase correction is applied to them. That is why they are referred to as *feedforward* (FF) algorithms. The main drawback of FF algorithms is that their field of operation is mainly restricted to the estimation of parameters that remain essentially constant over the observation interval. Throughout this thesis, we chose to consider a carrier phase shift that is linear in time over the duration of the burst. By making this assumption, the systematic derivation of carrier synchronization algorithms becomes a great deal easier. It also implies that it is sufficient to produce a single FF normalized frequency offset and phase shift estimate per burst, and to use these estimates for detecting all data symbols within the burst. On the other hand we should be aware of the fact that the carrier phase shift in practice is always subject to random fluctuations caused by imperfections in the transmitter and receiver oscillators. For long bursts, which is the common situation for codes with large coding gains, the fluctuation of the phase noise over the burst cannot always be ignored. This is shown in Fig. 7.1, reporting simulated BER results for a 2-PSK turbo-code detector when $N_s = 1000$ and feedforward carrier synchronization is performed. Ten joint PA CA SDD synchronization and MAP bit detection iterations are performed according to Algorithm 11. The initial estimate of (ν, θ) that is required to start the iterations of the receiver is obtained from the NPA NCA 2nd-power method (see section 6.5). The carrier phase shift is subject not only to a normalized frequency offset that is uniformly taken from $[-0.1, 0.1]$, but also to Wiener phase noise [82] that is characterized by independent and identically distributed Gaussian increments with a zero mean and a standard deviation of σ_u degrees.

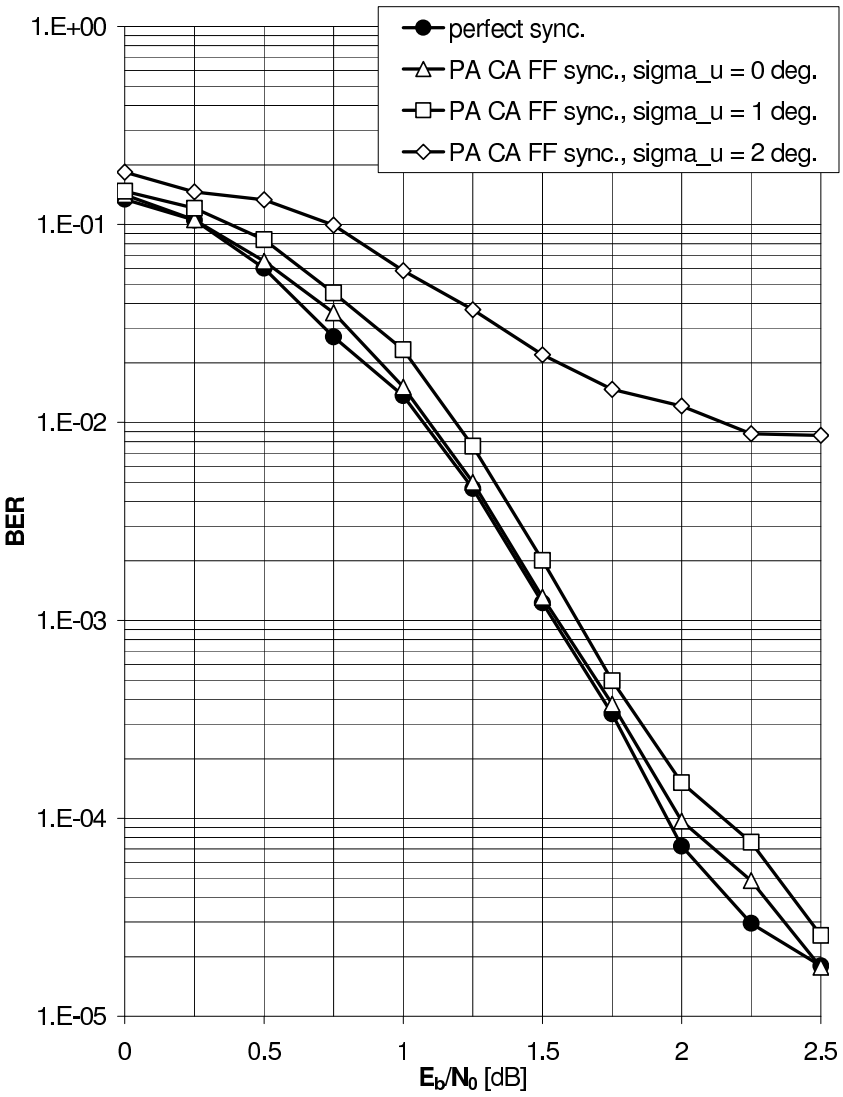

Figure 7.1: Effect of Wiener phase noise on the BER resulting from a feedforwardly synchronized rate-1/3 2-PSK turbo-code detector when $N_s = 1000$.

We observe that the resulting BER closely approaches that of the perfectly synchronized system when $\sigma_u = 0$ (no phase noise), but rapidly increases in the presence of phase noise ($\sigma_u = 1$, 2 degrees).

A possible solution to this problem consists of breaking down the problem of carrier synchronization into two parts: coarse normalized frequency offset correction and residual carrier phase shift tracking. A coarse estimate $\hat{\nu}_{coarse}$ of the normalized frequency offset ν can, for example, be obtained from a DA FF estimator that operates on a short pilot symbol sequence. Provided that the number of symbol periods between the transmission of the first and the last pilot symbol is significantly smaller than $1/\sigma_u^2$, it is plausible to assume that the carrier phase shift is approximately linear in time over the duration of the pilot symbol sequence, which means that the results from section 6.4 apply. After performing coarse normalized frequency offset correction of the received samples, the *tracker* makes multiple phase shift estimates per burst in order to track the *slow* variations of the residual carrier phase shift (due to the phase noise and the residual normalized frequency offset) over the burst.

A feedforward approach to tracking consists of using an observation window that is slid across the entire burst, one or more symbols at a time. At each window position, the currently observed symbols are used to produce a feedforward estimate of the instantaneous phase shift at the center of the observation window. In general, shorter observation windows can handle faster phase shift variations, but the part of the MSE caused by the AWGN increases, so that an optimum window length exists for a given SNR. Initially proposed for use with the Kth-power phase shift estimator in [63], this technique has been applied in [77] to a turbo-coded system. A related approach has also been suggested in [5], for use with the iterative PA CA SDD estimator.

An alternative for the sliding window technique consists of using a phase error *feedback* (FB) control loop. As opposed to FF synchronizers, FB synchronizers derive an estimate of the instantaneous phase error from the phase-corrected matched filter output samples and feed this estimate *back* to the phase correction block. The applied phase correction is updated at regular intervals during the burst interval, e.g., once per symbol period. As a result, FB structures are able to automatically track slow variations of the phase shift, whereas FF structures always have to start a memoryless estimation process from observation window to observation window. The typical block diagram of FF and FB carrier synchronizers is shown in Fig. 7.2 . The value $\hat{\Theta}(t)$ of the phase correction $\hat{\theta}$ applied at time t can be interpreted as an estimate of the instantaneous phase shift $\Theta(t)$.

Inspired by the EM algorithm, we are interested in designing a PA CA SDD FB scheme, i.e., a FB synchronizer that exploits the code properties and the pilot symbols in the phase error detection process. It is well known that (conventional NCA) FB schemes can be based upon (conventional NCA) FF algorithms [34], but complications arise when this approach is extended to PA CA SDD synchronization. In the latter case, computing the phase shift estimate $\hat{\Theta}(k_1 T)$ would imply evaluating the soft decision regarding the k_1th symbol based upon the

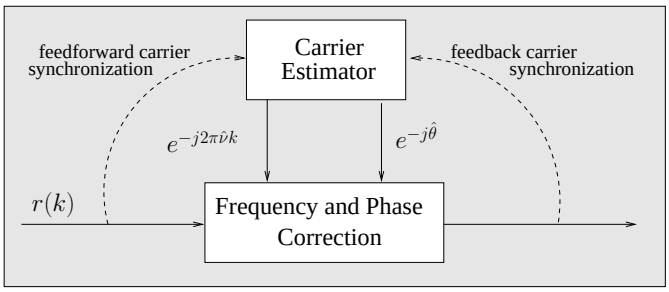

Figure 7.2: Feedforward and feedback carrier synchronization algorithms.

complete sequence $\left\{r\left(kT\right)e^{-j\hat{\Theta}(kT)}\right\}$ of phase corrected matched filter output samples, while, at instant k_1T, the phase shift estimates $\hat{\Theta}\left(k_2T\right)$ with $k_2 > k_1$ would not yet be available. In [23–27], we have presented and discussed an *iterative* PA CA FB synchronization structure that makes it possible to circumvent these difficulties. During the ith iteration we obtain the sequence of phase shift estimates $\left\{\hat{\Theta}^{(i)}\left(kT\right)\right\}$. The soft decisions needed in the ith iteration are computed from the sequence of phase shift estimates obtained during the $(i-1)$th iteration. The sequence of phase shift estimates $\left\{\hat{\Theta}^{(0)}\left(kT\right)\right\}$, needed to start the iterations can be obtained by a conventional NCA FB scheme. The interaction between the MAP detector and this iterative FB synchronizer closely resembles that between the MAP detector and the PA CA SDD FF synchronizer from section 6.6.2.3. As opposed to the FB phase synchronizers in [66, 83, 84], the derivation of the proposed scheme stems directly from the ML criterion. Moreover, the computational complexity of this PA CA synchronizer is lower than that of the algorithms in [85] and [78, 79], which modify the decoder operation by either taking into account the phase statistics or using a kind of 'per-survivor' phase estimation technique.

Assuming a zero normalized frequency offset, simulations run for a *first-order* FB loop in combination with an off-the-shelf 2-PSK turbo-code detector already yielded very promising performance results. Fig. 7.3 reports simulated BER results for $N_s = 1000$, $\nu \equiv 0$ and a first-order FB loop gain of 0.04. Seven joint PA CA FB synchronization and MAP bit detection iterations are performed. We observe that the receiver with FB synchronization is relatively robust against Wiener phase noise, and is able to cope with phase variations (such as $\sigma_u = 1$, 2 degrees) that are too large for proper operation of the receiver with FF synchronization (compare with Fig. 7.1). The BER degradation when the carrier phase shift is constant (i.e., $\sigma_u = 0$ degrees) is only caused by the AWGN component of the phase error. A slightly larger degradation occurs for $\sigma_u = 1$, 2 degrees, because a phase-noise component is added to the phase error.

A topic for future research is the use of a *second-order* FB loop in the presence of a nonzero frequency offset (which adds to the phase noise). The use

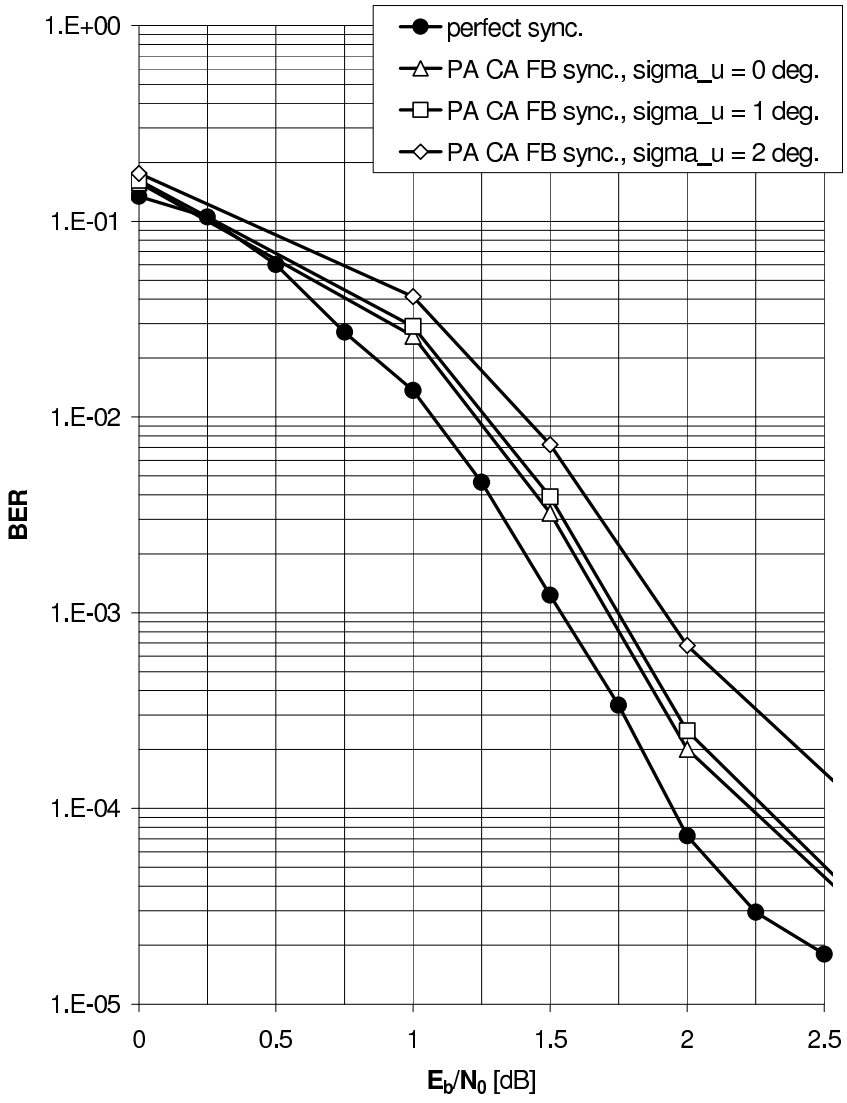

Figure 7.3: Effect of Wiener phase noise on the BER resulting from rate-1/3 2-PSK turbo-code detector with feedback synchronization when $N_s = 1000$.

of the proposed scheme in combination with higher-order signaling constellations also remains a topic for future research. The same goes for a comparison with another recently proposed scheme that estimates a time-varying carrier phase shift from coded signals using sequential Monte Carlo methods (see [28–31, 86]).

7.2.2 Time Delay

An important issue that fell outside the scope of this thesis is *timing synchronization*. So far, it was tacitly assumed that the receiver's local sampling clock was perfectly synchronized with the received baseband signal. However, we should be aware of the fact that the received signal in practice is always subject to an unknown (and possibly time-varying) propagation delay τ. In order to enable timing synchronization at the receiver, the value of τ has to be estimated (prior to or jointly with the carrier synchronization parameters) from the received baseband signal vector $\mathbf{r}$. The investigation of CRBs and practical algorithms pertaining to the estimation of τ, taking into account the low operating SNR of receivers with iterative MAP bit detection, is considered to be a very interesting topic for further research. A few preliminary steps in this direction have already been presented in [3, 5, 21, 22]. The difficulties encountered, the approach taken and the numerical results obtained for timing synchronization seem to closely resemble those for carrier synchronization.

APPENDICES

Appendix A

The purpose of this appendix is about showing that the majority of practical coded modulation schemes produce channel symbols that have the same first- and second-order moments as uncoded symbols.

Remember that we are dealing with channel symbols $d(k)$, $k \in \{0,\ 1,\ ...,\ N_d - 1\}$, which are described by a *linear* encoding rule:

$$c(j) = \mathbf{b}\mathbf{G}_j = \bigoplus_{i=0}^{N_b-1} b(i)\, G_{i,j}, \quad j \in \{0, 1, ..., N_c - 1\} \tag{A.1}$$

and a mapping rule:

$$d(k) = \mathcal{D}_m(\mathbf{c}_k), \quad k \in \{0, 1, ..., N_d - 1\}. \tag{A.2}$$

In (A.1), $\mathbf{b} = \begin{pmatrix} b(0) & b(1) & ... & b(N_b - 1) \end{pmatrix}$ denotes a sequence of N_b information bits, $\{c(j),\ j = 0, 1, ..., N_c - 1\}$ denote N_c coded bits, the matrix $\mathbf{G}$ with binary-valued elements $G_{i,j}$ and column vectors $\mathbf{G}_j = (G_{0,j}G_{1,j}...G_{N_b-1,j})^T$ denotes the $N_b \times N_c$ generator matrix of the code, and $\oplus$ denotes modulo-2 addition. In (A.2), $\mathcal{D}_m$ denotes a bijective time-invariant mapping function $\mathcal{D}_m : \{0,1\}^m \rightarrow \Omega_M$ with $M = 2^m$ that maps m bits to one complex number in $\Omega_M = \{\omega_0, \omega_1, ..., \omega_{M-1}\}$, and the vector $\mathbf{c}_k$ denotes the set of coded bits that

are mapped to the kth data symbol. We will use k_1, k_2, ..., k_m to denote the indices of the coded bits that constitute $\mathbf{c}_k$, i.e., $\mathbf{c}_k = (c(k_1)\ c(k_2)\ ...\ c(k_m))$ for $k \in \{0, 1, ..., N_d - 1\}$. Furthermore, it should be noted that

(prop1) All information bits are *statistically independent,* with *equal probability* for having a logical "1" and a logical "0".

(prop2) The signaling constellation Ω_M is *symmetric about the origin* of the complex plane and satisfies the *normalization* condition:

$$\frac{1}{M}\sum_{i=0}^{M-1}|\omega_i|^2 = 1.$$

In [87], it was shown that most convolutional codes with conventional Gray-mapping and trellis-coded modulations (TCM) encountered in the literature yield the same first- and second-order moments as for uncoded symbols. This proves our statement for the case of convolutional codes with conventional Gray-mapping and TCM. We will now extend this work to the case of turbo codes with conventional Gray-mapping, low-density parity-check (LDPC) codes with conventional Gray-mapping and bit-interleaved coded modulation (BICM). We will derive a simple condition on the generator matrix of the linear component code, which is sufficient for first- and second-order moment invariance and easy to verify.

Any linear channel code is described by its generator matrix $\mathbf{G}$. It follows directly from (A.1) and (prop1) that the coded bits $c(j)$ have equally likely possible values. A necessary and sufficient condition for a set of coded bits $\{c(j)\}$ to be statistically independent is then that the corresponding generator matrix column vectors $\{\mathbf{G}_j\}$ are linearly independent.

As the data symbols $d(k)$ and $d(l)$ are related to the coded bits $(c(k_1)\ c(k_2)\ ...\ c(k_m))$ and $(c(l_1)\ c(l_2)\ ...\ c(l_m))$, respectively, we find that the following condition is sufficient for first- and second-order moment invariance:

(cond) For each k and each $l \neq k$ $(0 \leq k, l \leq N_d - 1)$, the code generator matrix column vectors $\mathbf{G}_{k_1}$, $\mathbf{G}_{k_2}$, ..., $\mathbf{G}_{k_m}$, $\mathbf{G}_{l_1}$, $\mathbf{G}_{l_2}$, ..., $\mathbf{G}_{l_m}$ are linearly independent.

When this condition is fulfilled, the marginal probability mass function of any data symbols $d(k)$ is uniform over the signaling constellation, i.e., $\Pr[d(k) = \omega] = 1/\mathrm{M}$ for all $\omega \in \Omega_M$, and any two channel symbols $d(k)$ and $d(l)$ are statistically independent, i.e., $\Pr[d(k), d(l)] = \Pr[d(k)]\Pr[d(l)]$. This proves our statement.

It is intuitively clear that, for the usual case of small m and large N_b, (cond) is likely to hold when the generator matrix $\mathbf{G}$ exhibits a large degree of randomness (as is the case for LDPC codes and turbo codes with conventional Gray-mapping). We can now investigate the above condition for a BICM scheme

using an LDPC code, a convolutional code or a turbo code. In this case (cond) can be reformulated as follows:

(cond') For each k and each $l \neq k$, $0 \leq k, l \leq N_d - 1$, the code generator matrix column vectors $\mathbf{G}'_{km}$, $\mathbf{G}'_{km+1}$, ..., $\mathbf{G}'_{km+m-1}\mathbf{G}'_{lm}$, $\mathbf{G}'_{lm+1}$, ..., $\mathbf{G}'_{lm+m-1}$ are linearly independent.

where $\mathbf{G}'_j$ denotes the jth column of the matrix $\mathbf{G}'$, which in turn results from interleaving the columns of the generator matrix $\mathbf{G}$ in the same way as the coded bits are interleaved in the BICM scheme. Again, it is intuitively clear that, for the usual case of small m and large N_b, (cond') is likely to hold when the original generator matrix $\mathbf{G}$ is sparse (as is the case for convolutional codes and concatenated coding schemes using convolutional codes) or already exhibits a large degree of randomness (as is the case for LDPC codes and turbo codes).

Based on the above results it can be concluded that most practical coded modulation schemes produce channel symbols that have the same first- and second-order moments as uncoded symbols.

Appendix B

In this appendix, we present three methods to obtain a factor graph representation of a channel code (based on [45]).

B.1 Method 1: Factor Graph Construction from Generator Matrix

It follows directly from the discussion in section 2.2.1 that the code constraint $I[\mathbf{c} = \mathcal{C}(\mathbf{b})]$ of any linear code can be written as:

$$\begin{aligned} I[\mathbf{c} = C(\mathbf{b})] &= I[\mathbf{c} = \mathbf{bG}], \\ &= \prod_{k=0}^{N_c-1} I\left[c(k) = \bigoplus_{l=0}^{N_b-1} b(l) G_{l,k}\right], \\ &= \prod_{j=0}^{N_b-1} I\left[b(j) = b^0(j) = b^1(j) = ... = b^{N_c-1}(j)\right] \\ &\quad \cdot \prod_{k=0}^{N_c-1} I\left[0 = \left(\bigoplus_{l=0}^{N_b-1} b^k(l) G_{l,k}\right) \oplus c(k)\right], \end{aligned}$$

where $\oplus$ denotes modulo-2 addition, $\mathbf{G}$ is the generator matrix of the linear code, $G_{l,k}$ denotes the (l,k)-th element of $\mathbf{G}$. By introducing $N_b \times N_c$ dummy variables $\left\{b^j(l) : (0 \le l \le N_b - 1) \wedge (0 \le j \le N_c - 1)\right\}$, we have made sure that each variable appears in not more than two factors. Being 'clones' of the information bits, all these additional variables are defined over $\{0,1\}$. The factor graph corresponding to the above factorization of $I[\mathbf{c} = \mathcal{C}(\mathbf{b})]$ is depicted in Fig. B.1. The top "=" nodes enforce the N_b equality constraints $I\left[b(j) = b^0(j) = b^1(j) = ... = b^{N_c-1}(j)\right]$, and the bottom $'\oplus'$ nodes correspond to the N_c modulo-2 addition constraints $I\left[0 = \left(\bigoplus_{l=0}^{N_b-1} b^k(l) G_{l,k}\right) \oplus c(k)\right]$. Not all the edges are necessarily present. The edge between the l-th equality node and the k-th modulo-2 addition node is present when $G_{l,k}$ is equal to one, otherwise it is absent. In general, the factor graph will contain cycles.

B.2 Method 2: Factor Graph Construction from Parity-Check Matrix

It follows directly from the discussion in section 2.2.1 that the code constraint $I[\mathbf{c} \in \zeta_{\mathcal{C}}]$ of any linear code can be written as:

$$\begin{aligned} I[\mathbf{c} \in \zeta_{\mathcal{C}}] &= I\left[0 = \mathbf{H}\mathbf{c}^T\right], \\ &= \prod_{k=0}^{N_c-N_b-1} I\left[0 = \bigoplus_{l=0}^{N_b-1} H_{k,l} c(l)\right], \\ &= \prod_{j=0}^{N_c-1} I\left[c(j) = c^0(j) = c^1(j) = ... = c^{N_c-N_b-1}(j)\right] \\ &\quad \cdot \prod_{k=0}^{N_c-N_b-1} I\left[0 = \bigoplus_{l=0}^{N_b-1} H_{k,l} c^k(l)\right], \end{aligned}$$

where $H_{l,k}$ denotes the (l,k)-th element of $\mathbf{H}$, and $N_c \times (N_c - N_b)$ dummy variables $\left\{c^j(l) : (0 \le l \le N_c - 1) \wedge (0 \le j \le N_c - N_b - 1)\right\}$ defined over $\{0,1\}$ were introduced. The factor graph corresponding to the above factorization of $I[\mathbf{c} \in \zeta_{\mathcal{C}}]$ is shown in Fig. B.2. The bottom "=" nodes enforce the N_c equality constraints $I\left[c(j) = c^0(j) = c^1(j) = ... = c^{N_c-N_b-1}(j)\right]$, and the bottom "$\oplus$" nodes correspond to the $(N_c - N_b)$ modulo-2 addition constraints $I\left[0 = \bigoplus_{l=0}^{N_b-1} H_{k,l} c^k(l)\right]$. The edge between the l-th equality node and the k-th modulo-2 addition node is present when $H_{l,k}$ is equal to one, otherwise it is absent. This graph is essentially the same as a *Tanner graph* [88] for the code. In general, it will contain cycles.

B.3 Method 3: Factor Graph Construction from Trellis

A factor graph for a code can also be obtained as an abstraction and generalization of a trellis diagram. For example, from the trellis of a convolutional code we can derive the following factorization of $I[\mathbf{c} = C(\mathbf{b})]$:

$$\begin{aligned} & I[\mathbf{c} = C(\mathbf{b})] \\ & = f_0(s_0) \prod_{k=0}^{N_b-1} f_{trellis}\left(s_k, s_{k+1}, b(k), c^{(1)}(k), ..., c^{(\nu)}(k)\right), \end{aligned} \tag{B.1}$$

where we have used the notations from section 2.2.1.2. We have introduced N_b additional variables s_k, denoting the states at times k. Each additional variable may take 2^κ possible values. The function $f_0(s_0)$ in (B.1) enforces the initialization constraint, i.e.,

$$f_0(s_0) = I\left[s_0 = \underbrace{(\ 0 \quad 0 \quad ... \quad 0\ )}_{\kappa}\right].$$

The function $f_{trellis}()$ in (B.1) is related to the branches in the Viterbi trellis, it is always zero except when at time k an input $b(k)$ triggers a transition from state s_k to state s_{k+1}, thereby generating ν output bits $c^{(1)}(k)$, $c^{(2)}(k)$, ..., $c^{(\nu)}(k)$. The factor graph corresponding to the above factorization of $I[\mathbf{c} = C(\mathbf{b})]$ is depicted in Fig. B.3. In contrast with the methods based on the generator matrix or the parity-check matrix, we end up with a cycle-less graph. Also, it should be noted that the above procedure is not limited to convolutional codes and their trellis. In [52], Wolf has shown how a trellis can be derived from the parity-check matrix of an arbitrary linear code. The corresponding factor graph consists of $N_c + 1$ nodes and contains N_c additional variables that can each take $(N_c - N_b)$ possible values; this factor graph gives rise to an alternative factorization of the parity-check constraint $I\left[0 = \mathbf{H}\mathbf{c}^T\right]$.

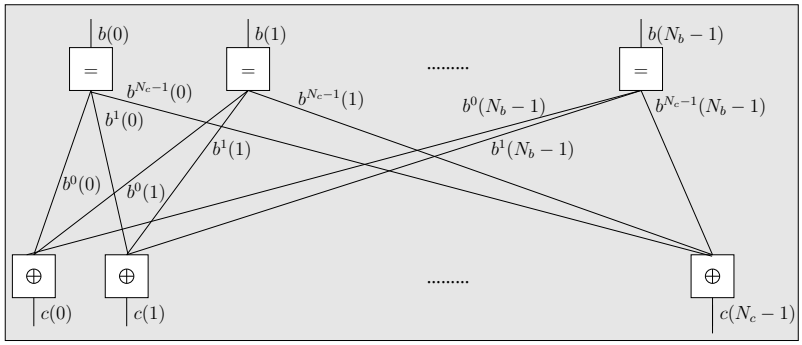

Figure B.1: Further decomposition of the code constraint node for coded transmission, using the code generator matrix.

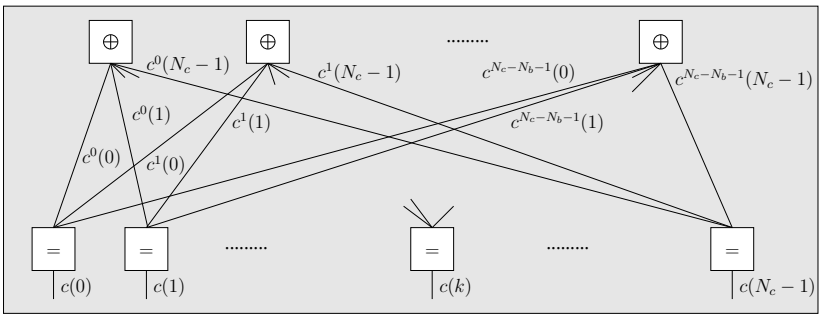

Figure B.2: Further decomposition of the code constraint node for coded transmission, using the code parity-check matrix.

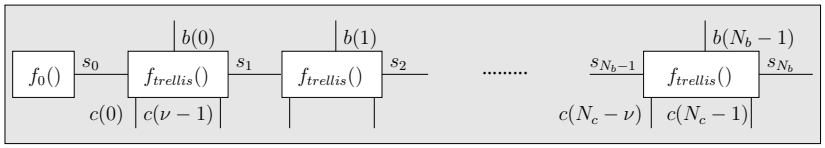

Figure B.3: Further decomposition of the code constraint node using the code trellis.

Appendix C

We consider the following two equivalent representations of the instantaneous phase shift:

$$\Theta_1 (kT; \mathbf{\Psi}_1) = 2\pi\nu (k - \kappa_1) + \theta_1 \tag{C.1}$$

and

$$\Theta_2 (kT; \mathbf{\Psi}_2) = 2\pi\nu (k - \kappa_2) + \theta_2, \tag{C.2}$$

where

$$\mathbf{\Psi}_1 = \begin{pmatrix} \nu \\ \theta_1 \end{pmatrix},$$

$$\mathbf{\Psi}_2 = \begin{pmatrix} \nu \\ \theta_2 \end{pmatrix}$$

and

$$\theta_1 = \theta_2 + 2\pi\nu (\kappa_1 - \kappa_2). \tag{C.3}$$

Subsequently, we consider two estimation scenarios:

1. The first estimation scenario chooses κ_1 as the normalized reference time instant. By denoting the estimate of $\mathbf{\Psi}_1$ that is derived from the observation vector $\mathbf{r}$ as $\hat{\mathbf{\Psi}}_1 = \left(\hat{\nu}\, \hat{\theta}_1\right)^T$, we obtain the following instantaneous

phase shift error:

$$\left(\Theta_1\left(kT;\mathbf{\Psi}_1\right)-\Theta_1\left(kT;\hat{\mathbf{\Psi}}_1\right)\right)=2\pi\left(\nu-\hat{\nu}\right)\left(k-\kappa_1\right)+\left(\theta-\hat{\theta}_1\right) \quad \text{(C.4)}$$

which is lower bounded by the CRB:

$$E\left[\left(\Theta_1\left(kT;\mathbf{\Psi}_1\right)-\Theta_1\left(kT;\hat{\mathbf{\Psi}}_1\right)\right)^2\right]\geq CRB_{\Theta_1(kT;\mathbf{\Psi}_1)}$$

with

$$CRB_{\Theta_1(kT;\mathbf{\Psi}_1)}=\left(\begin{array}{cc}2\pi\left(k-\kappa_1\right) & 1\end{array}\right)\left(\mathbf{J}_1\right)^{-1}\left(\begin{array}{c}2\pi\left(k-\kappa_1\right)\\ 1\end{array}\right) \quad \text{(C.5)}$$

In (C.5), $\left(\mathbf{J}_1\right)^{-1}$ denotes the inverse of the FIM $\mathbf{J}_1$ related to the likelihood function $p_1\left(\mathbf{r}\,|\nu,\theta_1\right)$.

2. The second estimation scenario chooses κ_2 as the normalized reference time instant. By denoting the estimate of $\mathbf{\Psi}_2$ that is derived from the observation vector $\mathbf{r}$ as $\hat{\mathbf{\Psi}}_2=\left(\hat{\nu}\,\hat{\theta}_2\right)^T$, we obtain the following instantaneous phase shift error:

$$\left(\Theta_2\left(kT;\mathbf{\Psi}_2\right)-\Theta_2\left(kT;\hat{\mathbf{\Psi}}_2\right)\right)=2\pi\left(\nu-\hat{\nu}\right)\left(k-\kappa_2\right)+\left(\theta-\hat{\theta}_2\right) \quad \text{(C.6)}$$

which is lower bounded by the CRB:

$$E\left[\left(\Theta_2\left(kT;\mathbf{\Psi}_2\right)-\Theta_2\left(kT;\hat{\mathbf{\Psi}}_2\right)\right)^2\right]\geq CRB_{\Theta_2(kT;\mathbf{\Psi}_2)},$$

with

$$CRB_{\Theta(kT;\mathbf{\Psi}_2)}\geq\left(\begin{array}{cc}2\pi\left(k-\kappa_2\right) & 1\end{array}\right)\left(\mathbf{J}_2\right)^{-1}\left(\begin{array}{c}2\pi\left(k-\kappa_2\right)\\ 1\end{array}\right). \quad \text{(C.7)}$$

In (C.7), $\left(\mathbf{J}_2\right)^{-1}$ denotes the inverse of the FIM $\mathbf{J}_2$ related to the likelihood function $p_2\left(\mathbf{r}\,|\nu,\theta_2\right)$.

The relation between the likelihood functions $p_1\left(\mathbf{r}\,|\nu,\theta_1\right)$ and $p_2\left(\mathbf{r}\,|\nu,\theta_2\right)$ follows directly from the relation between the phase shifts θ_1 and θ_2 at the normalized reference time instants κ_1 and κ_2 (see (C.3)). We obtain:

$$p_2\left(\mathbf{r}\,|\nu,\theta_2\right)=p_1\left(\mathbf{r}\,|\nu,\theta_2+2\pi\nu\left(\kappa_1-\kappa_2\right)\right). \quad \text{(C.8)}$$

Taking the logarithm of (C.8) and differentiating both sides with respect to ν and θ_2 subsequently yields:

$$\dot{\boldsymbol{\ell}}_2\left(\nu,\theta_2\right)=\left(\begin{array}{cc}1 & 2\pi\left(\kappa_1-\kappa_2\right)\\ 0 & 1\end{array}\right)\dot{\boldsymbol{\ell}}_1\left(\nu,\theta_2+2\pi\nu\left(\kappa_1-\kappa_2\right)\right), \quad \text{(C.9)}$$

with for i= 1,2:

$$\dot{\boldsymbol{\ell}}_i\left(\tilde{\nu},\tilde{\theta}\right)=\left(\begin{array}{c}\frac{\partial \ln p_i(\mathbf{r}|\nu,\theta_i)}{\partial\nu}\\ \frac{\partial \ln p_i(\mathbf{r}|\nu,\theta_i)}{\partial\theta_i}\end{array}\right)_{(\nu,\theta_i)=\left(\tilde{\nu},\tilde{\theta}\right)},$$

where $\tilde{\nu}$ and $\tilde{\theta}$ denote trial values of ν and θ_i with i=1,2. The following relation between the FIM $\mathbf{J}_1$ and $\mathbf{J}_2$ can now be derived from (C.9):

$$\begin{aligned}\mathbf{J}_2 &= E_{\mathbf{r},\nu,\theta_2}\left[\dot{\boldsymbol{\ell}}_2(\nu,\theta_2)\left(\dot{\boldsymbol{\ell}}_2(\nu,\theta_2)\right)^T\right],\\ &= \left(\begin{array}{cc}1 & 2\pi(\kappa_1-\kappa_2)\\ 0 & 1\end{array}\right)\mathbf{J}_1\left(\begin{array}{cc}1 & 0\\ 2\pi(\kappa_1-\kappa_2) & 1\end{array}\right),\end{aligned}\tag{C.10}$$

with

$$\begin{aligned}&\mathbf{J}_1 =\\ &E_{\mathbf{r},\nu,\theta_1}\left[\dot{\boldsymbol{\ell}}_1(\nu,\theta_1)\left(\dot{\boldsymbol{\ell}}_1(\nu,\theta_1)\right)^T\right] =\\ &E_{\mathbf{r},\nu,\theta_2}\left[\dot{\boldsymbol{\ell}}_1(\nu,\theta_2+2\pi\nu(\kappa_1-\kappa_2))\left(\dot{\boldsymbol{\ell}}_1(\nu,\theta_2+2\pi\nu(\kappa_1-\kappa_2))\right)^T\right].\end{aligned}$$

Applying matrix inversion to (C.10) we obtain:

$$(\mathbf{J}_2)^{-1} = \mathbf{K}(\mathbf{J}_1)^{-1}\mathbf{K}^T,\tag{C.11}$$

with

$$\mathbf{K}=\left(\begin{array}{cc}1 & 0\\ -2\pi(\kappa_1-\kappa_2) & 1\end{array}\right).$$

The computation of the CRBs according to (C.5) and (C.7) finally yields:

$$\left(\begin{array}{cc}2\pi(k-\kappa_2) & 1\end{array}\right)(\mathbf{J}_2)^{-1}\left(\begin{array}{c}2\pi(k-\kappa_2)\\ 1\end{array}\right)\tag{C.12}$$

$$=\left(\begin{array}{cc}2\pi(k-\kappa_2) & 1\end{array}\right)\mathbf{K}(\mathbf{J}_1)^{-1}\mathbf{K}^T\left(\begin{array}{c}2\pi(k-\kappa_2)\\ 1\end{array}\right),\tag{C.13}$$

$$=\left(\begin{array}{cc}2\pi(k-\kappa_1) & 1\end{array}\right)(\mathbf{J}_1)^{-1}\left(\begin{array}{c}2\pi(k-\kappa_1)\\ 1\end{array}\right),\tag{C.14}$$

which shows that the CRBs $CRB_{\Theta_1(kT;\boldsymbol{\Psi}_1)}$ and $CRB_{\Theta_2(kT;\boldsymbol{\Psi}_2)}$ are equal.

Appendix D

In what follows, we compute the elements $M_{1,1} \equiv M_{\nu,\nu}$, $M_{1,2} \equiv M_{\nu,\theta} = M_{2,1} \equiv M_{\theta,\nu}$ and $M_{2,2} \equiv M_{\theta,\theta}$ of the MFIM related to the log-likelihood function $L(\mathbf{\Psi}) = p(\mathbf{r}\,|\mathbf{\Psi})$ of the parameter vector $\mathbf{\Psi} = (\Psi_1 \Psi_2)^T = (\nu\,\theta)^T$ based on the observation vector $\mathbf{r}$ from model (5.8)-(5.54).

Taking in (3.41) $\mathbf{u} = \mathbf{\Psi}$, $\mathbf{v} = \mathbf{a}^T$, and $2\sigma^2 = \frac{N_0}{E_s}$, we obtain for $M_{i,j}$, $(i,j) \in \{1,2\}^2$:

$$M_{i,j} = \frac{2E_s}{N_0} E_{\mathbf{\Psi},\mathbf{a}} \left[\Re \left\{ \mathbf{s}_i(\mathbf{\Psi},\mathbf{a})\, \mathbf{s}_j^H(\mathbf{\Psi},\mathbf{a}) \right\} \right], \tag{D.1}$$

where $\mathbf{s}_i(\mathbf{\Psi},\mathbf{a})$ is a shorthand notation for the derivative of $\mathbf{s}(\mathbf{\Psi},\mathbf{a})$ with respect to Ψ_i, i.e.,

$$\mathbf{s}_i(\mathbf{\Psi},\mathbf{a}) = \frac{\partial \mathbf{s}(\mathbf{\Psi},\mathbf{a})}{\partial \Psi_i}. \tag{D.2}$$

D.1 Simplified Observation Model

In the case of the simplified observation model, we have:

$$M_{i,j} = \frac{2E_s}{N_0} E_{\mathbf{\Psi},\mathbf{a}} \left[\Re \left\{ \sum_{k \in I_s} s_i\left(k;\mathbf{\Psi},\mathbf{a}\right) s_j^*\left(k;\mathbf{\Psi},\mathbf{a}\right) \right\} \right], \tag{D.3}$$

with $s_1\left(k;\mathbf{\Psi},\mathbf{a}\right) = s_\nu\left(k;\mathbf{\Psi},\mathbf{a}\right)$ and $s_2\left(k;\mathbf{\Psi},\mathbf{a}\right) = s_\theta\left(k;\mathbf{\Psi},\mathbf{a}\right)$ given by:

$$s_\nu\left(k;\mathbf{\Psi},\mathbf{a}\right) = j2\pi\left(k-\kappa_0\right) a\left(k\right) e^{j\theta} e^{j2\pi\nu(k-\kappa_0)}, \tag{D.4}$$

$$s_\theta\left(k;\mathbf{\Psi},\mathbf{a}\right) = j a\left(k\right) e^{j\theta} e^{j2\pi\nu(k-\kappa_0)}. \tag{D.5}$$

Substitution of (D.4)-(D.5) into (D.3) yields (after some elementary calculations):

$$M_{\theta,\theta} = \frac{2E_s}{N_0} \sum_{k \in I_s} \xi_{k,k}, \tag{D.6}$$

$$M_{\nu,\theta} = M_{\theta,\nu} = 2\pi M_{\theta,\theta} \frac{1}{N_s} \sum_{k \in I_s} \xi_{k,k}\left(k-\kappa_0\right), \tag{D.7}$$

$$M_{\nu,\nu} = 4\pi^2 M_{\theta,\theta} \frac{1}{N_s} \sum_{k \in I_s} \xi_{k,k}\left(k-\kappa_0\right)^2, \tag{D.8}$$

where $\xi_{k,l}$ is a short-hand notation for $E_{\mathbf{a}}\left[a^*\left(k\right) a\left(l\right)\right]$, i.e.,

$$\xi_{k,l} = \sum_{\tilde{\mathbf{a}}} \Pr\left[\mathbf{a} = \tilde{\mathbf{a}}\right] \left(\tilde{\mathbf{a}}\right)_k^* \left(\tilde{\mathbf{a}}\right)_l, \tag{D.9}$$

where $\Pr\left[\mathbf{a}\right]$ denotes the priori probability mass function of $\mathbf{a}$.

When $\Pr\left[\mathbf{a}\right]$ satisfies (5.13), it is assumed (see chapter 2) that $\xi_{k,k} = 1$, for $k \in I_s$. The same holds when $\Pr\left[\mathbf{a}\right]$ satisfies (5.16) (PA NCA approximation) or (5.17) (NPA NCA approximation). By denoting the mean of the symbol index k over the set of indices I_s as $\kappa_{m,s}$, with

$$\kappa_{m,s} = \frac{1}{N_s} \sum_{k \in I_s} k$$

and noting that, for any set I_s of N_s consecutive indices k:

$$\sum_{k \in I_s} \left(k - \kappa_{m,s}\right)^2 = \frac{1}{12} N_s \left(N_s^2 - 1\right),$$

we obtain for the elements of the PA CA, the PA NCA and the NPA NCA MFIM:

$$M_{\theta,\theta} = \frac{2E_s}{N_0} N_s, \text{ PA CA, PA NCA, NPA NCA}, \tag{D.10}$$

$$\begin{aligned} M_{\nu,\theta} &= M_{\theta,\nu}, \\ &= 2\pi M_{\theta,\theta}\left(\kappa_{m,s}-\kappa_0\right), \text{ PA CA, PA NCA, NPA NCA,} \end{aligned} \tag{D.11}$$

$$\begin{aligned} &M_{\nu,\nu} \\ &= 4\pi^2 M_{\theta,\theta}\left(\frac{\left(N_s^2-1\right)}{12}+\left(\kappa_{m,s}-\kappa_0\right)^2\right), \text{ PA CA, PA NCA, NPA NCA.} \end{aligned} \tag{D.12}$$

Assuming that $\Pr[\mathbf{a}]$ satisfies (5.15) (DA approximation) yields $\xi_{k,k}=1$, for $k\in I_p\subset I_s$ and $\xi_{k,k}=0$, otherwise. By denoting the mean of the symbol index k over the set of indices $I_p\subset I_s$ as $\kappa_{m,p}$, with

$$\kappa_{m,p}=\frac{1}{N_p}\sum_{k\in I_p}k,$$

and by denoting the variance of the symbol index k over the set of indices $I_p\subset I_s$ as σ_p^2, with

$$\sigma_p^2=\frac{1}{N_p}\sum_{k\in I_p}\left(k-\kappa_{m,p}\right)^2, \tag{D.13}$$

we obtain for the elements of the DA MFIM:

$$M_{\theta,\theta} = \frac{2E_s}{N_0}N_p, \tag{D.14}$$

$$M_{\nu,\theta} = M_{\theta,F} = 2\pi M_{\theta,\theta}\left(\kappa_{m,p}-\kappa_0\right), \tag{D.15}$$

$$M_{\nu,\nu} = 4\pi^2 M_{\theta,\theta}\left(\sigma_p^2+\left(\kappa_{m,p}-\kappa_0\right)^2\right). \tag{D.16}$$

D.2 Correct Observation Model

In the case of the correct observation model, we have:

$$M_{i,j} = \frac{2E_s}{N_0}E_{\mathbf{\Psi},\mathbf{a}}\left[\Re\left\{\int s_i\left(t;\mathbf{\Psi},\mathbf{a}\right)s_j^*\left(t;\mathbf{\Psi},\mathbf{a}\right)\mathrm{d}t\right\}\right], \tag{D.17}$$

with the quantities $s_1\left(t;\mathbf{\Psi},\mathbf{a}\right)=s_\nu\left(t;\mathbf{\Psi},\mathbf{a}\right)$ and $s_2\left(t;\mathbf{\Psi},\mathbf{a}\right)=s_\theta\left(t;\mathbf{\Psi},\mathbf{a}\right)$ given by:

$$s_\nu\left(t;\mathbf{\Psi},\mathbf{a}\right)=-j2\pi\left(\frac{t}{T}-\kappa_0\right)\sum_{k\in I_s}a\left(k\right)p\left(t-kT\right)e^{j\theta}e^{j2\pi\nu\left(\frac{t}{T}-\kappa_0\right)}, \tag{D.18}$$

$$s_\theta\left(t;\mathbf{\Psi},\mathbf{a}\right)=-j\sum_{k\in I_s}a\left(k\right)p\left(t-kT\right)e^{j\theta}e^{j2\pi\nu\left(\frac{t}{T}-\kappa_0\right)}. \tag{D.19}$$

Substitution of (D.18)-(D.19) into (D.17) yields:

$$M_{\theta,\theta} = 2\frac{E_s}{N_0}\sum_{k,l\in I_s}\xi_{k,l}\int p\left(t-kT\right)p\left(t-lT\right)\mathrm{d}t, \tag{D.20}$$

$$\begin{aligned} M_{\nu,\theta}=M_{\theta,\nu} &= 4\pi\frac{E_s}{N_0}\sum_{k,l\in I_s}\xi_{k,l}\int\frac{t}{T}p\left(t-kT\right)p\left(t-lT\right)\mathrm{d}t \\ &\quad -2\pi\kappa_0 M_{\theta,\theta}, \end{aligned} \tag{D.21}$$

$$\begin{aligned} M_{\nu,\nu} &= 8\pi^2\frac{E_s}{N_0}\sum_{k,l\in I_s}\xi_{k,l}T\int\left(\frac{t}{T}\right)^2 p\left(t-kT\right)p\left(t-lT\right)\mathrm{d}t \\ &\quad -4\pi\kappa_0\left(M_{\nu,\theta}+\pi\kappa_0 M_{\theta,\theta}\right), \end{aligned} \tag{D.22}$$

where $\xi_{k,l}=E_{\mathbf{a}}\left[a^*\left(k\right)a\left(l\right)\right]=\sum_{\tilde{\mathbf{a}}}\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}\right]\left(\tilde{\mathbf{a}}\right)_k^*\left(\tilde{\mathbf{a}}\right)_l$.

- Note that the pulse $\int p\left(t-kT\right)p\left(t-\left(k+m\right)T\right)\mathrm{d}t$ is nothing but the Nyquist pulse $g\left(mT\right)$ from (2.14), such that:

$$\int p\left(t-kT\right)p\left(t-\left(k+m\right)T\right)dt=\delta_m.$$

  This yields (see (D.20)):

$$M_{\theta,\theta} = 2\frac{E_s}{N_0}\sum_{k\in I_s}\xi_{k,k}, \tag{D.23}$$

  such that the entry $M_{\theta,\theta}$ of the PA CA, the PA NCA or the NPA NCA MFIM is given by ($\xi_{k,k}$ equals 1 for all $k\in I_s$):

$$M_{\theta,\theta} = 2\frac{E_s}{N_0}N_s, \text{ PA CA, PA NCA, NPA NCA}, \tag{D.24}$$

  while the entry $M_{\theta,\theta}$ of the DA MFIM is given by ($\zeta_{k,k}$ equals 1 for $k\in I_p\subset I_s$, and 0 otherwise):

$$M_{\theta,\theta} = 2\frac{E_s}{N_0}N_p, \text{ DA}. \tag{D.25}$$

- Taking into account that the function $p\left(t-kT\right)p\left(t-\left(k+m\right)T\right)$ is symmetric with respect to $t=\left(k+m/2\right)T$, it follows that:

$$\begin{aligned} &\int\frac{t}{T}p\left(t-kT\right)p\left(t-\left(k+m\right)T\right)\mathrm{d}t \\ &=\left(k+\frac{m}{2}\right)\underbrace{\int p\left(t-kT\right)p\left(t-\left(k+m\right)T\right)\mathrm{d}t}_{=\delta_m}, \\ &=k\delta_m. \end{aligned}$$

Using this equality, we find for the non-diagonal elements of the MFIM (see (D.21)):

$$M_{\nu,\theta} = M_{\theta,\nu} \quad = \quad 4\pi \frac{E_s}{N_0} \sum_{k \in I_s} \zeta_{k,k} k - 2\pi\kappa_0 M_{\theta,\theta}, \tag{D.26}$$

such that the entry $M_{\nu,\theta} = M_{\theta,\nu}$ of the PA CA, the PA NCA or the NPA NCA MFIM is given by:

$$M_{\nu,\theta} = M_{\theta,\nu} = 2\pi M_{\theta,\theta} \left(\kappa_{m,s} - \kappa_0\right), \text{ PA CA, PA NCA, NPA NCA,} \tag{D.27}$$

where $\kappa_{m,s} = \frac{1}{N_s} \sum_{k \in I_s} k$, while the entry $M_{\nu,\theta} = M_{\theta,\nu}$ of the DA MFIM is given by:

$$M_{\nu,\theta} = M_{\theta,\nu} \quad = \quad 2\pi M_{\theta,\theta} \left(\kappa_{m,p} - \kappa_0\right), \text{ DA,} \tag{D.28}$$

where $\kappa_{m,p} = \frac{1}{N_p} \sum_{k \in I_p} k$.

- Performing the variable substitution $\kappa = \kappa' + \left(k + \frac{m}{2}\right)$, it is easily shown that:

$$\begin{aligned} &\int \left(\frac{t}{T}\right)^2 p\left(t - kT\right) p\left(t - (k+m)\,T\right) \mathrm{d}t \\ &= \int \left(\frac{t}{T} + k + \frac{m}{2}\right)^2 p\left(t + \frac{mT}{2}\right) p\left(t - \frac{mT}{2}\right) \mathrm{d}t. \end{aligned}$$

Now, because $p\left(t + \frac{m}{2}T\right) p\left(t - \frac{m}{2}T\right)$ is symmetric with respect to $t = 0$, it follows that

$$\begin{aligned} &\int \left(\frac{t}{T}\right)^2 p\left(t - kT\right) p\left(t - (k+m)\,T\right) \mathrm{d}t \\ &= \mathrm{f}\left(lT\right) + \left(k + \frac{l}{2}\right)^2 \underbrace{\int p\left(t + \frac{mT}{2}\right) p\left(t - \frac{mT}{2}\right) \mathrm{d}t}_{=\delta_m}, \\ &= \mathrm{f}\left(mT\right) + k^2 \delta_m, \end{aligned}$$

where the function $\mathrm{f}\left(t\right)$ is defined as :

$$\mathrm{f}\left(t\right) = \int \tilde{t}^2 p\left(\tilde{t} + \frac{t}{2}\right) p\left(\tilde{t} - \frac{t}{2}\right) \mathrm{d}\tilde{t}.$$

As a result, the quantity $M_{\nu,\nu}$ from (D.22) can be rewritten as:

$$M_{\nu,\nu} = \delta M_{\nu,\nu} + 8\pi^2 \frac{E_s}{N_0} \sum_{k \in I_s} \xi_{k,k} k^2 - 4\pi\kappa_0 \left(M_{\nu,\theta} + \pi\kappa_0 M_{\theta,\theta}\right), \tag{D.29}$$

with

$$\delta M_{\nu,\nu} = 8\pi^2 \frac{E_s}{N_0} \sum_{k,l \in I_s} \xi_{k,l} \mathrm{f}\left((k-l)T\right). \quad \text{(D.30)}$$

When $\Pr[\mathbf{a}]$ satisfies (5.13) (correct PA CA expression), we have $\xi_{k,l} = \delta_{k-l}$ when both $a(k)$ and $a(l)$ are unknown data symbols. We obtain $\xi_{k,l} = 0$ when $a(k)$ is an unknown data symbol while $a(l)$ is a known pilot symbol (or vice versa) and $\xi_{k,l} = a^*(k)\, a(l)$ (with $\xi_{k,k} = \zeta_{l,l} = 1$) when both $a(k)$ and $a(l)$ are known pilot symbols. The same holds when $\Pr[\mathbf{a}]$ satisfies (5.16) (PA NCA approximation) or (5.17) (NPA NCA approximation). We thus obtain for the element $M_{\nu,\nu}$ of the PA CA, the PA NCA and the NPA NCA MFIM:

$$M_{\nu,\nu} = \delta M_{\nu,\nu} + M_{\nu,\nu}^{(si)}, \quad \text{(D.31)}$$

with

$$\begin{aligned} M_{\nu,\nu}^{(si)} &= 4\pi^2 M_{\theta,\theta} \frac{1}{N_s} \sum_{k \in I_s} (k-\kappa_0)^2, \quad \text{(D.32)} \\ &= 4\pi^2 M_{\theta,\theta} \left(\frac{\left(N_s^2-1\right)}{12} + (\kappa_{m,s}-\kappa_0)^2 \right), \end{aligned}$$

where $\kappa_{m,s} = \frac{1}{N_s} \sum_{k \in I_s} k$, and

$$\delta M_{\nu,\nu} = \frac{8\pi^2 E_s}{N_0} \left(N_s \mathrm{f}(0) + 2 \sum_{k \in I_s} \sum_{m=1}^{N_s-1} \Xi_{k,m} \mathrm{f}(mT) \right), \quad \text{(D.33)}$$

where $\Xi_{k,m}$ is defined as follows:

$$\Xi_{k,m} = \Re\left\{\xi_{k,k+m}\right\} I\left[(k+m) \in I_s\right],$$

with $I[P] = 1$ when P is true and $I[P] = 0$ when P is false. In arriving at (D.33), we have taken into account that the pulse $\mathrm{f}(t)$ is even in t. When $\Pr[\mathbf{a}]$ satisfies (5.15) (DA approximation), we have $\xi_{k,l} = 0$ when $a(k)$ or $a(l)$ (or both $a(k)$ and $a(l)$) are unknown data symbols, and $\xi_{k,l} = a^*(k)\, a(l)$ (with $\xi_{k,k} = \zeta_{l,l} = 1$) when both $a(k)$ and $a(l)$ are known pilot symbols. We obtain for the entry $M_{\nu,\nu}$ of the DA MFIM:

$$M_{\nu,\nu} = \delta M_{\nu,\nu} + M_{\nu,\nu}^{(si)}, \quad \text{(D.34)}$$

with

$$\begin{aligned} M_{\nu,\nu}^{(si)} &= 4\pi^2 M_{\theta,\theta} \frac{1}{N_p} \sum_{k \in I_p} (k-\kappa_0)^2, \\ &= 4\pi^2 M_{\theta,\theta} \left(\sigma_p^2 + (\kappa_{m,p}-\kappa_0)^2 \right), \quad \text{(D.35)} \end{aligned}$$

where $\kappa_{m,p} = \frac{1}{N_p}\sum_{k\in I_p} k$ and $\sigma_p^2 = \frac{1}{N_p}\sum_{k\in I_p}(k-\kappa_{m,p})^2$, and

$$\delta M_{\nu,\nu} = \frac{8\pi^2 E_s}{N_0}\left(N_p \mathrm{f}(0) + 2\sum_{k\in I_p}\sum_{m=1}^{N_s-1}\Xi_{k,m}\mathrm{f}(mT)\right), \quad \text{(D.36)}$$

where

$$\Xi_{k,m} = \Re\left\{a^*(k)\,a(k+m)\right\} I\left[(k+m)\in I_p\right].$$

D.3 Remark

The MFIM entries $M_{\theta,\theta}$ (see (D.10),(D.24) and (D.14),(D.25)) and $M_{\nu,\theta} = M_{\theta,\nu}$ (see (D.11),(D.27) and (D.15),(D.28)) do not depend on the observation model. The MFIM entry $M_{\nu,\nu}$ resulting from the correct observation model (see (D.32) and (D.35)) is equal to the sum of the MFIM entry $M_{\nu,\nu}^{(si)}$ resulting from the simplified observation model (see (D.12) and (D.16)) and an additional term $\delta M_{\nu,\nu}$ (see (D.33) and (D.36)). We will now show that this additional term can be neglected for a large number of transmitted symbols N_s (in the case of the PA CA, the PA NCA and the NPA NCA estimation modes), or a large number of pilot symbols (in the case of the DA estimation mode).

D.3.1 PA CA, PA NCA and NPA NCA Estimation Modes

We obtain a valid upper bound on $\delta M_{\nu,\nu}$ from (D.33) by assuming that $\Xi_{k,m} = 1$ for all $m = 1, ..., N_s - 1$ and all $k \in I_s$: $\delta M_{\nu,\nu} \leq \bar{\delta} M_{\nu,\nu}$, with

$$\bar{\delta} M_{\nu,\nu} = 8\pi^2 \frac{E_s}{N_0}\left(N_s \mathrm{f}(0) + 2N_s C\right), \text{PA CA, PA NCA, NPA NCA} \quad \text{(D.37)}$$

and C given by:

$$\begin{aligned} C &= \sum_{m=1}^{N_s-1}\mathrm{f}(mT), \\ &\approx \sum_{m=1}^{\infty}\mathrm{f}(mT), \end{aligned}$$

where the approximation is valid when the observation interval N_sT is much larger than the duration of $\mathrm{f}(t)$ (which amounts to a few symbol intervals), in which case C becomes independent of N_s. Hence, for large N_s, the bound $\bar{\delta} M_{\nu,\nu}$ is proportional to N_s because $\sum_{m=1}^{\infty}\mathrm{f}(mT)$ is a constant. It then follows from (D.37) that the first term in (D.33), for $N_s \gg 1$, is upper bounded by a function that is linear in N_s. On the other hand, it follows directly from (D.32) that the second term in (D.24) is proportional to $(N_s)^3$ for large N_s, i.e., $N_s \gg 1$. Taking this into account it follows that, in (D.32), the first term can be neglected as compared to the second term, or equivalently, the MFIM entry $M_{\nu,\nu}$ resulting from the correct observation model can be approximated by the MFIM entry $M_{\nu,\nu}^{(si)}$ resulting from the simplified observation model, for large values of N_s.

D.3.2 DA Estimation Mode

A similar result can be derived for the DA case. We will first consider the special case of N_p *consecutive* pilot symbols. In this case, (D.35) and (D.36) reduce to:

$$M_{\nu,\nu}^{(si)} = 4\pi^2 M_{\theta,\theta}\left(\frac{\left(N_p^2-1\right)}{12}+\left(\kappa_{m,p}-\kappa_0\right)^2\right), \tag{D.38}$$

$$\delta M_{\nu,\nu} = 8\pi^2\frac{E_s}{N_0}\left(N_p \mathrm{f}\left(0\right)+2\sum_{k\in I_p}\sum_{m=1}^{N_p-1}\Xi_{k,m}\mathrm{f}\left(mT\right)\right), \tag{D.39}$$

with

$$\Xi_{k,m}=\Re\left\{a^*\left(k\right)a\left(k+m\right)\right\}I\left[\left(k+m\right)\in I_p\right].$$

Note the similarities between (D.32),(D.33) and (D.38),(D.39). Replacing, in the discussion for the PA CA, PA NCA and NPA NCA estimation modes (previous section), $(N_s, I_s, \kappa_{m,s})$ with $(N_p, I_p, \kappa_{m,p})$ we obtain that (D.38) is proportional to $(N_p)^3$, whereas (D.39) is proportional to N_p for $N_p \gg 1$. As a consequence, in (D.34) the first term can be neglected as compared to the second term, or equivalently, the DA MFIM entry $M_{\nu,\nu}$ resulting from the correct observation model can be approximated by the DA MFIM entry $M_{\nu,\nu}^{(si)}$ resulting from the simplified observation model, for N_p consecutive pilot symbols and large values of N_p (i.e., $N_p \gg 1$).

Let us now reconsider the general case. For a given number of pilot symbols N_p, the quantity (D.35) achieves its *minimum* value when all pilot symbols are transmitted consecutively (not separated by any data symbols). This minimum value is given by (D.38). On the other hand, for large values of N_p, the quantity (D.33) may be safely replaced by (D.39). This is because the pulse $\mathrm{f}\left(t\right)$ becomes negligibly small for values of t larger than or equal to N_pT, when $N_p \gg 1$. Consequently, since (D.38) dominates (D.39) for large N_p the first term in (D.34) always dominates the second term for large values of N_p (*no matter what* the *pilot symbol insertion rule* is).

Appendix E

Assuming that $\mathbf{r} = \mathbf{s}(\mathbf{a}, \mathbf{\Psi}) + \mathbf{w}$, with $\mathbf{w}$ Gaussian noise with $E\left[\mathbf{w}^H \mathbf{w}\right] = (N_0/E_s)\,\mathbf{I}$ and $(\mathbf{a}, \mathbf{\Psi})$ uniformly distributed over the domain $(\zeta_{\mathcal{A}}, [-\nu_{max}, \nu_{max}] \times [-\pi, \pi])$, we obtain for the joint symbol APPs $\Pr[\mathbf{a} = \tilde{\mathbf{a}} \,|\mathbf{r}, \mathbf{\Psi}]$, $\tilde{\mathbf{a}} \in \zeta_{\mathcal{A}}$:

$$\begin{aligned} \Pr[\mathbf{a} = \tilde{\mathbf{a}} \,|\mathbf{r}, \mathbf{\Psi}] &\propto \frac{p(\mathbf{r} \,|\mathbf{a} = \tilde{\mathbf{a}}, \mathbf{\Psi})}{\sum_{\check{\mathbf{a}} \in \zeta_{\mathcal{A}}} p(\mathbf{r} \,|\mathbf{a} = \check{\mathbf{a}}, \mathbf{\Psi})}, \\ &= \frac{e^{-\frac{E_s}{N_0}|\mathbf{r} - \mathbf{s}(\tilde{\mathbf{a}}, \mathbf{\Psi})|^2}}{\sum_{\check{\mathbf{a}} \in \zeta_{\mathcal{A}}} e^{-\frac{E_s}{N_0}|\mathbf{r} - \mathbf{s}(\check{\mathbf{a}}, \mathbf{\Psi})|^2}}, \end{aligned}$$

where $\check{\mathbf{a}} \in \zeta_{\mathcal{A}}$ denotes summation over all legitimate symbol sequences. For large E_s/N_0, this converges to:

$$\Pr[\mathbf{a} = \tilde{\mathbf{a}} \,|\mathbf{r}, \mathbf{\Psi}] = \begin{cases} 1 & , \tilde{\mathbf{a}} = \hat{\mathbf{a}}(\mathbf{r}, \mathbf{\Psi}) \\ 0 & , \tilde{\mathbf{a}} \neq \hat{\mathbf{a}}(\mathbf{r}, \mathbf{\Psi}) \end{cases},$$

with

$$\hat{\mathbf{a}}(\mathbf{r}, \mathbf{\Psi}) = \arg \min_{\check{\mathbf{a}} \in \zeta_{\mathcal{A}}} |\mathbf{r} - \mathbf{s}(\check{\mathbf{a}}, \mathbf{\Psi})|^2 .$$

This is illustrated by Fig. E.1 showing the symbol APP for the simple example

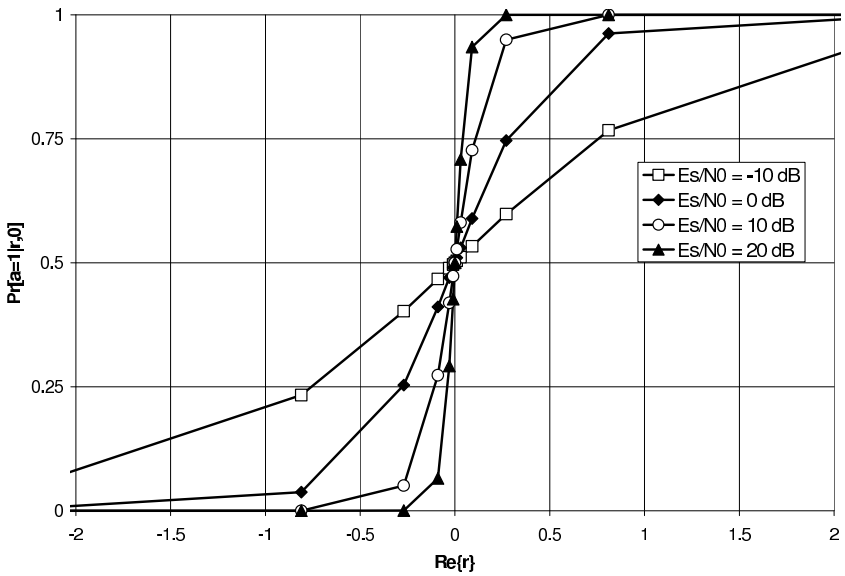

Figure E.1: Symbol APP $\Pr[a=1\,|r,\boldsymbol{\Psi}=\mathbf{0}]$ when $\boldsymbol{\Psi}=\mathbf{0}$, $r=a+w$ and $a\in\{-1,1\}$, as a function of $\Re\{r\}$.

where $\boldsymbol{\Psi}=\mathbf{0}$ and $r=a+w$, with $a\in\{-1,1\}$. In this case, one obtains:

$$\hat{a}\,(r,\mathbf{0})=\begin{cases}1 & ,\,\Re\{r\}>0\\ -1 & ,\,\Re\{r\}<0\end{cases}$$

and $\Pr[a=-1\,|r,\boldsymbol{\Psi}=\mathbf{0}]=1-\Pr[a=1\,|r,\boldsymbol{\Psi}=\mathbf{0}]$, with

$$\Pr[a=1\,|r,\boldsymbol{\Psi}=\mathbf{0}]=\frac{1}{1+\exp\left(-4\frac{E_s}{N_0}\Re\{r\}\right)}. \tag{E.1}$$

We recall that the FIM entries are given by

$$\begin{aligned}&J_{i,j}\\&=E_{\mathbf{r},\boldsymbol{\Psi}}\left[\ell_i\left(\boldsymbol{\Psi};\mathbf{r}\right)\ell_j\left(\boldsymbol{\Psi};\mathbf{r}\right)\right],\\&=\int\sum_{\check{\mathbf{a}}\in\zeta_{\mathcal{A}}}\Pr[\mathbf{a}=\check{\mathbf{a}}]\,I_{i,j}\left(\boldsymbol{\Psi};\check{\mathbf{a}}\right)p\left(\boldsymbol{\Psi}\right)\mathrm{d}\boldsymbol{\Psi},\end{aligned}$$

with

$$\begin{aligned}&I_{i,j}\left(\boldsymbol{\Psi};\check{\mathbf{a}}\right)\\&=\int\ell_i\left(\boldsymbol{\Psi};\mathbf{r}\right)\ell_j\left(\boldsymbol{\Psi};\mathbf{r}\right)p\left(\mathbf{r}\,|\mathbf{a}=\check{\mathbf{a}},\boldsymbol{\Psi}\right)\mathrm{d}\mathbf{r}\end{aligned}$$

and

$$\ell_i\left(\mathbf{\Psi};\mathbf{r}\right)=\sum_{\tilde{\mathbf{a}}\in\zeta_{\mathcal{A}}}\frac{\partial\ln p\left(\mathbf{r}\left|\tilde{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_i}\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}\left|\mathbf{r},\mathbf{\Psi}\right.\right]. \tag{E.2}$$

For given $(\mathbf{a}=\check{\mathbf{a}},\mathbf{\Psi})$, the vector $\mathbf{r}$ is distributed according to $p\left(\mathbf{r}\left|\mathbf{a}=\check{\mathbf{a}},\mathbf{\Psi}\right.\right)$. For large E_s/N_0, $\mathbf{r}$ converges to $\mathbf{s}\left(\check{\mathbf{a}},\mathbf{\Psi}\right)$, such that $\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}\left|\mathbf{r},\mathbf{\Psi}\right.\right]$ converges to

$$\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}\left|\mathbf{r},\mathbf{\Psi}\right.\right] \quad = \quad \begin{cases} 1 & ,\,\tilde{\mathbf{a}}=\check{\mathbf{a}} \\ 0 & ,\,\tilde{\mathbf{a}}\neq\check{\mathbf{a}} \end{cases} \tag{E.3}$$

This is illustrated by Fig. E.2 which shows the symbol APP $\Pr\left[a=1\left|r,\mathbf{\Psi}=\mathbf{0}\right.\right]$ (E.1) as a function of E_s/N_0 when r is generated according to $p\left(r\left|a=1,\mathbf{\Psi}=\mathbf{0}\right.\right)$. From (E.3) and (E.2) it follows that, for given $(\mathbf{a}=\check{\mathbf{a}},\mathbf{\Psi})$ and increasing E_s/N_0:

$$\ell_i\left(\mathbf{\Psi};\mathbf{r}\right)\approx\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_i}$$

such that

$$\begin{aligned} &I_{i,j}\left(\mathbf{\Psi};\check{\mathbf{a}}\right) \\ &\approx\int\mathrm{d}\mathbf{r}\left(\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_i}\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_j}\right)p\left(\mathbf{r}\left|\mathbf{a}=\check{\mathbf{a}},\mathbf{\Psi}\right.\right) \\ &=E_{\mathbf{r}\left|\mathbf{a}=\check{\mathbf{a}},\mathbf{\Psi}\right.}\left[\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_i}\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_j}\right] \end{aligned}$$

and finally

$$\begin{aligned} &J_{i,j} \\ &\approx E_{\mathbf{r},\mathbf{a},\mathbf{\Psi}}\left[\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_i}\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\mathbf{\Psi}\right.\right)}{\partial\Psi_j}\right] \\ &=M_{i,j} \end{aligned}$$

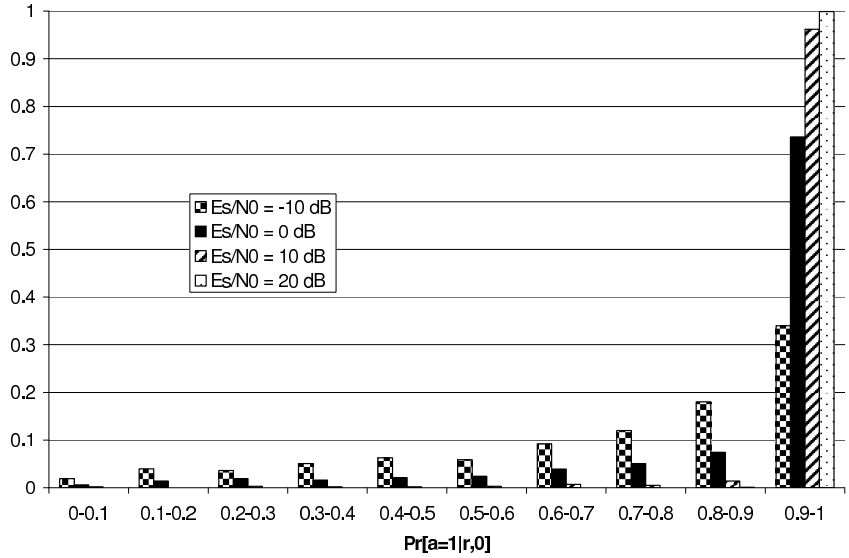

Figure E.2: Histogram of the symbol APP $\Pr[a = 1 \,|r, \mathbf{\Psi} = \mathbf{0}]$ when $\mathbf{\Psi} = \mathbf{0}$ and $r = 1 + w$.

Appendix F

F.1 Statistical Properties of z, $\frac{\partial \mathbf{z}}{\partial \nu}$ and $\frac{\partial \mathbf{z}}{\partial \theta}$ for Given $(\mathbf{a}, \nu, \theta)$

For both observation models, the quantities $\mathbf{z} = (z(k):\ k \in I_s)$, $\mathbf{z}_i = (z_i(k):\ k \in I_s)$ are given by:

$$\mathbf{z} = \mathbf{z}(\mathbf{r}, \mathbf{\Psi}) \quad = \quad \mathbf{r}\mathbf{S}^H(\mathbf{\Psi}), \tag{F.1}$$

$$\mathbf{z}_i = \mathbf{z}_i(\mathbf{r}, \mathbf{\Psi}) \quad = \quad \frac{\partial \mathbf{z}(\mathbf{r}, \mathbf{\Psi})}{\partial \Psi_i}. \tag{F.2}$$

From $\mathbf{r} = \mathbf{a}\mathbf{S}(\mathbf{\Psi}) + \mathbf{w}$ with $E\left[\mathbf{w}^H\mathbf{w}\right] = (N_0/E_s)\,\mathbf{I}$, it follows that for a given value of $(\mathbf{a}, \mathbf{\Psi})$,

$$\mathbf{z} \quad = \quad \mathbf{a}\mathbf{S}(\mathbf{\Psi})\,\mathbf{S}^H(\mathbf{\Psi}) + \mathbf{n}, \tag{F.3}$$

$$\mathbf{z}_i \quad = \quad \mathbf{a}\mathbf{S}(\mathbf{\Psi})\,\mathbf{S}_i^H(\mathbf{\Psi}) + \mathbf{n}_i, \tag{F.4}$$

with

$$\mathbf{n} \quad = \quad \mathbf{w}\mathbf{S}^H(\mathbf{\Psi}), \tag{F.5}$$

$$\mathbf{n}_i = \mathbf{w}\mathbf{S}_i^H(\mathbf{\Psi}), \tag{F.6}$$

where $\mathbf{S}_i(\mathbf{\Psi})$ is a short-hand notation for the derivative of $\mathbf{S}(\mathbf{\Psi})$ with respect to Ψ_i. Taking into account that $\mathbf{S}(\mathbf{\Psi})\mathbf{S}^H(\mathbf{\Psi}) = \mathbf{I}$, it is easily found that:

$$\mathbf{z} = \mathbf{a} + \mathbf{n}, \tag{F.7}$$

with

$$E\left[\mathbf{n}^H\mathbf{n}\right] = (N_0/E_s)\mathbf{I}. \tag{F.8}$$

Note that the statistics of $\mathbf{z}$, for a given $(\mathbf{a}, \mathbf{\Psi})$, do not depend on $\mathbf{\Psi}$.

F.1.1 Simplified Observation Model

For the simplified observation model, the quantities $z(k)$, $z_1(k) = z_\nu(k) = \frac{\partial z(k)}{\partial \nu}$ and $z_2(k) = z_\theta(k) = \frac{\partial z(k)}{\partial \theta}$, for $k \in I_s$, are given by:

$$z(k) = r(k)\, e^{-j\theta} e^{-j2\pi\nu(k-\kappa_0)}, \tag{F.9}$$

$$z_\theta(k) = \frac{\partial z(k)}{\partial \theta} = -jz(k), \tag{F.10}$$

$$z_\nu(k) = \frac{\partial z(k)}{\partial \nu} = -j2\pi(k-\kappa_0)z(k). \tag{F.11}$$

The quantities $z_\nu(k)$ and $z_\theta(k)$ are deterministic (linear) functions of $z(k)$.

F.1.2 Correct Observation Model

For the correct observation model, the components $z(k)$, $k \in I_s$ of the vector $\mathbf{z}$, and the quantities $z_1(k) = z_\nu(k) = \frac{\partial z(k)}{\partial \nu}$ and $z_2(k) = z_\theta(k) = \frac{\partial z(k)}{\partial \theta}$, $k \in I_s$ are given by:

$$z(k) = \frac{1}{\sqrt{E_s}} \int r(t)\, p(t-kT)\, e^{-j\theta} e^{-j2\pi\nu\left(\frac{t}{T}-\kappa_0\right)} \mathrm{d}t, \tag{F.12}$$

$$z_\theta(k) = \frac{\partial z(k)}{\partial \theta} = -jz(k), \tag{F.13}$$

$$\begin{aligned} z_\nu(k) &= \frac{\partial z(k)}{\partial \nu} \\ &= \frac{-j2\pi}{\sqrt{E_s}} \int \left(\frac{t}{T} - \kappa_0\right) r(t)\, p(t-kT)\, e^{-j\theta} \mathrm{e}^{-\mathrm{j}2\pi\nu\left(\frac{t}{T}-\kappa_0\right)} \mathrm{d}t, \end{aligned} \tag{F.14}$$

where we have used the continuous time signal model:

$$r(t) = \sqrt{E_s} \sum_n a(n)\, p(t-nT)\, e^{j\theta} e^{j2\pi\nu\left(\frac{t}{T}-\kappa_0\right)} + w(t), \tag{F.15}$$

with $E\left[w^*\left(t+\tau\right)w\left(t\right)\right] = \frac{N_0}{E_s}\delta_\tau$.

The quantity $z_\theta\left(k\right)$ is a deterministic (linear) function of $z\left(k\right)$. The quantity $z_\nu\left(k\right)$ is not a deterministic function of $z\left(k\right)$, but we might decompose the vector $\mathbf{z}_\nu = \left(z_\nu\left(k\right) : \, k \in I_s\right)$ as the sum of a deterministic linear function of the vector $\mathbf{z} = \left(z\left(k\right) : \, k \in I_s\right)$, and a vector $\delta\mathbf{z}_\nu$ that is uncorrelated with $\mathbf{z}$. We obtain:

$$\mathbf{z}_\nu = \mathbf{z}\mathbf{L} + \delta\mathbf{z}_\nu, \tag{F.16}$$

with

$$\mathbf{L} = Cov\left[\mathbf{z}_\nu, \mathbf{z}\right]\left(Cov\left[\mathbf{z}, \mathbf{z}\right]\right)^{-1}, \tag{F.17}$$

where $Cov\left[\mathbf{x}, \mathbf{y}\right] = E\left[\mathbf{x}\mathbf{y}^H\right] - E\left[\mathbf{x}\right]E\left[\mathbf{y}^H\right]$ denotes the covariance matrix of two vectors $\mathbf{x}$ and $\mathbf{y}$, and (F.17) follows directly from the constraint that $Cov\left[\delta\mathbf{z}_\nu, \mathbf{z}\right] = \mathbf{0}$. Using (F.16) and (F.17), the mean and the autocorrelation matrix of $\delta\mathbf{z}_\nu$ can be expressed in terms of the means and covariance matrices of $\mathbf{z}$ and $\mathbf{z}_\nu$. This yields:

$$E\left[\delta\mathbf{z}_\nu\right] = \mathbf{m}_\nu - \mathbf{m}\mathbf{L}, \tag{F.18}$$

$$Cov\left[\delta\mathbf{z}_\nu, \delta\mathbf{z}_\nu\right] = Cov\left[\mathbf{z}_\nu, \mathbf{z}_\nu\right] - Cov\left[\mathbf{z}_\nu, \mathbf{z}\right]\left(Cov\left[\mathbf{z}, \mathbf{z}\right]\right)^{-1}Cov\left[\mathbf{z}, \mathbf{z}_\nu\right], \tag{F.19}$$

where $\mathbf{m}$ and $\mathbf{m}_\nu$ are short-hand notations for the statistical expectations of $\mathbf{z}$ and $\mathbf{z}_\nu$, respectively.

It follows directly from (F.7) and (F.8) that:

$$\mathbf{m} = \mathbf{a} \tag{F.20}$$

and

$$Cov\left[\mathbf{z}, \mathbf{z}\right] = \frac{N_0}{E_s}\mathbf{I}_{N_s}. \tag{F.21}$$

It will be shown in the next subsection that:

$$Cov\left[z_\nu\left(k\right), z\left(l\right)\right] = -j2\pi\left(k - \kappa_0\right)\frac{N_0}{E_s}\delta_{k-l}, \tag{F.22}$$

$$m_\nu\left(k\right) = -j2\pi k\, a\left(k\right), \tag{F.23}$$

$$Cov\left[z_\nu\left(k\right), z_\nu\left(l\right)\right] = \frac{N_0}{E_s}4\pi^2\left(\left(k - \kappa_0\right)^2\delta_{k-l} + \mathrm{f}\left(\left(k-l\right)T\right)\right), \tag{F.24}$$

with $\mathrm{f}\left(t\right) = \int\left(\frac{\tilde{t}}{T}\right)^2 p\left(\tilde{t} + \frac{t}{2}\right)p\left(\tilde{t} - \frac{t}{2}\right)\mathrm{d}\tilde{t}$.

Substituting (F.21) and (F.22) into (F.17) we find that $\mathbf{L}$ is a diagonal matrix with the elements of the vector $\left(-j2\pi k : \, k \in I_s\right)$ on the main diagonal. This will be denoted as follows:

$$\mathbf{L} = diag\left(-j2\pi k : \, k \in I_s\right). \tag{F.25}$$

Substituting (F.20), (F.23) and (F.25) into (F.18), and substituting (F.21), (F.24) and (F.22) into (F.19) further yields:

$$E\left[\delta\mathbf{z}_\nu\right] = \mathbf{0} \tag{F.26}$$

and

$$\begin{aligned} Cov\left[\delta z_\nu(k), \delta z_\nu(l)\right] &= Cov\left[z_\nu(k), z_\nu(l)\right], \\ &\quad -\frac{N_0}{E_s} 4\pi^2 (k-\kappa_0)^2 \delta_{k-l}, \end{aligned} \tag{F.27}$$

$$= \frac{N_0}{E_s} 4\pi^2 \mathrm{f}\left((k-l)T\right). \tag{F.28}$$

F.1.3 Computation of $\mathbf{m}_\nu$, $Cov\left[\mathbf{z}_\nu, \mathbf{z}\right]$ and $Cov\left[\mathbf{z}_\nu, \mathbf{z}_\nu\right]$ for the Correct Observation Model

Substitution of (F.15) into (F.12) and (F.14) yields:

$$z(k) = \sum_l a(l) \int p(t-kT)\, p(t-lT)\, \mathrm{d}t + n(k), \tag{F.29}$$

with

$$n(k) = \int w(t)\, p(t-kT)\, e^{j\theta} e^{j2\pi\nu\left(\frac{t}{T}-\kappa_0\right)} \mathrm{d}t \tag{F.30}$$

and

$$\begin{aligned} z_\nu(k) &= -j2\pi \sum_l a(l) \int \left(\frac{t}{T}-\kappa_0\right) p(t-kT)\, p(t-lT)\, \mathrm{d}t \\ &\quad + n_\nu(k), \end{aligned} \tag{F.31}$$

with

$$n_\nu(k) = -j2\pi \int w(t) \left(\frac{t}{T}-\kappa_0\right) p(t-kT)\, e^{j\theta} e^{j2\pi\nu\left(\frac{t}{T}-\kappa_0\right)} \mathrm{d}t. \tag{F.32}$$

From $E\left[w(t)\right] = 0$ it follows that $E\left[n(k)\right] = E\left[n_\nu(k)\right] = 0$. Taking into account that $p(t-kT)\, p(t-lT)$ is symmetric with respect to $\frac{(l+k)T}{2}$, and that $\int p(t-kT)\, p(t-lT)\, \mathrm{d}t = \delta_{k-l}$, we obtain:

$$m_\nu(k) = -j2\pi \sum_l a(l) \left(\frac{k+l}{2}-\kappa_0\right) \delta_{k-l}, \tag{F.33}$$

$$= -j2\pi a(k) (k-\kappa_0) \delta_{k-l} \tag{F.34}$$

and

$$Cov\left[z_\nu(k), z(l)\right] = j2\pi \frac{N_0}{E_s} \int \left(\frac{t}{T}-\kappa_0\right) p(t-kT)\, p(t-lT)\, \mathrm{d}t, \tag{F.35}$$

$$= j2\pi \frac{N_0}{E_s} (k-\kappa_0) \delta_{k-l}. \tag{F.36}$$

Finally, we have (similar to the computation of $M_{\nu,\nu}$ in Appendix D):

$$\begin{aligned} Cov\left[z_\nu(k), z_\nu(l)\right] &= 4\pi^2 \frac{N_0}{E_s} \\ &\quad \cdot \int \left(\frac{t}{T} - \kappa_0\right)^2 p(t-kT)\, p(t-lT)\, dt, \end{aligned} \tag{F.37}$$

$$= 4\pi^2 \frac{N_0}{E_s} \left(\mathrm{f}\left((k-l)T\right) + (k-\kappa_0)^2 \delta_{k-l}\right). \tag{F.38}$$

F.2 Analytical Computation of $H_{i,j}(k,l;\mathbf{z})$

We now compute the quantity $H_{i,j}(k,l;\mathbf{z})$ from (5.68), i.e.,

$$H_{i,j}(k,l;\mathbf{z}) = E_{z_i(k),z_j(l)|\mathbf{z}}\left[\Re\left\{z_i(k)\,\mu^*(k;\mathbf{z})\right\}\Re\left\{z_j(l)\,\mu^*(l;\mathbf{z})\right\}\right]. \tag{F.39}$$

Collecting the results from subsections F.1.1 and F.1.2 it may be concluded that, for a given $(\mathbf{\Psi}, \mathbf{a})$, the following representation can be used:

$$z_i(k) = d_i(k)\, z(k) + \delta z_i(k), \tag{F.40}$$

where (for both observation models), with $(\Psi_1, \Psi_2) = (\nu, \theta)$,

$$d_1(k) = -j2\pi(k-\kappa_0),$$

$$d_2(k) = -j$$

and $\delta z_i(k)$ equal to zero except for the variable $\Psi_1 = \nu$ in the correct observation model in which case:

$$E\left[\delta z_\nu(k)\right] = 0,$$

$$Cov\left[\delta z_\nu(k), \delta z_\nu(k+m)\right] = \frac{N_0}{E_s} 4\pi^2 \mathrm{f}(mT).$$

From (F.40) it follows that:

$$\begin{aligned} H_{i,j}(k,l;\mathbf{z}) &= \frac{1}{2}\Re\left\{\mu^*(k;\mathbf{z})\,\mu^*(l;\mathbf{z})\,E\left[z_i(k)\,z_j(l)\,|\mathbf{z}\right]\right\} \\ &\quad + \frac{1}{2}\Re\left\{\mu(k;\mathbf{z})\,\mu^*(l;\mathbf{z})\,E\left[z_i^*(k)\,z_j(l)\,|\mathbf{z}\right]\right\}, \end{aligned} \tag{F.41}$$

where the correlations $E\left[z_i(k)\,z_j(l)\,|\mathbf{z}\right]$ and $E\left[z_i^*(k)\,z_j(l)\,|\mathbf{z}\right]$ of $z_i(k)$ and $z_j(l)$, conditioned on $\mathbf{z}$, are given by:

$$E\left[z_i(k)\,z_j(l)\,|\mathbf{z}\right] = d_i(k)\,d_j(l)\,z(k)\,z(l), \tag{F.42}$$

$$E\left[z_i^*(k)\,z_j(l)\,|\mathbf{z}\right] = d_i^*(k)\,d_j(l)\,z^*(k)\,z(l) + Cov\left[\delta z_i(k), \delta z_j(l)\right] \tag{F.43}$$

Appendix G

When $\Pr[\mathbf{a}]$ satisfies (5.15) (DA approximation), we have:

$$\Pr[\mathbf{a}=\tilde{\mathbf{a}}] \;=\; \begin{cases} 1 & ,\tilde{\mathbf{a}}=\boldsymbol{\alpha} \\ 0 & ,\tilde{\mathbf{a}}\neq\boldsymbol{\alpha} \end{cases}, \tag{G.1}$$

where $\boldsymbol{\alpha}=(\alpha(k):k\in I_s)$, with $\alpha(k)=A_k$, i.e., the actual value of the pilot symbol $a(k)$ for $k\in I_p$, and $\alpha(k)=0$ for $k\in I_d$. Consequently, the MFIM entries are given by:

$$\begin{aligned} M_{i,j} &= E_{\mathbf{r},\mathbf{a},\boldsymbol{\Psi}}\left[\frac{\partial \ln p(\mathbf{r}|\mathbf{a},\boldsymbol{\Psi})}{\partial\Psi_i}\frac{\partial \ln p(\mathbf{r}|\mathbf{a},\boldsymbol{\Psi})}{\partial\Psi_j}\right], \\ &= E_{\mathbf{r},\boldsymbol{\Psi}}\left[\frac{\partial \ln p(\mathbf{r}|\mathbf{a}=\boldsymbol{\alpha},\boldsymbol{\Psi})}{\partial\Psi_i}\frac{\partial \ln p(\mathbf{r}|\mathbf{a}=\boldsymbol{\alpha},\boldsymbol{\Psi})}{\partial\Psi_j}\right]. \end{aligned} \tag{G.2}$$

At the same time, we obtain:

$$\Pr[\mathbf{a}=\tilde{\mathbf{a}}|\mathbf{r},\boldsymbol{\Psi}] \;=\; \begin{cases} 1 & ,\tilde{\mathbf{a}}=\boldsymbol{\alpha} \\ 0 & ,\tilde{\mathbf{a}}\neq\boldsymbol{\alpha} \end{cases}, \tag{G.3}$$

such that

$$
\begin{aligned}
&\ell_i\left(\boldsymbol{\Psi};\mathbf{r}\right)\\
&=\sum_{\tilde{\mathbf{a}}\in\zeta_{\mathcal{A}}}\frac{\partial\ln p\left(\mathbf{r}\left|\mathbf{a}=\boldsymbol{\alpha},\boldsymbol{\Psi}\right.\right)}{\partial\Psi_i}\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}\left|\mathbf{r},\boldsymbol{\Psi}\right.\right],\\
&=\frac{\partial\ln p\left(\mathbf{r}\left|\mathbf{a}=\boldsymbol{\alpha},\boldsymbol{\Psi}\right.\right)}{\partial\Psi_i}
\end{aligned}
$$

and

$$
\begin{aligned}
&J_{i,j}\\
&=E_{\mathbf{r},\boldsymbol{\Psi}}\left[\ell_i\left(\boldsymbol{\Psi};\mathbf{r}\right)\ell_j\left(\boldsymbol{\Psi};\mathbf{r}\right)\right],\\
&=E_{\mathbf{r},\boldsymbol{\Psi}}\left[\frac{\partial\ln p\left(\mathbf{r}\left|\mathbf{a}=\boldsymbol{\alpha},\boldsymbol{\Psi}\right.\right)}{\partial\Psi_i}\frac{\partial\ln p\left(\mathbf{r}\left|\mathbf{a}=\boldsymbol{\alpha},\boldsymbol{\Psi}\right.\right)}{\partial\Psi_j}\right],\\
&=M_{i,j}.
\end{aligned}
$$

Appendix H

H.1 Further Computation of $\mu\left(k;\mathbf{z}\left(\mathbf{r},\mathbf{\Psi}\right)\right)$

According to (5.65), we have:

$$\mu\left(k;\mathbf{z}\right) \quad = \quad \sum_{\omega\in\Omega}\Pr\left[a\left(k\right)=\omega\left|\mathbf{z}\right.\right]\omega, \tag{H.1}$$

where $\mathbf{z}$ is a short-hand notation for $\mathbf{z}\left(\mathbf{r},\mathbf{\Psi}\right)$. Since the statistics of $\mathbf{z}$, for a given $\left(\mathbf{a},\mathbf{\Psi}\right)$, do not depend on $\mathbf{\Psi}$ (see Appendix F), the following holds:

$$\mu\left(k;\mathbf{z}\right) \quad = \quad \frac{\sum_{\mathbf{a}\in\Omega^{N_s}}a\left(k\right)\Pr\left[\mathbf{a}\right]p\left(\mathbf{z}\left|\mathbf{a},\mathbf{\Psi}\right.\right)}{\sum_{\mathbf{a}\in\Omega^{N_s}}\Pr\left[\mathbf{a}\right]p\left(\mathbf{z}\left|\mathbf{a},\mathbf{\Psi}\right.\right)}. \tag{H.2}$$

Appendix E further shows that the conditional pdf $p\left(\mathbf{z}\left|\mathbf{a},\mathbf{\Psi}\right.\right)$ in (H.2) can be decomposed as:

$$p\left(\mathbf{z}\left|\mathbf{a},\mathbf{\Psi}\right.\right) \quad = \quad p\left(\mathbf{z}_d\left|\mathbf{a}_d,\mathbf{\Psi}\right.\right)\prod_{k\in I_p}p\left(z\left(k\right)\left|a\left(k\right),\mathbf{\Psi}\right.\right),$$

with $\mathbf{z}_d=\left(z\left(k\right):\,k\in I_d\right)$, and the a priori probability mass function $\Pr\left[\mathbf{a}\right]$ also satisfies (5.13), i.e.,

$$\Pr\left[\mathbf{a}\right]\propto I\left[\mathbf{a}_d\in\zeta_{\mathcal{D}}\right]\prod_{k\in I_p}I\left[a\left(k\right)=A_k\right],\,\forall\mathbf{a}\in\Omega^{N_s},$$

with $\zeta_{\mathcal{D}}$ the set of legitimate data symbol sequences $\mathbf{a}_d = (a(k) : k \in I_d)$, A_k the actual value of the pilot symbol $a(k) \in I_p$, and $I[.]$ the indicator function defined in (2.2), i.e., $I[P] = 0$ when P is false and $I[P] = 1$ when P is true. Taking this into account, we can rewrite the numerator in (H.2) as:

$$\begin{aligned}&\sum_{\mathbf{a}\in\Omega^{N_s}} a(k)\Pr[\mathbf{a}]\, p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})\\ &= \begin{cases} A_k \prod_{l\in I_p} p(z(l)|A_l,\mathbf{\Psi})\, p(\mathbf{z}_d|\mathbf{a}_d,\mathbf{\Psi}) & , k\in I_p \\ \prod_{l\in I_p} p(z(l)|A_l,\mathbf{\Psi}) \sum_{\mathbf{a}_d\in\zeta_{\mathcal{D}}} a(k)\, p(\mathbf{z}_d|\mathbf{a}_d,\mathbf{\Psi}) & , k\in I_d \end{cases}\end{aligned}$$

and the denominator as

$$\begin{aligned}&\sum_{\mathbf{a}\in\Omega^{N_s}} \Pr[\mathbf{a}]\, p(\mathbf{r}|\mathbf{a},\mathbf{\Psi})\\ &= \prod_{l\in I_p} p(z(l)|A_l,\mathbf{\Psi})\, p(\mathbf{z}_d|\mathbf{a}_d,\mathbf{\Psi}).\end{aligned}$$

The marginal symbol a posteriori probabilities $\mu(k;\mathbf{z})$ (H.2) thus become:

$$\mu(k;\mathbf{z}) = \begin{cases} a(k) & , k\in I_p \\ E_{a(k)}[a(k)|\mathbf{z}_d] & , k\in I_d \end{cases}, \tag{H.3}$$

where $E_{a(k)}[a(k)|\mathbf{z}_d]$ denotes the average of the symbol $a(k)$, conditioned on $\mathbf{z}_d(\mathbf{r},\mathbf{\Psi})$.

H.2 Computation of $H_{i,j}(k,l)$ for $k \in I_p$, $l \in I_d$

It follows from Appendix D that:

$$\begin{aligned}H_{i,j}(k,l) &= d_i(k)\, d_j(l)\, E_{\mathbf{z}}[\Im\{\mu^*(k;\mathbf{z})\, z(k)\}\, \Im\{\mu^*(l;\mathbf{z})\, z(l)\}]\\ &\quad + \frac{1}{2} Cov[\delta z_i(k)\, \delta z_j(l)]\, \Re\{E_{\mathbf{z}}[\mu(k;\mathbf{z})\, \mu^*(l;\mathbf{z})]\}.\end{aligned}$$

Taking into account (H.3), this yields:

$$\begin{aligned}H_{i,j}(k,l) &= d_i(k)\, d_j(l) \underbrace{E_{z(k)}[\Im\{A_k^* z(k)\}]}_{=0} E_{\mathbf{z}_d}\left[\Im\left\{E_{a(l)}[a^*(l)|\mathbf{z}_d]\, z(l)\right\}\right]\\ &\quad + \frac{1}{2} Cov[\delta z_i(k)\, \delta z_j(l)]\, \Re\left\{a(k) \underbrace{E_{\mathbf{z}_d}\left[E_{a(l)}[a^*(l)|\mathbf{z}_d]\right]}_{=0}\right\},\\ &= 0,\end{aligned}$$

where the first expectation follows from $z(k) = a(k) + w(k)$ with $E[w(k)] = 0$:

$$\begin{aligned} E_{z(k)}\left[\Im\left\{A_k^* z(k)\right\}\right] &= E_{z(k)|a(k)=A_k}\left[\Im\left\{A_k^* z(k)\right\}\right], \\ &= \Im\left\{|A_k|^2\right\} = 0 \end{aligned}$$

and the last expectation is computed as follows

$$E_{\mathbf{z}_d}\left[E_{a(l)}\left[a^*(l)\,|\mathbf{z}_d\right]\right] = E_{a(l)}\left[a^*(l)\right] = 0,$$

where the outer average removes the conditioning on $\mathbf{z}_d$ in the inner average, and the resulting average of the data symbol $a(l)$, $l \in I_d$ equals zero.

Appendix I

In the case of linear modulation, the log-likelihood function $\ell(\mathbf{\Psi};\mathbf{r}) = \ln p(\mathbf{r}\,|\mathbf{\Psi}\,)$ can be written as follows:

$$\ell(\mathbf{\Psi};\mathbf{r}) \quad = \quad \ln\left(E_{\mathbf{a}}\left[\prod_{k\in I_s} F(a(k), z(k;\mathbf{r},\mathbf{\Psi}))\right]\right), \tag{I.1}$$

with $\mathbf{z}(\mathbf{r},\mathbf{\Psi}) = \mathbf{r}\mathbf{S}(\mathbf{\Psi})$ and

$$F(a(k), z(k)) \quad = \quad \exp\left(\frac{E_s}{N_0}\left(2\Re\left\{a^*(k)\, z(k;\mathbf{r},\mathbf{\Psi})\right\} - |a(k)|^2\right)\right). \tag{I.2}$$

For small E_s/N_0, we obtain an approximation of $\ell(\mathbf{\Psi};\mathbf{r})$ by expanding the exponential function from (I.2) into a Taylor series, working out the product in (I.1) over the set of indices I_s, averaging each resulting term with respect to the symbol vector $\mathbf{a}$, and keeping only the relevant terms that correspond to the smallest powers of E_s/N_0. Expanding $F(.,.)$ yields:

$$\begin{aligned} &F(a(k), z(k;\mathbf{r},\mathbf{\Psi})) = \\ &1 + \sum_{p=1}^{\infty}\sum_{q=0}^{p}\sum_{r=0}^{p-q}\left(\frac{E_s}{N_0}\right)^p f(p,q,r;z(k))\,(a^*(k))^{p-r}\,(a(k))^{q+r}, \end{aligned} \tag{I.3}$$

with

$$f\left(p,q,r;z\left(k\right)\right)=\frac{\left(-1\right)^{q}\left(z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)^{p-q-r}\left(z^{*}\left(k;\mathbf{r},\mathbf{\Psi}\right)\right)^{r}}{q!r!\left(p-q-r\right)!}. \tag{I.4}$$

Assuming that the set of indices I_p is not empty, or equivalently, the symbol vector $\mathbf{a}$ contains at least one pilot symbol, the log-likelihood function can be approximated as follows:

$$\begin{aligned}\ell\left(\mathbf{\Psi};\mathbf{r}\right) &\approx \ln\left(1+2\frac{E_s}{N_0}\sum_{k\in I_s}\Re\left\{E\left[a^{*}\left(k\right)\right]z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right\}\right), &\text{(I.5)}\\ &= \ln\left(1+2\frac{E_s}{N_0}\sum_{k\in I_p}\Re\left\{a^{*}\left(k\right)z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right\}\right), &\text{(I.6)}\\ &\approx 2\frac{E_s}{N_0}\sum_{k\in I_p}\Re\left\{a^{*}\left(k\right)z\left(k;\mathbf{r},\mathbf{\Psi}\right)\right\}. &\text{(I.7)}\end{aligned}$$

All other terms in (I.1) are either independent of $\mathbf{\Psi}$ (for small E_s/N_0 these terms can be neglected as compared to the term 1 in (I.5)), or dependent of $\mathbf{\Psi}$ but containing a power of E_s/N_0 that is larger than 1. Straightforward application of (I.7) to (3.26) yields (the dominant part of) the FIM, which turns out to be equal to the DA (M)FIM with entries (5.25)-(5.27).

When the set of indices I_p is empty, the right-hand side of (I.7) reduces to zero and therefore additional terms containing a power of E_s/N_0 larger than 1 (and depending on Ψ) must be included in (I.5). In general, these terms depend on moments of the data symbols of the following type:

$$E\left[\prod_{i}\left(a^{*}\left(k_i\right)\right)^{p_i-r_i}\left(a\left(k_i\right)\right)^{q_i+r_i}\right], \tag{I.8}$$

with $\sum_i p_i \geq 1$ determining the power of E_s/N_0. When the symbols $a\left(k_i\right)$ are statistically independent, the order of the expectation operation and the multiplication operation in (I.8) may be reversed so that only moments of a single data symbol remain. Besides, when the symbol constellation is rotationally symmetric over $2\pi/K$ (K=2 for M-PAM, K=4 for M-QAM, K=M for M-PSK), these moments reduce to zero for $2r_i+q_i-p_i \notin \{0,\pm K,\pm 2K,...\}$. Taking this into account, a simple closed-form expression for (the dominant part of) the FIM has been derived in [57]. The computation of (the dominant part of) the FIM is considerably more difficult in the case of coded transmission, because of the statistical dependence between the data symbols. As the evaluation of the expectations in (I.8) can become quite tedious for an arbitrary encoding rule/mapping rule combination, a general closed-form expression for the ACRBs will not be presented. We note however that in the case of coded 2-PAM (2-PSK) it is sufficient to assume that the data symbols are pairwise uncorrelated, i.e., $E\left[a\left(k_1\right)a\left(k_2\right)\right]=\delta_{k_1-k_2}$, in order to obtain the same ACRBs as for uncoded 2-PAM. Pairwise uncorrelated data symbols occur not only for statistically independent $\{a\left(k\right)\}$, but also for the large majority of practical codes.

Appendix J

From (5.76), (5.71) and (5.70) it follows that the elements $J_{i,j}$, $(i,j) \in \{1,2\}^2$of the FIM (5.48) can be rewritten as:

$$J_{i,j} = M_{i,j}^{(DA)} + \left(\frac{2E_s}{N_0}\right)^2 \sum_{k,l \in I_d} E_{z(k),z(l)} \left[H_{i,j}\left(k,l;z\left(k\right),z\left(l\right)\right)\right], \quad \text{(J.1)}$$

where $M_{1,1}^{(DA)} = M_{\nu,\nu}^{(DA)}$, $M_{1,2}^{(DA)} = M_{2,1}^{(DA)} = M_{\nu,\theta}^{(DA)} = M_{\theta,\nu}^{(DA)}$ and $M_{\theta,\theta}^{(DA)}$ denote the elements of the DA MFIM (5.25)-(5.27), the notation $z\left(k\right)$, $k \in I_s$ is a short-hand writing of $z\left(k;\mathbf{r},\mathbf{\Psi}\right)$, and the quantity $H_{i,j}\left(k,l;z\left(k\right),z\left(l\right)\right)$ is defined as:

$$\begin{aligned} &H_{i,j}\left(k,l;z\left(k\right),z\left(l\right)\right) = \\ &E_{z_i(k),z_j(l)|z(k),z(l)}\left[\Re\left(z_i\left(k\right)\mu^*\left(k;z\left(k\right)\right)\right)\Re\left(z_j\left(l\right)\mu^*\left(l;z\left(l\right)\right)\right)\right], \end{aligned} \quad \text{(J.2)}$$

where $z_i\left(k\right)$, $(i,k) \in \{1,2\} \times I_s$ is a short-hand writing of $z_i\left(k;\mathbf{r},\mathbf{\Psi}\right)$. By taking into account the statistical properties of $\mathbf{z}$, $\frac{\partial \mathbf{z}}{\partial \nu}$ and $\frac{\partial \mathbf{z}}{\partial \theta}$ (see Appendix F), we are able to perform analytically the average in (J.2) over $\left(z_i\left(k\right),z_j\left(l\right)\right)$, conditioned

on $(z(k), z(l))$. It follows from (F.40) that:

$$\begin{aligned} &H_{i,j}(k,l;z(k),z(l)) = \\ &\frac{1}{2}\Re\left\{\mu^*(k;z(k))\,\mu^*(l;z(l))\,E\left[z_i(k)\,z_j(l)\,|z(k),z(l)\right]\right\} \\ &+\frac{1}{2}\Re\left\{\mu(k;z(k))\,\mu^*(l;z(l))\,E\left[z_i^*(k)\,z_j(l)\,|z(k),z(l)\right]\right\}, \end{aligned} \tag{J.3}$$

with

$$E\left[z_i(k)\,z_j(l)\,|z(k),z(l)\right] \quad = \quad d_i(k)\,d_j(l)\,z(k)\,z(l), \tag{J.4}$$

$$\begin{aligned} &E\left[z_i^*(k)\,z_j(l)\,|z(k),z(l)\right] \\ &= d_i^*(k)\,d_j(l)\,z^*(k)\,z(l) + Cov\left[\delta z_i(k),\delta z_j(l)\right] \end{aligned} \tag{J.5}$$

and where (for both observation models), with $(\Psi_1,\Psi_2)=(\nu,\theta)$,

$$d_1(k) = -j2\pi(k-\kappa_0),$$

$$d_2(k) = -j$$

and $\delta z_i(k)=0$ except for the variable $\Psi_1=\nu$ in the correct observation model in which case:

$$E\left[\delta z_\nu(k)\right]=0,$$

$$Cov\left[\delta z_\nu(k),\delta z_\nu(k+m)\right] = \frac{N_0}{E_s}4\pi^2\mathrm{f}(mT).$$

Further manipulation yields for the expectation $E_{z(k),z(l)}\left[H_{i,j}(k,l;z(k),z(l))\right]$ in (J.1):

$$\begin{aligned} &E_{z(k),z(l)}\left[H_{i,j}(k,l;z(k),z(l))\right] \\ &= -d_i(k)\,d_j(l)\,\Gamma(k,l) + \frac{1}{2}Cov\left[\delta z_i(k)\,\delta z_j(l)\right]\Upsilon(k,l), \end{aligned} \tag{J.6}$$

with

$$\Gamma(k,l) = E\left[\Im\left\{\mu^*(k;z(k))\,z(k)\right\}\Im\left\{\mu^*(l;z(l))\,z(l)\right\}\right], \tag{J.7}$$

$$\Upsilon(k,l) = \Re\left\{E\left[\mu(k;z(k))\,\mu^*(l;z(l))\right]\right\}, \tag{J.8}$$

where the quantities $z(k)$, $k\in I_d$ can be decomposed as $z(k)=a(k)+n(k)$ with $E[a(k)]=0$, $E\left[a(k)\,a^*(l)\right]=\delta_{k-l}$, $E[n(k)]=0$ and $E\left[n(k)\,n^*(l)\right]=(N_0/E_s)\,\delta_{k-l}$, and the quantities $\mu(k;z(k))$, $k\in I_d$ denote the a posteriori average of $a(k)$, i.e., $\mu(k;z(k))=E\left[a(k)\,|z(k)\right]$, with $E\left[X\,|Y\right]$ denoting the statistical expectation of X conditioned on the value of Y. Using the law of total expectation, i.e., $E[X]=E\left[E\left[X\,|Y\right]\right]$, the quantities $\Gamma(k,l)$ (J.7) and $\Upsilon(k,l)$ (J.8), $k,l\in I_d$ can be further computed as follows. We will individually consider the cases ($k=l$, $k\in I_d$) and ($k\neq l$, $k,l\in I_d$).

1. For the case ($k = l$, $k \in I_d$) we simply obtain:

$$\Gamma(k,k) = E_{z(k)}\left[\left(\Im\left\{\mu^*\left(k; z\left(k\right)\right) z\left(k\right)\right\}\right)^2\right] = \frac{N_0}{2E_s} R_\Omega\left(\frac{E_s}{N_0}\right),$$

$$\Upsilon(k,k) \quad = \quad E_{z(k)}\left[\left|\mu\left(k; z\left(k\right)\right)\right|^2\right] = Q_\Omega\left(\frac{E_s}{N_0}\right).$$

2. For the case ($k \neq l$, $k, l \in I_d$) the quantities $\Gamma(k,l)$ and $\Upsilon(k,l)$ reduce to zero. This can be shown as follows:

$$\begin{aligned}
\Gamma(k,l) &= \Im^2\left\{E\left[\mu^*\left(k; z\left(k\right)\right) z\left(k\right)\right]\right\}, \\
&= \Im^2\left\{E\left[E\left[a^*\left(k\right)\left|z\left(k\right)\right.\right] z\left(k\right)\right]\right\}, \\
&= \Im^2\left\{E\left[E\left[a^*\left(k\right) z\left(k\right)\left|z\left(k\right)\right.\right]\right]\right\}, \\
&= \Im^2\left\{E\left[a^*\left(k\right) z\left(k\right)\right]\right\}, \\
&= \Im^2\left\{E\left[E\left[a^*\left(k\right) z\left(k\right)\left|a\left(k\right)\right.\right]\right]\right\}, \\
&= \Im^2\left\{E\left[\left|a\left(k\right)\right|^2\right]\right\}, \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
\Upsilon(k,l) &= \left|E\left[\mu\left(k; z\left(k\right)\right)\right]\right|^2, \\
&= \left|E\left[E\left[a\left(k\right)\left|z\left(k\right)\right.\right]\right]\right|^2, \\
&= \left|E\left[a\left(k\right)\right]\right|^2, \\
&= 0.
\end{aligned}$$

Substituting these expressions into (J.6), and further into (J.1) we obtain:

$$J_{i,j} \quad = \quad M_{i,j}^{(DA)} + \frac{2E_s}{N_0}\left(C_{i,j}^{(1)} R_\Omega\left(\frac{E_s}{N_0}\right) + C_{i,j}^{(2)} Q_\Omega\left(\frac{E_s}{N_0}\right)\right), \qquad \text{(J.9)}$$

where (for both observation models), with $(\Psi_1, \Psi_2) = (\nu, \theta)$,

$$C_{i,j}^{(1)} = 4\pi^2 \sum_{k \in I_d} \left(k - \kappa_0\right)^2,$$

$$C_{1,2}^{(1)} = C_{2,1}^{(1)} = 2\pi \sum_{k \in I_d} \left(k - \kappa_0\right),$$

$$C_{2,2}^{(1)} = N_d,$$

and $C_{i,j}^{(2)} = 0$ except for the variable $i = j = 1$ in the correct observation model in which case $C_{1,1}^{(2)} = 4\pi^2 N_d \mathrm{f}\left(0\right)$. Substituting the expressions of the DA MFIM entries (5.25)-(5.27) into (J.9) finally yields (5.78)-(5.80).

Appendix K

In Appendix E it was shown that the FIM equals the MFIM when, for large E_s/N_0, the symbol vector a posteriori probability density function $\Pr[\mathbf{a}=\tilde{\mathbf{a}}\,|\mathbf{r},\boldsymbol{\Psi}]$ converges to:

$$\Pr[\mathbf{a}=\tilde{\mathbf{a}}\,|\mathbf{r},\boldsymbol{\Psi}] \;=\; \begin{cases} 1 & ,\tilde{\mathbf{a}}=\mathbf{a}_0 \\ 0 & ,\tilde{\mathbf{a}}\neq\mathbf{a}_0 \end{cases}, \tag{K.1}$$

where $\mathbf{a}_0$ denotes the actual value of the symbol vector $\mathbf{a}$ (i.e., $\mathbf{r}$ is distributed according to $p(\mathbf{r}\,|\mathbf{a}=\mathbf{a}_0,\boldsymbol{\Psi})$).

When the a priori distribution $\Pr[\mathbf{a}]$ of the symbol vector $\mathbf{a}$ satisfies (5.17), these a posteriori symbol averages can be expressed as:

$$\Pr[\mathbf{a}=\tilde{\mathbf{a}}\,|\mathbf{r},\boldsymbol{\Psi}] \;=\; \frac{p(\mathbf{r}\,|\mathbf{a}=\tilde{\mathbf{a}},\boldsymbol{\Psi})}{E_{\mathbf{a}}[p(\mathbf{r}\,|\mathbf{a},\boldsymbol{\Psi})]}, \tag{K.2}$$

$$\;=\; \frac{\prod\limits_{k\in I_s}\exp\left(2\Re\{\tilde{a}^*(k)\,z(k;\mathbf{r},\boldsymbol{\Psi})\}-|\tilde{a}(k)|^2\right)}{\prod\limits_{k\in I_s}\sum\limits_{\omega\in\Omega}\exp\left(2\Re\{\omega z(k;\mathbf{r},\boldsymbol{\Psi})\}-|\omega|^2\right)}, \tag{K.3}$$

where, for given $(\mathbf{a},\boldsymbol{\Psi})$, $\mathbf{z}(\mathbf{r},\boldsymbol{\Psi})=\mathbf{a}+\mathbf{w}$ with the complex-valued zero-mean Gaussian noise vector $\mathbf{w}$ having statistically independent components with vari-

ance N_0/E_s. Dividing both numerator and denominator of the right-hand side of (K.2) by $\prod_{k\in I_s} \exp\left(2\Re\left\{a_0^*(k)\, z(k;\mathbf{r},\mathbf{\Psi})\right\} - |a_0(k)|^2\right)$, we obtain :

$$\Pr\left[\mathbf{a}=\tilde{\mathbf{a}}\,|\mathbf{r},\mathbf{\Psi}\right] \;=\; \frac{\prod_{k\in I_s} \check{F}\left(a(k)=\tilde{a}(k), a_0(k)\right)}{\prod_{k\in I_s}\sum_{\omega\in\Omega}\check{F}\left(a(k)=\omega, a_0(k)\right)}, \tag{K.4}$$

with

$$\check{F}\left(a(k), a_0(k)\right) = \frac{\exp\left(2\Re\left\{a^*(k)\, z(k;\mathbf{r},\mathbf{\Psi})\right\} - |a(k)|^2\right)}{\exp\left(2\Re\left\{a_0^*(k)\, z(k;\mathbf{r},\mathbf{\Psi})\right\} - |a_0(k)|^2\right)}. \tag{K.5}$$

Note that $\check{F}\left(a(k), a_0(k)\right) = 1$ when $a(k) = a_0(k)$. For a given value of $a_0(k)$, the quantity $\check{F}\left(a(k), a_0(k)\right)$ is log-normally distributed with median value[1] equal to:

$$\exp\left(E\left[\ln \check{F}\left(a(k), a_0(k)\right)\right]\right) = \exp\left(-\frac{E_s}{N_0}\left|a(k)-a_0(k)\right|^2\right). \tag{K.6}$$

This implies that the quantities $\check{F}\left(a(k), a_0(k)\right)$ with $a(k) \neq a_0(k)$ are likely to be much smaller than 1 when $\frac{E_s}{N_0}\left|a(k)-a_0(k)\right|^2$ is large. In this case, the sum in the denominator of the right-hand side of (K.4) is, with high probability, strongly dominated by the term with $\omega = a_0(k)$ (for which $\check{F}\left(a(k)=\omega, a_0(k)\right) = 1$), so that the left-hand side of (K.4) is indeed well approximated by the right-hand side of (K.1), and the FIM converges to the MFIM.

Now, let d_M denote the minimum Euclidean distance between the constellation points. At high E_s/N_0, the quantities $\check{F}\left(a(k)=\omega, a_0(k)\right)$ with $|\omega - a_0(k)| = d_M$ are most likely to be the second largest of all $\check{F}\left(a(k)=\omega, a_0(k)\right)$. The convergence of the FIM to the MFIM is therefore mainly determined by the value of $\frac{E_s}{N_0}\left(d_M\right)^2$. This establishes a direct relation between the convergence of the FIM to the MFIM and the uncoded SER from (4.19), which mainly depends on $\frac{E_s}{N_0}\left(d_M\right)^2$.

We will now verify that the quantities $\check{F}\left(a(k)=\omega, a_0(k)\right)$ with $\omega \neq a_0(k)$ are indeed likely to be much smaller than 1 when E_s/N_0 is such that the SER is about 10^{-3}. The cumulative distribution function CDF (x) = $\Pr\left[\check{F}\left(a(k)=\omega, a_0(k)\right) \leq \mathrm{x}\right]$ corresponding to $|\omega - a_0(k)| = d_M$ is plotted in Fig. K.1 as a function of x. For $\frac{E_s}{N_0}\left(d_M\right)^2$=12.8 dB (which corresponds to a SER of about 10^{-3}), we observe that the considered $\check{F}\left(a(k)=\omega, a_0(k)\right)$ are less than 10^{-3} (10^{-2}) in about 97.5% (99%) of all cases, which indicates that they can be neglected as compared to $\check{F}\left(a(k)=\omega, a_0(k)\right) = 1$.

[1]The log-normal distribution is highly skewed to the right (asymmetrical with right tail longer then left). The median is less affected by extreme scores than the mean and is therefore a better measure of central tendency for such an extremely skewed distribution.

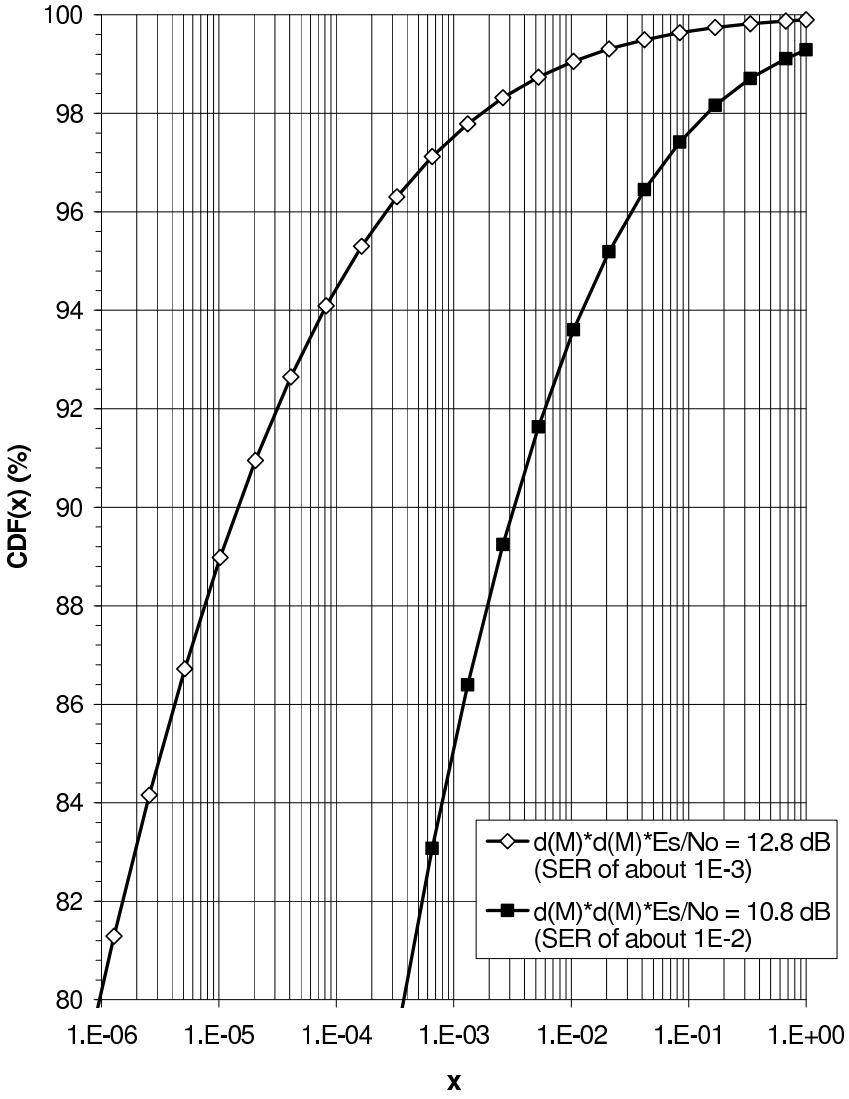

Figure K.1: Cumulative distribution function of $\check{F}\left(a\left(k\right)=\omega,a_0\left(k\right)\right)$ $\left(\mathrm{CDF}\left(\mathrm{x}\right)=\Pr\left[\check{F}\left(a\left(k\right)=\omega,a_0\left(k\right)\right)\leq\mathrm{x}\right]\right)$ corresponding to $\left|\omega-a_0\left(k\right)\right|=d_M$ as a function of x.

Appendix L

The procedure for computing the high-SNR limit of the FIMs related to $L(\nu;\mathbf{r})$ and $L(\theta;\mathbf{r})$ is very similar to the one for computing the high-SNR limit of the FIM related to $L(\mathbf{\Psi};\mathbf{r})$ (Appendix E). Assuming that $\mathbf{r} = \mathbf{s}(\mathbf{a},\Psi_i,\Psi_{3-i}) + \mathbf{w}$, with $\mathbf{w}$ Gaussian noise with $E\left[\mathbf{w}^H\mathbf{w}\right] = (N_0/E_s)\,\mathbf{I}$ and $(\mathbf{a},\Psi_1 \equiv \nu, \Psi_2 \equiv \theta)$ uniformly distributed over the domain $(\zeta_{\mathcal{A}}, [-\nu_{max},\nu_{max}], [-\pi,\pi])$, we obtain for the APPs $p\left(\mathbf{a}=\tilde{\mathbf{a}}, \Psi_{3-i} = \tilde{\Psi}_{3-i}\,|\mathbf{r},\Psi_i\right)$, $i \in \{1,2\}$ and $\tilde{\mathbf{a}} \in \zeta_{\mathcal{A}}$:

$$
\begin{aligned}
& p\left(\mathbf{a}=\tilde{\mathbf{a}}, \Psi_{3-i} = \tilde{\Psi}_{3-i}\,|\mathbf{r},\Psi_i\right) \\
&= \frac{p\left(\mathbf{r}\,\middle|\mathbf{a}=\tilde{\mathbf{a}},\Psi_i,\Psi_{3-i}=\tilde{\Psi}_{3-i}\right)}{\int \sum_{\check{\mathbf{a}}\in\zeta_{\mathcal{A}}} p\left(\mathbf{r}\,\middle|\mathbf{a}=\check{\mathbf{a}},\Psi_i,\Psi_{3-i}=\tilde{\Psi}_{3-i}\right) \mathrm{d}\tilde{\Psi}_{3-i}}, \\
&= \frac{e^{-\frac{E_s}{N_0}\left|\mathbf{r}-\mathbf{s}\left(\tilde{\mathbf{a}},\Psi_i,\tilde{\Psi}_{3-i}\right)\right|^2}}{\int \sum_{\check{\mathbf{a}}\in\zeta_{\mathcal{A}}} e^{-\frac{E_s}{N_0}\left|\mathbf{r}-\mathbf{s}\left(\check{\mathbf{a}},\Psi_i,\check{\Psi}_{3-i}\right)\right|^2} \mathrm{d}\check{\Psi}_{3-i}}.
\end{aligned}
$$

For large E_s/N_0, this converges to:

$$
p\left(\mathbf{a}=\tilde{\mathbf{a}}, \Psi_{3-i} = \tilde{\Psi}_{3-i}\,|\mathbf{r},\Psi_i\right) \;=\; \begin{cases} \delta\left(\tilde{\Psi}_{3-i} - \hat{\Psi}_{3-i}(\mathbf{r})\right) & , \tilde{\mathbf{a}} = \hat{\mathbf{a}}(\mathbf{r}) \\ 0 & , \tilde{\mathbf{a}} \neq \hat{\mathbf{a}}(\mathbf{r}) \end{cases},
$$

with

$$\left(\hat{\mathbf{a}}\left(\mathbf{r}\right),\hat{\Psi}_{3-i}\left(\mathbf{r}\right)\right)=\arg\min_{\check{\mathbf{a}}\in\zeta_{\mathcal{A}},\check{\Psi}_{3-i}}\left|\mathbf{r}-\mathbf{s}\left(\check{\mathbf{a}},\Psi_i,\check{\Psi}_{3-i}\right)\right|^2$$

We recall that the FIMs J_i, $i=1,2$ are given by

$$\begin{aligned}&J_i\\&=E_{\mathbf{r},\Psi_i}\left[\left(\ell'\left(\Psi_i;\mathbf{r}\right)\right)^2\right]\\&=\int\sum_{\check{\mathbf{a}}\in\zeta_{\mathcal{A}}}\Pr\left[\mathbf{a}=\check{\mathbf{a}}\right]I\left(\Psi_i;\check{\mathbf{a}},\check{\Psi}_{3-i}\right)p\left(\Psi_{3-i}=\check{\Psi}_{3-i}\right)p\left(\Psi_i\right)\mathrm{d}\Psi_i\mathrm{d}\check{\Psi}_{3-i}\end{aligned}$$

with

$$\begin{aligned}&I\left(\Psi_i;\check{\mathbf{a}},\check{\Psi}_{3-i}\right)\\&=\int\left(\ell'\left(\Psi_i;\mathbf{r}\right)\right)^2p\left(\mathbf{r}\left|\mathbf{a}=\check{\mathbf{a}},\Psi_i,\Psi_{3-i}=\check{\Psi}_{3-i}\right.\right)\mathrm{d}\mathbf{r}\end{aligned}$$

and

$$\ell'\left(\Psi_i;\mathbf{r}\right)=\int\sum_{\tilde{\mathbf{a}}\in\zeta_{\mathcal{A}}}f\left(\Psi_i;\tilde{\mathbf{a}},\tilde{\Psi}_{3-i},\mathbf{r}\right)\mathrm{d}\tilde{\Psi}_{3-i},\tag{L.1}$$

with

$$\begin{aligned}&f\left(\Psi_i;\tilde{\mathbf{a}},\tilde{\Psi}_{3-i},\mathbf{r}\right)\\&=\frac{\partial\ln p\left(\mathbf{r}\left|\mathbf{a}=\tilde{\mathbf{a}},\Psi_i,\Psi_{3-i}=\tilde{\Psi}_{3-i}\right.\right)}{\partial\Psi_i}p\left(\mathbf{a}=\tilde{\mathbf{a}},\Psi_{3-i}=\tilde{\Psi}_{3-i}\left|\mathbf{r},\Psi_i\right.\right).\end{aligned}$$

For given $\left(\mathbf{a}=\check{\mathbf{a}},\Psi_i,\Psi_{3-i}=\check{\Psi}_{3-i}\right)$, the vector $\mathbf{r}$ is distributed according to $p\left(\mathbf{r}\left|\mathbf{a}=\check{\mathbf{a}},\Psi_i,\Psi_{3-i}=\check{\Psi}_{3-i}\right.\right)$. For large E_s/N_0, $\mathbf{r}$ converges to $\mathbf{s}\left(\check{\mathbf{a}},\Psi_i,\check{\Psi}_{3-i}\right)$, such that $p\left(\mathbf{a}=\tilde{\mathbf{a}},\Psi_{3-i}=\tilde{\Psi}_{3-i}\left|\mathbf{r},\Psi_i\right.\right)$ converges to:

$$p\left(\mathbf{a}=\tilde{\mathbf{a}},\Psi_{3-i}=\tilde{\Psi}_{3-i}\left|\mathbf{r},\Psi_i\right.\right)\quad=\quad\begin{cases}\delta\left(\tilde{\Psi}_{3-i}-\check{\Psi}_{3-i}\right) & ,\tilde{\mathbf{a}}=\check{\mathbf{a}}\\0 & ,\tilde{\mathbf{a}}\neq\check{\mathbf{a}}\end{cases}.\tag{L.2}$$

From (L.2) and (L.1) it follows that, for given $\left(\mathbf{a}=\check{\mathbf{a}},\Psi_i,\Psi_{3-i}=\check{\Psi}_{3-i}\right)$ and increasing E_s/N_0:

$$\ell'\left(\Psi_i;\mathbf{r}\right)\approx\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\Psi_i,\check{\Psi}_{3-i}\right.\right)}{\partial\Psi_i},$$

such that:

$$\begin{aligned}&I\left(\Psi_i;\check{\mathbf{a}},\check{\Psi}_{3-i}\right)\\&\approx\int\left(\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\Psi_i,\check{\Psi}_{3-i}\right.\right)}{\partial\Psi_i}\right)^2p\left(\mathbf{r}\left|\mathbf{a}=\check{\mathbf{a}},\Psi_i,\Psi_{3-i}=\check{\Psi}_{3-i}\right.\right)\mathrm{d}\mathbf{r},\\&=E_{\mathbf{r}\left|\mathbf{a}=\check{\mathbf{a}},\Psi_i,\Psi_{3-i}=\check{\Psi}_{3-i}\right.}\left[\left(\frac{\partial\ln p\left(\mathbf{r}\left|\check{\mathbf{a}},\Psi_i,\check{\Psi}_{3-i}\right.\right)}{\partial\Psi_i}\right)^2\right]\end{aligned}$$

and finally

$$
\begin{aligned}
& J_i \\
& \approx E_{\mathbf{r},\mathbf{a},\Psi_i,\Psi_{3-i}} \left[\frac{\partial \ln p\left(\mathbf{r}\,|\mathbf{a},\Psi_i,\Psi_{3-i}\right)}{\partial \Psi_i} \frac{\partial \ln p\left(\mathbf{r}\,|\mathbf{a},\Psi_i,\Psi_{3-i}\right)}{\partial \Psi_j} \right], \\
& = M_i.
\end{aligned}
$$

Appendix M

First, let us compute the quantity $H_\nu(k,l;\mathbf{z})$ from (5.166), i.e.,

$$H_\nu(k,l;\mathbf{z}) = E_{z_\nu(k),z_\nu(l)|\mathbf{z}}\left[\Re\left\{K(k;\mathbf{z})\,z_\nu(k)\right\}\Re\left\{K(l;\mathbf{z})\,z_\nu(l)\right\}\right],$$

where $z_\nu(k)=\frac{\partial z(k)}{\partial \nu}$.

From (F.40) it follows that:

$$\begin{aligned} H_\nu(k,l;\mathbf{z}) &= \frac{1}{2}\Re\left\{K(k;\mathbf{z})\,K(l;\mathbf{z})\,E\left[z_\nu(k)\,z_\nu(l)\,|\mathbf{z}\right]\right\} \\ &\quad +\frac{1}{2}\Re\left\{K^*(k;\mathbf{z})\,K(l;\mathbf{z})\,E\left[z_\nu^*(k)\,z_\nu(l)\,|\mathbf{z}\right]\right\}, \end{aligned} \tag{M.1}$$

where the correlations $E\left[z_\nu(k)\,z_\nu(l)\,|\mathbf{z}\right]$ and $E\left[z_\nu^*(k)\,z_\nu(l)\,|\mathbf{z}\right]$ of $z_\nu(k)$ and $z_\nu(l)$, conditioned on $\mathbf{z}$, are given by:

$$E\left[z_\nu(k)\,z_\nu(l)\,|\mathbf{z}\right] = -4\pi^2(k-\kappa_0)(l-\kappa_0)\,z(k)\,z(l), \tag{M.2}$$

$$\begin{aligned} E\left[z_\nu^*(k)\,z_\nu(l)\,|\mathbf{z}\right] &= 4\pi^2(k-\kappa_0)(l-\kappa_0)\,z^*(k)\,z(l) \\ &\quad +Cov\left[\delta z_\nu(k),\delta z_\nu(l)\right], \end{aligned} \tag{M.3}$$

with $Cov\left[\delta z_\nu(k),\delta z_\nu(l)\right]$ equal to 0 for the simplified observation model, and $Cov\left[\delta z_\nu(k),\delta z_\nu(l)\right]$ equal to $4\pi^2\frac{N_0}{E_s}\mathrm{f}\left((k-l)\,T\right)$ for the correct observation model, where $\mathrm{f}(t)=\int\left(\frac{\tilde{t}}{T}\right)^2 p\left(\tilde{t}-\frac{t}{2}\right)p\left(\tilde{t}+\frac{t}{2}\right)\mathrm{d}t$.

Then, further manipulation of (M.1) yields for the average in (5.165):

$$E_{\mathbf{z}}\left[\sum_{k,l\in I_s} H_\nu\left(k,l;\mathbf{z}\right)\right] \quad = \quad 4\pi^2\left(\frac{N_0}{2E_s}E_2+E_1\right), \tag{M.4}$$

with

$$E_1 = E\left[\left(\sum_{k\in I_s}\left(k-\kappa_0\right)\Im\left\{K\left(k;\mathbf{z}\right)z\left(k\right)\right\}\right)^2\right],$$

$$E_2 = E\left[\sum_{k,l\in I_s}\mathrm{f}\left(\left(k-l\right)T\right)\Re\left\{K^*\left(k;\mathbf{z}\right)K\left(l;\mathbf{z}\right)\right\}\right].$$

Appendix N

We consider the function $Y_{DA}(\nu;\mathbf{r})$ (6.23) for $\kappa_0 = \kappa_{m,p}$ (i.e., the phase shift is estimated at the center of the pilot part of the symbol burst) and evaluated at a trial value $\tilde{\nu}$ of the normalized frequency offset ν:

$$Y_{DA}(\tilde{\nu};\mathbf{r}) = \sum_{k\in I_p} A_k^* r(k)\, e^{-j2\pi\tilde{\nu}(k-\kappa_{m,p})}, \tag{N.1}$$

with

$$\kappa_{m,p} = \frac{1}{N_p}\sum_{k\in I_p} k \tag{N.2}$$

and

$$r(k) = A_k e^{j(2\pi\nu(k-\kappa_{m,p})+\theta)} + w(k),\ \forall k \in I_p, \tag{N.3}$$

where $|A_k|^2 = 1$, $E[w(k)] = 0$ and $E[w(k)\,w^*(l)] = (N_0/E_s)\,\delta_{k-l}$.

It follows that the $Y_{DA}(\tilde{\nu};\mathbf{r})$ (N.1) can be decomposed as:

$$Y_{DA}(\tilde{\nu};\mathbf{r}) = Y_0(\tilde{\nu}) + n_Y, \tag{N.4}$$

where $Y_0(\tilde{\nu})$ is the noiseless version of $Y_{DA}(\tilde{\nu};\mathbf{r})$:

$$Y_0(\tilde{\nu}) \quad = \quad e^{j\theta}\sum_{k\in I_p} e^{j2\pi(\nu-\tilde{\nu})(k-\kappa_{m,p})} \tag{N.5}$$

and n_Y is a zero-mean statistical fluctuation with a variance σ_Y^2, given by:

$$\sigma_Y^2 = E\left[\left|\sum_{k\in I_p} A_k^* w(k)\, e^{-j2\pi\tilde{\nu}(k-\kappa_{m,p})}\right|^2\right] = N_p\frac{N_0}{E_s}.$$

Assuming that the symbol vector $\mathbf{a}$ contains two blocks of $N_p/2$ consecutive pilot symbols spaced with S data symbols, it can be shown that [60]:

$$\begin{aligned}&|Y_0(\tilde{\nu})|\\ &= N_p\left|\operatorname{sinc}\left(\frac{N_p}{2}(\tilde{\nu}-\nu)\right)\cos\left(\pi\left(\frac{N_p}{2}+S\right)(\tilde{\nu}-\nu)\right)\right|,\end{aligned} \tag{N.6}$$

where $\operatorname{sinc}(x)$ is defined as $\sin(\pi x)/\pi x$. The function $|Y_0(\tilde{\nu})|$ (N.6) is symmetric about $\tilde{\nu}=\nu$. It has the shape of the absolute value of a sinc function modulated by a cosine function. The global maximum occurs at $\tilde{\nu}=\nu$ and has value N_p. There are also numerous lower amplitude maxima. The null-to-null width of the main sinc lobe is $\frac{4}{N_p}$ and the peaks of the cosine function are separated by $\frac{2}{N_p+2S_p} = \frac{4}{N_p}\left(\frac{1}{2+4\epsilon_{Sp}}\right)$, with $\epsilon_{Sp} = S/N_p$. The central peak of (N.6) gets narrower as the number of pilot symbols N_p or the spacing S increases. When $S>0$, there will be more than one peak of the cosine function within the main lobe interval of the sinc function. The larger the spacing ratio ϵ_{Sp} the smaller the difference in magnitude between the main peak and the secondary peaks. For a fixed spacing ratio, the difference in magnitude between the main peak and the secondary peaks is proportional to N_p. Fig. N.1 shows the function $|Y_0(\tilde{\nu})|$ (N.6) scaled by N_p as a function of $|\tilde{\nu}-\nu|$ for $N_p\in\{32,64\}$ and $\epsilon_{Sp}=0.5$. The sinc-shaped envelopes of (N.6) are also shown. Modulation of the sinc envelop with a cosine function produces peaks in $|Y_0(\tilde{\nu})|$ at virtually regular intervals. The effect of increasing the spacing ratio ϵ_{Sp} is illustrated in Fig. N.2 where $N_p=32$. The curve labeled $S=0$ corresponds to the case of 32 consecutive pilot symbols. This curve has a sinc shape. Separating two blocks of each 16 pilot symbols by 16 symbol periods produces the curve labeled $S=16$. The central peak of this function is narrower than the one in the consecutive case. The separation also introduces secondary peaks into the function at $|\tilde{\nu}-\nu|$ about 0.028. As the spacing S is increased, these peaks move closer to the center peak in both position and amplitude. At $S=48$ the secondary peaks have moved to $|\tilde{\nu}-\nu|\approx 0.015$ and are only 10% lower in magnitude than the main peak.

We now show that the curvature of $|Y_0(\tilde{\nu})|$ at $\tilde{\nu}=\nu$ is proportional to $E_{\mathbf{r},\mathbf{\Psi}}\left[\partial^2\ell_{DA}(\mathbf{\Psi};\mathbf{r})/\partial\nu^2\right]$, where $\ell_{DA}(\mathbf{\Psi};\mathbf{r})$ denotes the DA log-likelihood function (6.19) of $\mathbf{\Psi}$, which can be written as:

$$\ell_{DA}(\mathbf{\Psi};\mathbf{r}) \quad\propto\quad \Re\left\{Y_{DA}(\nu;\mathbf{r})\, e^{-j\theta}\right\}. \tag{N.7}$$

Substituting (N.3) into (N.1) and then into (N.7), it can be verified that:

$$E_{\mathbf{r}|\mathbf{\Psi}}\left[\ell_{DA}(\mathbf{\Psi};\mathbf{r})\right] \propto \Re\left\{Y_1(\nu;\theta)\right\}, \tag{N.8}$$

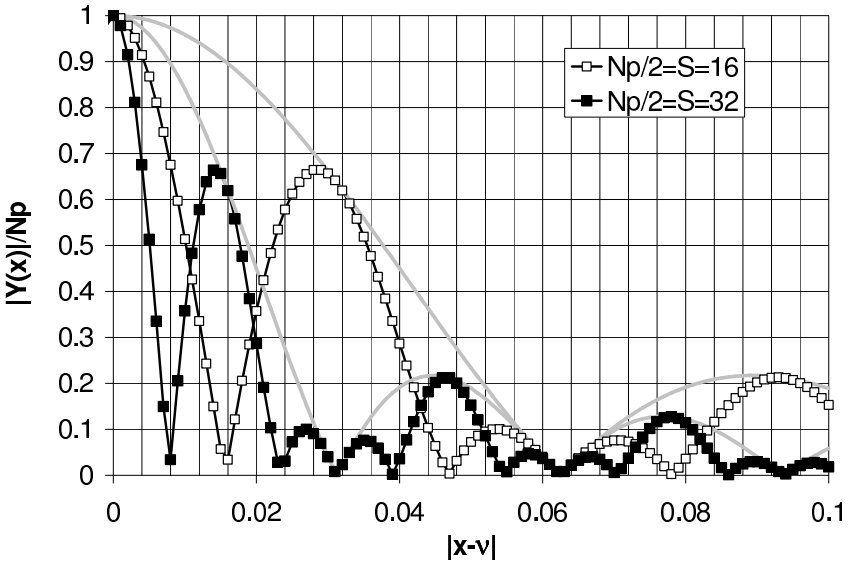

Figure N.1: Effect of increasing the number of pilot symbols for a fixed spacing ratio $\epsilon_{Sp} = 0.5$ on the objective function for DA normalized frequency offset estimation.

such that

$$E_{\mathbf{r}|\mathbf{\Psi}}\left[\frac{\partial^2}{\partial\nu^2}\ell_{DA}\left(\mathbf{\Psi};\mathbf{r}\right)\right] \propto \Re\left\{Y_1''\left(\nu,\theta\right)\right\}. \tag{N.9}$$

Here, $E_{\mathbf{r}|\mathbf{\Psi}}\left[.\right]$ denotes averaging with respect to $p\left(\mathbf{r}\left|\mathbf{\Psi}\right.\right)$, $Y_1''\left(\nu;\theta\right) = \left.\frac{\mathrm{d}^2 Y_1(\tilde{\nu};\theta)}{\mathrm{d}\tilde{\nu}^2}\right|_{\tilde{\nu}=\nu}$ and

$$Y_1\left(\tilde{\nu};\theta\right) = Y_0\left(\tilde{\nu}\right)e^{-j\theta}. \tag{N.10}$$

It follows from (N.10) that:

$$\left|Y_1\left(\tilde{\nu};\theta\right)\right| = \left|Y_0\left(\tilde{\nu}\right)\right|. \tag{N.11}$$

It further follows from (N.10) and (N.5) and (N.2) that, for given (ν,θ):

$$Y_1\left(\nu;\theta\right) = Y_0\left(\tilde{\nu}\right)|_{\tilde{\nu}=\nu}\, e^{-j\theta} = \sum_{k\in I_p} 1 = N_p, \tag{N.12}$$

$$Y_1'\left(\nu;\theta\right) \;\; = \left.\frac{\mathrm{d}Y_0(\tilde{\nu})}{\mathrm{d}\tilde{\nu}}\right|_{\tilde{\nu}=\nu} e^{-j\theta} = \;\; -j2\pi\sum_{k\in I_p}\left(k-\kappa_{m,p}\right) = 0 \tag{N.13}$$

and $Y_1''\left(\nu;\theta\right) = \left.\frac{\mathrm{d}^2 Y_0(\tilde{\nu})}{\mathrm{d}\tilde{\nu}^2}\right|_{\tilde{\nu}=\nu} e^{-j\theta}$ does not depend on (ν,θ), such that (N.9) reduces to:

$$E_{\mathbf{r},\mathbf{\Psi}}\left[\frac{\partial^2}{\partial\nu^2}\ell_{DA}\left(\mathbf{\Psi};\mathbf{r}\right)\right] \propto \Re\left\{Y_1''\left(\nu,\theta\right)\right\}, \tag{N.14}$$

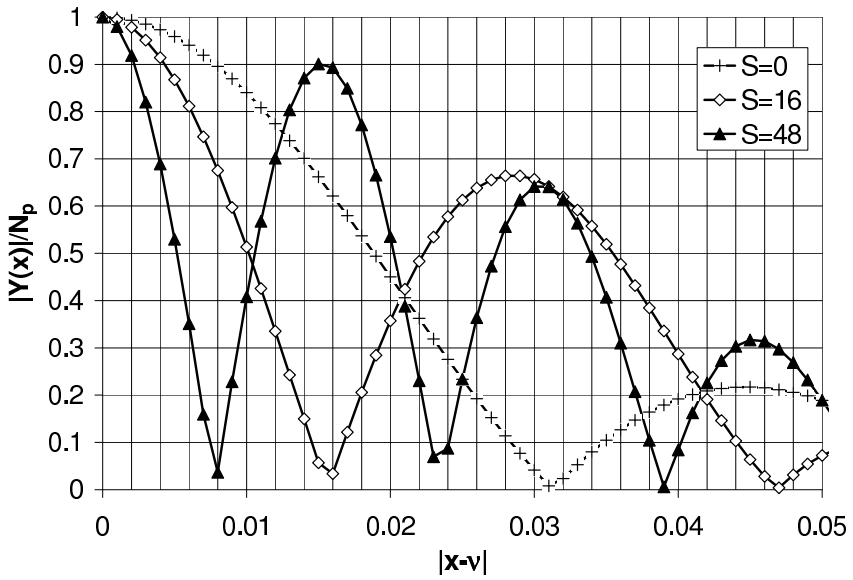

Figure N.2: Effect of increasing the spacing for a fixed number $N_p = 32$ of pilot symbols on the objective function for DA normalized frequency offset estimation.

where $E_{\mathbf{r},\mathbf{\Psi}}[.]$ denotes averaging with respect to $p(\mathbf{r},\mathbf{\Psi})$ (and not $p(\mathbf{r}|\mathbf{\Psi})$). Now, taking into account (N.13) it follows from:

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{\nu}}\left|Y_1(\tilde{\nu};\theta)\right| = \frac{\Re\left\{Y_1(\tilde{\nu};\theta)\,Y_1'(\tilde{\nu};\theta)\right\}}{\left|Y_1(\tilde{\nu};\theta)\right|}, \tag{N.15}$$

that

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{\nu}}\left|Y_1(\tilde{\nu};\theta)\right|\Big|_{\tilde{\nu}=\nu} = 0. \tag{N.16}$$

Similarly, it follows from (N.13), (N.16) and

$$\frac{\mathrm{d}^2\left|Y_1(\tilde{\nu};\theta)\right|^2}{\mathrm{d}\tilde{\nu}^2} = 2\left(\left(\frac{\mathrm{d}}{\mathrm{d}\tilde{\nu}}\left|Y_1(\tilde{\nu};\theta)\right|\right)^2 + \left|Y_1(\tilde{\nu};\theta)\right|\frac{\mathrm{d}^2}{\mathrm{d}\tilde{\nu}^2}\left|Y_1(\tilde{\nu};\theta)\right|\right) \tag{N.17}$$

$$= 2\left(\left|Y_1'(\tilde{\nu};\theta)\right|^2 + \Re\left\{Y_1^*(\tilde{\nu};\theta)\,Y_1''(\tilde{\nu};\theta)\right\}\right), \tag{N.18}$$

that

$$\frac{\mathrm{d}^2}{\mathrm{d}\tilde{\nu}^2}\left|Y_1(\tilde{\nu};\theta)\right|\Big|_{\tilde{\nu}=\nu} = \Re\left\{Y_1''(\nu;\theta)\right\}. \tag{N.19}$$

When we gather the above results (N.14), (N.19) and (N.11) we find that:

$$E_{\mathbf{r},\mathbf{\Psi}}\left[\frac{\partial^2}{\partial\nu^2}\ell_{DA}(\nu,\theta_{m,p})\right] \propto \frac{\partial^2}{\partial\tilde{\nu}^2}\left|Y_0(\tilde{\nu})\right|\Big|_{\tilde{\nu}=\nu},$$

which is exactly what was to be demonstrated.

Appendix O

We consider the function $Y_{K\text{th}-\text{power}}(\nu;\mathbf{r})$ (6.37) for $\kappa_0 = \kappa_{m,s}$ (i.e., the phase shift is estimated at the center of the symbol burst) and evaluated at a trial value $\tilde{\nu}$ of the normalized frequency offset ν:

$$Y_{K\text{th}-\text{power}}(\tilde{\nu};\mathbf{r}) = \sum_{k\in I_s} r(k)^K e^{-j2\pi\tilde{\nu}(k-\kappa_{m,s})}, \tag{O.1}$$

with

$$\kappa_{m,s} = \frac{1}{N_s}\sum_{k\in I_s} k$$

and

$$r(k)^K = (a(k))^K e^{j(2\pi\nu K(k-\kappa_{m,s})+K\theta)} + w_K(k), \forall k \in I_s,$$

where

$$w_K(k) = \sum_{i=1}^{K} (a(k))^{K-i} e^{j(2\pi\nu(K-i)(k-\kappa_{m,s})+(K-i)\theta)} (w(k))^i.$$

We note that $E[w(k)] = 0$ and $E[w(k)w^*(l)] = (N_0/E_s)\delta_{k-l}$. Assuming that the symbol constellation is K-PSK, i.e., for $k \in I_s$:

$$a(k) \in \Omega = \left\{e^{j\frac{2\pi i}{K}} : i = 0, 1, ..., (K-1)\right\},$$

we have:

$$\left(a\left(k\right)\right)^{K}=1$$

and

$$E\left[\left(a^{*}\left(k\right)\right)^{K-i_1}\left(a\left(k\right)\right)^{K-i_2}\right]=\delta_{i_1-i_2}, \tag{O.2}$$

for $(i_1, i_2) \in \{1, 2, ..., K\}$. Taking this into account, $Y_{K\text{th}-\text{power}}\left(\tilde{\nu};\mathbf{r}\right)$ (O.1) can be decomposed as:

$$Y_{K\text{th}-\text{power}}\left(\tilde{\nu};\mathbf{r}\right)=Y_0\left(\tilde{\nu}\right)+n_Y,$$

where $Y_0\left(\tilde{\nu}\right)$ is the noiseless version of $Y_{K\text{th}-\text{power}}\left(\tilde{\nu};\mathbf{r}\right)$ with:

$$\left|Y_0\left(\tilde{\nu}\right)\right| \quad = \quad N_s\left|\text{sinc}\left(N_sK\left(\tilde{\nu}-\nu\right)\right)\right|, \tag{O.3}$$

where $\text{sinc}\left(x\right)$ is defined as $\sin\left(\pi x\right)/\pi x$ and n_Y is a zero-mean statistical fluctuation with variance σ_Y^2, given by:

$$\begin{aligned}\sigma_Y^2 &= E\left[\left|\sum_{k\in I_s} w_K\left(k\right)\right|^2\right],\\ &= N_sE\left[\left|w_K\left(k\right)\right|^2\right],\\ &= N_s\sum_{i=1}^{K}\frac{\left(2i\right)!}{2^ii!}\left(\frac{N_0}{E_s}\right)^i.\end{aligned}$$

The function $\left|Y_0\left(\tilde{\nu}\right)\right|$ (O.3) has the same shape as for the DA case. In fact, the curve labeled $S = 0$ in Fig. N.2 can also be interpreted as the function $\left|Y_0\left(\tilde{\nu}\right)\right|$ (O.3) scaled by N_s as a function of $\frac{N_sK}{32}\left|\tilde{\nu}-\nu\right|$.

Bibliography

[1] N. Noels, H. Steendam and M. Moeneclaey. "The Cramer-Rao bound for phase estimation from coded linearly modulated signals". *IEEE Communications Letters*, 7(5):207–209, May 2003.

[2] N. Noels, H. Steendam and M. Moeneclaey. "The True Cramer-Rao bound for carrier frequency estimation from a PSK signal". *IEEE Transactions on Communications*, 52(5):834–844, May 2004.

[3] N. Noels, H. Steendam and M. Moeneclaey. "Carrier and clock recovery in (turbo) coded systems: Cramer-Rao bound and synchronizer performance". *EURASIP Journal on Applied Signal Processing*, (6):972–980, May 2005.

[4] N. Noels, H. Steendam, M. Moeneclaey and H. Bruneel. "Carrier phase and frequency estimation for pilot-symbol assisted transmission: bounds and algorithms". *IEEE Transactions on Signal Processing*, 53(12):4578–4587, December 2005.

[5] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise and L. Vandendorpe. "A theoretical framework for soft information based synchronization in iterative (turbo) receivers". *EURASIP Journal on Wireless Communications and Networking*, 2005(2):117–129, April 2005.

[6] C. Herzet, N. Noels, V. Lottici, H. Wymeersch, M. Luise, M. Moeneclaey and L. Vandendorpe. "Code-aided turbo synchronization". *Proceedings of the IEEE*, 95(6):1255–1271, June 2007.

[7] N. Noels, H. Steendam and M. Moeneclaey. "True Cramer-Rao bounds for carrier and symbol synchronization". In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Toulouse, France, September 2002.

[8] N. Noels, H. Steendam and M. Moeneclaey. "The true Cramer-Rao bound for phase-independent carrier frequency estimation from a PSK signal". In *Proc. IEEE lobal Telecommunications Conference (GLOBECOM)*, Taipei, Taiwan, November 2002.

[9] N. Noels, H. Steendam and M. Moeneclaey. "The true Cramer-Rao bound for estimating the carrier phase of a convolutionally encoded PSK signal". In *Proc. 9th Symposium on Communications and Vehicular Technology (SCVT)*, Louvain-la-Neuve, Belgium, October 2002.

[10] N. Noels, H. Steendam and M. Moeneclaey. "The impact of the observation model on the Cramer-Rao bound for carrier phase and frequency synchronization". In *Proc. IEEE International Conference on Communications (ICC)*, Anchorage, AK USA, May 2003.

[11] N. Noels, H. Steendam and M. Moeneclaey. "Carrier phase recovery in turbo receivers: Cramer-Rao bound and synchronizer performance". In *Proc. International Symposium on Turbo Codes and Related Topics*, Brest, France, September 2003.

[12] N. Noels, H. Steendam and M. Moeneclaey. "Pilot-symbol assisted carrier synchronization: Cramer-Rao bound and synchronizer performance". In *Proc. IEEE 10th Symposium on Communications and Vehicular Technology (SCVT)*, Eindhoven, The Netherlands, November 2003.

[13] N. Noels, H. Steendam and M. Moeneclaey. "Pilot-symbol assisted iterative carrier synchronization for burst transmission". In *Proc. IEEE International Conference on Communications (ICC)*, Paris, France, June 2004.

[14] N. Noels, H. Steendam and M. Moeneclaey. "On the Cramer-Rao lower bound and the performance of synchronizers for (turbo) encoded systems". In *Proc. IEEE Workshop on Signal Processing Advances in Wireless communications (SPAWC)*, Lisbon, Portugal, July 2004.

[15] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise and L. Vandendorpe. "Turbo-synchronization: an EM algorithm interpretation". In *Proc. IEEE International Conference on Communications (ICC)*, Anchorage, AK USA, May 2003.

[16] S. Steendam, N. Noels and M. Moeneclaey. "Iterative carrier phase synchronization for low-density parity-check coded systems". In *Proc. IEEE*

International Conference on Communications (ICC), Anchorage, AK USA, May 2003.

[17] F. Simoens, H. Wymeersch, N. Noels, H. Steendam and M. Moeneclaey. "Turbo channel estimation for bit-interleaved coded modulation". In *Proc. XI National Symposium of Radio Science*, Poznan, Poland, April 2005.

[18] N. Noels and M. Moeneclaey. "Iterative carrier synchronization techniques in transmission systems protected by a powerful error-correcting code". In *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA USA, November 2007.

[19] N. Noels and M. Moeneclaey. "True Cramer-Rao bound for estimating synchronization parameters from a linearly modulated bandpass signal with unknow data symbols". In *Proc. 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Saint Thomas, VI USA, December 2007.

[20] H. Wymeersch, N. Noels, H. Steendam and M. Moeneclaey. "Synchronization at low SNR: performance bounds and algorithms".
Invited presentation at the Communication Theory Workshop, Capri, Italy, May 2004. Available at http://telin.ugent.be/∼hwymeers.

[21] N. Noels, H. Wymeersch, H. Steendam and M. Moeneclaey. "True Cramer-Rao bound for timing recovery from a bandlimited linearly modulated waveform with unknown carrier phase and frequency". *IEEE Transactions on Communications*, 52(3):473–483, March 2004.

[22] N. Noels, H. Steendam and M. Moeneclaey. "The true Cramer-Rao bound for estimating the time delay of a linearly modulated waveform". In *Proc. IEEE International Conference on Communications (ICC)*, New York, NY USA, May 2002.

[23] N. Noels, H. Steendam and M. Moeneclaey. "Carrier phase tracking from turbo and LDPC coded signals affected by a frequency offset". *IEEE Communications Letters*, 9(10):915–917, October 2005.

[24] N. Noels, H. Steendam, M. Moeneclaey and H. Bruneel. "A maximum-likelihood based feedback carrier synchronizer for turbo-coded systems". In *Proc. IEEE 61st Vehicular Technology Conference (VTC-Spring)*, Stockholm, Sweden, May 2005.

[25] N. Noels, H. Steendam and M. Moeneclaey. "Effectiveness study of code-aided and non-code-aided ML-based feedback phase synchronizers". In *Proc. IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, June 2006.

[26] N. Noels, M. Dervin, M. Moeneclaey and M.-L. Boucheret. "Carrier phase tracking at low signal-to-noise ratio: a performance comparison of a parity-code-aided and a pilot-sybol-assisted approach". In *Proc. Ninth International Workshop on Signal Processing for Space Communications (SPSC)*, Noordwijk, The Netherlands, September 2006.

[27] N. Noels, H. Steendam and M. Moeneclaey. "Performance analysis of ML-based feedback carrier phase synchronizers for coded signals". *IEEE Transactions on Signal Processing*, 55(3):1129–1136, March 2007.

[28] E. Panayırcı, H. Cırpan, M. Moeneclaey and N. Noels. "Blind phase noise estimation in OFDM systems by sequential Monte Carlo method". In *Proc. 5th International Workshop on Multi-Carrier Spread-Spectrum (MC-SS)*, Oberpfaffenhoven, Germany, September 2005.

[29] E. Panayırcı, H. Cırpan, M. Moeneclaey and N. Noels. "Blind phase noise estimation in OFDM systems by sequential Monte Carlo method". *European Transactions on Telecommunications*, 17(6):685–693, December 2006.

[30] E. Panayırcı, H. Cırpan, M. Moeneclaey and N. Noels. "Blind phase noise estimation and data detection based on SMC technique and unscented filtering". In *Proc. IEEE European Signal Processing Conference (EUSIPCO)*, Florence, Italy, September 2006.

[31] E. Panayırcı, H. Cırpan, M. Moeneclaey and N. Noels. "Blind data detection in the presence of PLL phase noise by sequential Monte Carlo method". In *Proc. IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, June 2006.

[32] U. Mengali and A.N. D'Andrea. *Synchronization techniques for digital receivers*. Plenum Press, 1997.

[33] J.G. Proakis. *Digital communications*. McGraw-Hill, 4th edition, 2001.

[34] H. Meyr, M. Moeneclaey and S.A. Fechtel. *Synchronization, channel estimation, and signal processing*, volume 2 of *Digital Communication Receivers*. John Wiley & Sons, 1997.

[35] R.G. Gallager. "Low density parity-check codes". *IRE Transactions on Information Theory*, 8:21–29, January 1962.

[36] D.J.C. MacKay. "Good error-correcting codes based on very sparse matrices". *IEEE Transactions on Information Theory*, 45(2):399–431, March 1999.

[37] A. Viterbi. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm". *IEEE Transactions on Information Theory*, 13(2):260–269, April 1967.

[38] C. Berrou, A. Glavieux and P. Thitimajsima. "Near Shannon limit error-correcting coding and decoding: turbo codes". In *Proc. IEEE International Conference on Communications (ICC)*, pages 1064–1070, Geneva, Switzerland, May 1993.

[39] S. Benedetto and G. Montorsi. "Unveiling turbo codes: some results on parallel concatenated coding schemes". *IEEE Transactions on Information Theory*, 42(2):409–428, March 1996.

[40] C.E. Shannon. "A mathematical theory of communication". *Bell System Technical Journal*, 27:379–423, July 1948.

[41] G. Ungerboeck. "Channel coding with multilevel/phase signal". *IEEE Transactions on Information Theory*, 28(1):55–66, January 1982.

[42] E. Zehavi. "8-PSK trellis codes for Rayleigh fading channels". *IEEE Transactions on Commumications*, 41:873–883, May 1992.

[43] G. Caire, G. Taricco and E. Biglieri. "Bit-interleaved coded modulation". *IEEE Transactions on Information Theory*, 44(3):927–946, May 1998.

[44] F. Simoens, H. Wymeersch, H. Bruneel and M. Moeneclaey. "Multi-dimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling". *IEEE Communications Letters*, 9(5):453–455, May 2005.

[45] H.-A. Loeliger. "An introduction to factor graphs". *IEEE Signal Processing Magazine*, 21(1):28–41, January 2004.

[46] H. Wymeersch. "Software radio algorithms for coded transmission". *Phd. dissertation, Faculty of Engineering, Ghent University*, 2005.
Available at http://telin.ugent.be/∼hwymeers.

[47] H.L. Van Trees. *Detection, estimation, and modulation theory, Part I.* John Wiley & Sons, 2001.

[48] J. Dauwels. "On graphical models for communications and machine learning: algorithms, bounds, and analog implementation". *Phd. dissertation, Swiss Federal Institute of Technology, Zurich*, 2006.
Available at http://www.dauwels.com.

[49] B. Bobrovsky, E. Mayer-Wolf and M. Zakai. "Some classes of global Cramer-Rao bounds". *Annals of Statistics*, 15(4):1421–1438, December 1987.

[50] N.A. D'Andrea, U. Mengali and R. Reggiannini. "The modified Cramer-Rao bound and its application to synchronization problems". *IEEE Transactions on Communications*, 42(2/3/4):1391–1399, March 1994.

[51] F. Gini, R. Reggiannini and U. Mengali. "The modified Cramer-Rao bound in vector parameter estimation". *IEEE Transactions on Communications*, 46(1):52–60, January 1998.

[52] J.-K. Wolf. "Efficient maximum likelihood decoding of linear block codes using a trellis". *IEEE Transactions on Information Theory*, 31(1):76–80, January 1978.

[53] L.R. Bahl, J. Cocke, F. Jelinek and J. Raviv. "Optimal decoding of linear codes for minimising symbol error rate". *IEEE Transactions on Information Theory*, 20(3):284–287, March 1974.

[54] M. Moeneclaey. "On the true and the modified Cramer-Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters". *IEEE Transactions on Communications*, 46(11):1536–1544, November 1998.

[55] W.G. Cowley. "Phase and frequency estimation for PSK packets: bounds and algorithms". *IEEE Transactions on Communications*, 44(1):26–28, January 1996.

[56] F. Rice, B. Cowley, B. Moran and M. Rice. "Cramer-Rao lower bounds for QAM phase and frequency estimation". *IEEE Transactions on Communications*, 49(9):1582–1591, September 2001.

[57] H. Steendam and M. Moeneclaey. "Low-SNR limit of the Cramer-Rao bound for estimating the carrier phase and frequency of a PAM, PSK or QAM waveform". *IEEE Communications Letters*, 5(5):215–217, May 2001.

[58] J. W. Cooley and J. W. Tukey. "An algorithm for the machine calculation of complex Fourier series". *Mathematics of Computation*, 19:297–301, April 1965.

[59] D. C. Rife and R. R. Boorstyn. "Single-tone parameter estimation from discrete-time observations". *IEEE Transactions on Information Theory*, 20(5):591–598, September 1974.

[60] B. Beahan. "Frequency estimation of partitioned reference symbol sequences". *Master thesis, University of South Australia*, 2001. Available at http://www.itr.unisa.edu.au/research/pubs/thesis.php.

[61] H. Steendam and M. Moeneclaey. "Low-SNR limit of the Cramer-Rao bound for estimating the time delay of a PAM, PSK or QAM waveform". *IEEE Communications Letters*, 5(1):31–33, January 2001.

[62] M. Moeneclaey and G. de Jonghe. "ML-oriented NDA carrier synchronization for general rotationally symmetric signal constellations". *IEEE Transactions on Communications*, 42(8):2531–2533, August 1994.

[63] A.J. Viterbi and A.M. Viterbi. "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission". *IEEE Transactions on Information Theory*, 29(4):543–551, July 1983.

[64] H. Wymeersch and M. Moeneclaey. "Iterative Code-Aided ML Phase Estimation and Phase Ambiguity Resolution". *Eurasip Journal on Applied Signal Processing*, 2005(6):981–988, May 2005.

[65] E. Cacciamani and C. Wolejsza. "Phase-ambiguity resolution in a four-phase PSK communications system". *IEEE Transactions on Communications*, 19(6):1200–1210, December 1971.

[66] C. Langlais and M. Helard. "Phase carrier recovery for turbo codes over a satellite link with the help of tentative decisions". In *Proc. International Symposium on Turbo Codes and Related Topics*, Brest, France, September 2000.

[67] C. Morlet, M.-L. Boucheret and I. Buret. "A carrier phase estimator for multi-media satellite payloads suited for RCS coding schemes". In *Proc. International Conference on Communications (ICC)*, pages 455–459, New Orleans, USA, June 2000.

[68] V. Lottici and M. Luise. "Embedding carrier phase recovery into iterative decoding of turbo-coded linear modulations". *IEEE Transactions on Communications*, 52(4):661–669, April 2004.

[69] L. Zhang and A. Burr. "A novel carrier phase recovery method for turbo coded QPSK systems". In *Proc. European Wireless 2000*, Florence, Italy, February 2000.

[70] T. Moon. "The expectation-maximization algorithm". *IEEE Signal Processing Magazine*, 13(6):47–60, 1996.

[71] A.P. Dempster, N.M. Laird and D.B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society*, 39(1):1–38, 1977. Series B.

[72] C. F. J. Wu. "On the convergence properties of the EM algorithm". *The Annals of Statistics*, 11(1):95–109, 1983.

[73] X. Wu and H. Xiang. "Iterative carrier phase recovery methods in turbo receivers". *IEEE Communications Letters*, 9(8):735–737, August 2005.

[74] C. Herzet, X. Wautelet, V. Ramon and L. Vandendorpe. "Iterative synchronization: EM algorithm versus Newton-Raphson approach". In *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 393–396, Toulouse, France, May 2006.

[75] J. Dauwels and H.-A. Loeliger. "Phase estimation by message passing". In *Proc. IEEE International Conference on Communications (ICC)*, Paris, France, June 2004.

[76] S. Godtmann, A. Pollok, N. Hadaschik, W. Steinert, G. Ascheid and H. Meyr. "Joint iterative synchronization and decoding assisted by pilot symbols". In *Proc. IST Mobile and Wireless Communication Summit*, Myconos, Greece, June 2006.

[77] A. D'Amico, A. N. D'Andrea and R. Reggiannini. "Efficient non-data-aided carrier and clock recovery for satellite DVB at very low signal-to-noise ratios". *IEEE Journal on selected areas in communications*, 19(12):2320–2330, December 2001.

[78] A. Anastasopoulos and K.M. Chugg. "Adaptive iterative detection for phase tracking in turbo-coded systems". *IEEE Transactions on Communications*, 49(12):2135–2144, December 2001.

[79] G. Colavolpe, G. Ferrari and R. Raheli. "Noncoherent iterative (turbo) decoding". *IEEE Transactions on Communications*, 48(9):1488–1498, September 2000.

[80] B. Mielczarek and A. Svensson. "Phase offset estimation using enhanced turbo decoders". In *Proc. IEEE International Conference on Communications (ICC)*, New York, NY USA, May 2002.

[81] C. Herzet, V. Ramon and L. Vandendorpe. "A theoretical framework for iterative synchronization based on the sum-product and the expectation-maximization algorithms". *IEEE Transactions on Signal Processing*, 55(5-1):1644–1658, May 2007.

[82] H. Stark and J. W. Woods. *Probability and Random Processes with Applications to Signal Processing*. Prentice Hall, 3rd edition, 2002.

[83] W. Oh and K. Cheun. "Joint decoding and carrier phase recovery algorithm for turbo codes". *IEEE Communications Letters*, 5(9):375–377, September 2001.

[84] L. Lu and G. Wilson. "Synchronization of turbo coded modulation systems at low SNR". In *Proc. International Conference on Communications (ICC)*, pages 428–432, Atlanta, GA USA, June 1998.

[85] G. Colavolpe, A. Barbieri and G. Caire. "Algorithms for iterative decoding in the presence of strong phase noise". *IEEE Journal on Selected Areas in Communications on Communications*, 23(9):1748–1757, September 2005.

[86] E. Panayırcı, H. Cırpan and M. Moeneclaey. "A sequential Monte-Carlo method for blind phase noise estimation and data detection". In *Proc. IEEE European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, September 2005.

[87] M. Moeneclaey and U. Mengali. "Sufficient conditions on trellis-coded modulation for code-independent synchronizer performance". *IEEE Transactions on Communications*, 38(5):595–601, May 1990.

[88] R.-M. Tanner. "A recursive approach to low complexity codes". *IEEE Transactions on Information Theory*, 27(5):76–80, September 1981.